

RESEARCH ARTICLE

Characterization of two related *Erwinia* myoviruses that are distant relatives of the PhiKZ-like Jumbo phages

Daniel K. Arens¹, T. Scott Brady¹, John L. Carter¹, Jenny A. Pape¹, David M. Robinson¹, Kerri A. Russell¹, Lyndsay A. Staley¹, Jason M. Stettler², Olivia B. Tateoka¹, Michelle H. Townsend¹, Kiara V. Whitley¹, Trevor M. Wienclaw¹, Taryn L. Williamson², Steven M. Johnson¹, Julianne H. Grose^{1*}

1 Microbiology and Molecular Biology Department, Brigham Young University, Provo, Utah, United States of America, **2** Plant and Wildlife Sciences Department, Brigham Young University, Provo, Utah, United States of America

* julianne_grose@byu.edu



OPEN ACCESS

Citation: Arens DK, Brady TS, Carter JL, Pape JA, Robinson DM, Russell KA, et al. (2018) Characterization of two related *Erwinia* myoviruses that are distant relatives of the PhiKZ-like Jumbo phages. PLoS ONE 13(7): e0200202. <https://doi.org/10.1371/journal.pone.0200202>

Editor: Szabolcs Semsey, Niels Bohr Institute, DENMARK

Received: November 3, 2017

Accepted: June 21, 2018

Published: July 6, 2018

Copyright: © 2018 Arens et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was graciously funded by the Department of Microbiology and Molecular Biology and the College of Life Sciences at Brigham Young University, as well as a private individual donor. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Bacteriophages are a major force in the evolution of bacteria due to their sheer abundance as well as their ability to infect and kill their hosts and to transfer genetic material. Bacteriophages that infect the *Enterobacteriaceae* family are of particular interest because this bacterial family contains dangerous animal and plant pathogens. Herein we report the isolation and characterization of two jumbo myovirus *Erwinia* phages, RisingSun and Joad, collected from apple trees. These two genomes are nearly identical with Joad harboring two additional putative gene products. Despite mass spectrometry data that support the putative annotation, 43% of their gene products have no significant BLASTP hit. These phages are also more closely related to *Pseudomonas* and *Vibrio* phages than to published *Enterobacteriaceae* phages. Of the 140 gene products with a BLASTP hit, 81% and 63% of the closest hits correspond to gene products from *Pseudomonas* and *Vibrio* phages, respectively. This relatedness may reflect their ecological niche, rather than the evolutionary history of their host. Despite the presence of over 800 *Enterobacteriaceae* phages on NCBI, the uniqueness of these two phages highlights the diversity of *Enterobacteriaceae* phages still to be discovered.

Introduction

The existence of bacteriophages has been known since the early 1900's when Frederick Twort and Felix d'Herelle independently isolated phage [1]. Phages are now considered the most abundant source of biomass on the planet [2] and contribute heavily to the evolution of bacteria because of their ability to infect and lyse different bacterial strains as well as their ability to transfer genetic information [3, 4]. Due to this transfer of genetic information, phages have been shown to be required for the pathogenicity of several bacterial strains, such as pathogenic *V. cholerae*, *C. diphtheriae*, and *E. coli* strains [5–7]. This incredible bacterial host specificity

Competing interests: JHG has a licensing agreement for a bacteriophage treatment of fire blight, for which *Erwinia amylovora* is the causative agent. Although the private donor is affiliated with a company that has licensed other *Erwinia* phages for commercial use, the funds provided for this specific study were a gift and the specific phages used in this study are not commercialized. There are no patents, products in development or marketed products to declare. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

has led to many attempts to use them as diagnostic and therapeutic agents [8]. d'Herelle was one of the first to put this into practice by using phage to treat and cure patients with *Bacillus* based dysentery [9].

Enterobacteriaceae is one of the most highly studied bacterial families with over 50 accepted genera. These genera include several well-known animal pathogens such as *Enterobacter*, *Escherichia*, *Klebsiella*, *Salmonella*, *Shigella*, and *Yersinia*, as well as plant pathogens such as *Erwinia* and *Dickeya* (for a recent discussion on the classification of this family see [10]). This family is a major health concern in the United States with the Centers for Disease Control and Prevention (CDC) citing carbapenem resistant and extended spectrum β -lactamase (ESBL) producing *Enterobacteriaceae* as “urgent” and “serious” threats [11]. Plant pathogens are also of great concern, with *Erwinia* infections causing over \$100 million in agricultural loss per year [12]. Studying the evolution of this family of bacteria, including the phages that infect them, is critical to controlling many health and agricultural concerns.

Over 800 phages that infect members of the *Enterobacteriaceae* family have been isolated, sequenced, and deposited in NCBI. Most of these phages infect genera with common animal pathogens including *Escherichia*, *Salmonella*, *Shigella*, and *Klebsiella*. Phages of the plant pathogens have also been recently deposited, including those that infect *Erwinia amylovora*, the causative agent of fire blight [13]. When infected, fire blight causes the leaves of the *Rosaceae* plants to dry and shrivel up, giving the appearance of being scorched. Another similar wilting disease infecting the *Cucurbita* genus, which includes squash and pumpkin, is caused by *Erwinia tracheiphila*. It has been reported to cause millions of dollars of agricultural loss in the northeastern United States [14]. Therefore, the study of *Erwinia* phages may aid in understanding and treating multiple devastating agricultural diseases. Currently, 45 *Erwinia* specific phages have been isolated and deposited on NCBI, 25 of which were discovered by our group.

Herein we report the isolation and characterization of two of these *Erwinia* phages, vB_EamM_RisingSun (RisingSun) and vB_EamM_Joad (Joad). These phages have only very distant relationships to other published phages and are highly similar to one another, with Joad containing two additional genes. BLASTP hits to putative annotated ORFs indicate that much of their genomes are composed of novel proteins with no BLASTP hit, while many of those having BLASTP hits harbor closer relationships to *Pseudomonas* and *Vibrio* phages as opposed to *Enterobacteriaceae* phages.

Methods

Phage isolation, sequencing and host range

Both Joad and RisingSun were isolated from apple blossom samples collected in Payson, Utah. Blossoms were crushed with a mortar and pestle and the resulting debris was added to an exponential culture of *Erwinia amylovora* ATCC 29780 [15, 16]. The enrichment culture was harvested by centrifugation 48 hours later, filtered with a 0.45 μ M filter, and used to infect a fresh culture of bacteria. Three plaque purifications were performed, after which a high titer lysate was made in liquid broth. Phage DNA was extracted from this lysate using the Phage DNA Isolation Kit (Norgen Biotek Corporation), and was sequenced, assembled, and annotated as previously described [15]. Host range was performed by spotting 5 μ L of lysate onto 0.5 mL of bacteria imbedded in LB top agar. Positive spots were verified by plaque assay. Bacterial strains used included *Erwinia amylovora* ATCC 29780, *Erwinia amylovora* EA110 [16], *Pantoea agglomerans* E325 [17], *Pantoea vegans* C9-1 [18], *Dickeya chrysanthemi* ATCC 11663 [19], the common clinical strain *Pseudomonas aeruginosa* Boston 41501 ATCC 27853 [20], *Pseudomonas chlororaphis* ATCC13985 [21], *Vibrio cholerae* ATCC 14035 (originally deposited by Standards Lab., London) [22], *Salmonella enterica typhimurium* LT2 (a generous gift

from John Roth, UC Davis), *Enterobacter cloacae* ATCC 13047 (deposited by the CDC) [23], and *E. coli* BW25113 [24].

Electron microscopy

Samples were prepared for transmission electron microscopy by placing 20 μ L of high-titer phage lysate on a 200-mesh copper carbon type-B electron microscope grid for two to five minutes. Excess lysate was wicked away and the grid was then stained for one minute using 2% phosphotungstic acid. The grid was then briefly dipped into distilled water and excess liquid was wicked away. Phages were imaged at the BYU Microscopy Center. The phages were measured for capsid width as well as tail length and width using ImageJ software [25].

Mass spectrometry

Samples were prepared according to the methods of Guttman et al. [26]. Briefly, (all concentrations are final concentrations) fresh lysates were diluted with TNE (50 mM Tris pH 8.0, 100 mM NaCl, 1 mM EDTA) buffer and RapiGest SF reagent (Waters Corp.) was added to 0.1% before 5 min of boiling. Next, samples were incubated at 37°C for 30 min in the presence of 1 mM TCEP (Tris (2-carboxyethyl) phosphine). Iodoacetamide (0.5 mg/ml) was used to carboxymethylate the samples for 30 min at 37°C. Carboxymethylation was neutralized with 2 mM TCEP. Using a trypsin:protein ratio of 1:50 samples were digested overnight at 37°C. Next, 250 mM HCl was used to degrade the RapiGest for 1 hr at 37°C. Samples were then centrifuged for 30 min at 4°C and 14000 rpm. In new tubes, peptides were extracted from the soluble fractions by desalting using Aspire RP30 desalting columns (Thermo Scientific).

High pressure liquid chromatography (HPLC) coupled with tandem mass spectrometry (LC-MS/MS) using nano-spray ionization was used to analyze the trypsin-digested peptides according to the method of McCormack et al. [27]. Experiments were performed on a Triple-TOF 5600 hybrid mass spectrometer (ABSCIEX) interfaced with nano-scale reversed-phase HPLC (Tempo) using a 10 cm-100 μ m ID glass capillary packed with 5- μ m C18 Zorbax™ beads (Agilent Technologies, Santa Clara, CA). The peptides were eluted from the C18 packed capillary tubes into the mass spectrometer using a linear gradient of Acetonitrile (ACN) (5–60% generated from two buffers: buffer A with 98% H₂O, 2% ACN, 0.2% formic acid, and 0.005% TFA, and buffer B with 100% ACN, 0.2% formic acid, and 0.005% TFA) at a flow rate of 250 μ L/min for 1 hr.

MS/MS data were acquired in a data-dependent manner in which the MS1 data were acquired for 250 ms at m/z of 400 to 1250 Da and the MS/MS data were acquired from m/z of 50 to 2,000 Da. For independent data acquisition (IDA) parameters of MS1-TOF for 250 milliseconds, followed by 50 MS2 events of 25 milliseconds each were used. The IDA criteria were set at over 200 counts threshold, charge state of plus 2–4 with 4 seconds exclusion window. Finally, the collected data were analyzed using MASCOT® (Matrix Sciences) and Protein Pilot 4.0 (ABSCIEX) for peptide identifications.

Analysis of RisingSun and Joad phage genomes

Gepard [28] was used to create three dot plots, one with whole genome sequences, one with major capsid protein (MCP) sequences, and one with terminase sequences. Putative major capsid proteins and terminase proteins from phage Joad were used in BLASTP [29–31] analysis to find related phages [2, 32, 33]. Accession numbers that were used are: whole genome accession numbers (Joad [MF459647], RisingSun [MF459646], Pseudomonas phage EL [NC_007623.1], Pseudomonas phage OBP [NC_016571.1], Vibrio phage pTD1 [AP017972.1], Vibrio phage VP4B [KC131130.1], Pseudomonas phage phiKZ [AF399011.1]), MCP accession

numbers (Joad [ASU03832.1], RisingSun [ASU03587.1], *Pseudomonas* phage EL [YP_418111.1], *Vibrio* phage pTD1 [BAW98274.1], *Vibrio* phage VP4B [AGB07257.1], *Pseudomonas* phage OBP [YP_004958031.1], *Pseudomonas* phage phiKZ [AAL83021.1]) and terminase accession numbers (Joad [ASU03673.1], RisingSun [ASU03430.1] *Pseudomonas* phage EL [YP_418044.1], *Vibrio* phage pTD1 [BAW98365.1], *Vibrio* phage VP4B [AGB07167.1], *Pseudomonas* phage OBP [YP_004957913.1], *Pseudomonas* phage phiKZ [NP_803591.1]). Kalign [34–38] was used to determine average nucleotide identity (ANI) of the phage genomes.

Motif analysis and identification

The Center for Phage Technology Galaxy Server (<https://cpt.tamu.edu/galaxy-pub/>) and MEME [39] were used to scan the phage genome for significant motifs with an e-value less than 10^{-7} . The Galaxy Server was able to scan the entire genome at once. FIMO [40] was used to search the phage genome for motifs found in the Galaxy results that passed our significance threshold (q value < 0.01) and determined the exact positions of the motif(s) in the entire genome. We then used DNA Master [41] and Phamerator [42] to analyze the genes neighboring the motifs to determine putative transcription patterns.

Results and discussion

Phage isolation and sequencing

Joad and RisingSun were isolated from apple tree samples that appeared to be infected with fire blight. DNA analysis suggests Joad and RisingSun are Jumbo phages [43] with genome sizes of 235374 bp and 235108 bp respectively (a summary of their genomes is provided in Table 1). A search for tRNA's using tRNA ScanSE [44] returned no tRNA results. No rigorous testing for lysogeny formation has been performed, but their clear plaque morphology and ease in obtaining higher titers suggest they may be lytic phages. This conclusion is supported by a BLASTP analysis of the putative Major Capsid Protein (MCP) from Joad, which has only close BLASTP hits to phage and none to bacterial genomes. In a recent analysis of phage, Casjens et al. found that MCP's from temperate phages generally have BLASTP hits that are >70% identity in bacterial genomes [45].

Phage morphology

Transmission electron microscopy (TEM) analysis revealed the large nature of these phages. From three independent TEM images of RisingSun (Fig 1), the average capsid width was 143.2 +/- 6.0 nm, tail width was 23.2 +/- 2.4 nm, and tail length was 206.8 +/- 3.6 nm, consistent with the large genome size reported above. The measurements from the single Joad image obtained were within the standard deviation of RisingSun measurements. The viral morphology, icosahedral capsids attached to long contractile tails, is consistent with these phages belonging to the jumbo *myoviridae* bacteriophages [32].

Table 1. General characteristics of *Erwinia amylovora* phages Joad and RisingSun.

Phage Name	GenBank Accession #	Fold Coverage	Genome Length (bps)	ORFs	tRNAs	GC%
Joad	MF459647	522.2	235374	245	none	48.29
RisingSun	MF459646	138.6	235108	243	none	48.32

<https://doi.org/10.1371/journal.pone.0200202.t001>

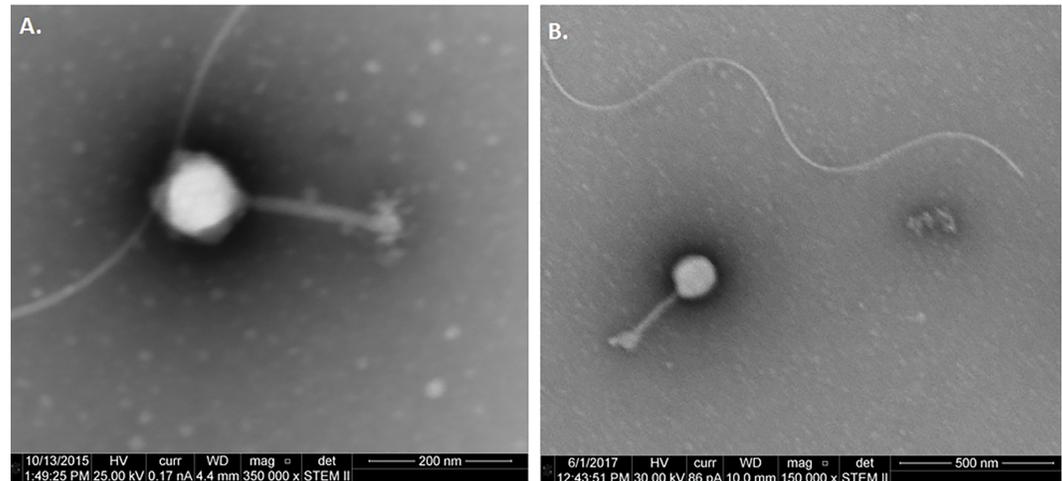


Fig 1. Transmission electron microscopy revealed (A) RisingSun and (B) Joad as *Myoviridae*.

<https://doi.org/10.1371/journal.pone.0200202.g001>

Genomic analysis

Whole genome dot plot comparisons (Fig 2A) were performed using the whole genome sequences of Joad, RisingSun, and any phages that were retrieved from a BLASTP analysis of their putative major capsid (MCP) and terminase proteins. The dot plot reveals two distinct clusters of phages, and low relatedness between singletons (unique phages unrelated to others in the group). Clusters are defined here similarly to other studies, as two or more phages with sequence similarity over at least half of the genome [2]. Joad and RisingSun constitute one cluster with *Pseudomonas* phage EL being a distant relation in the gray area of cluster boundary, while the second cluster is comprised of two *Vibrio* phages, VP4B and pTD1. Whole genome Average Nucleotide Identity (ANI) data shown in Table 2A support the clusters identified in Fig 2A with a 96% identity match between Joad and RisingSun. These two phages also had 46% similarity compared to EL and only 36% similarity compared to OBP and phiKZ, while low similarity is seen between Joad and the singleton phages. A 75% identity match is also observed within the vibrio phage cluster of VP4B and pTD1. ANI data show that Joad and RisingSun differ in genome size by 266 nucleotides, Joad having the larger genome, in an otherwise highly similar genome (~96.6% nucleotide identity). In addition to the distinct clusters, Fig 2A shows a distant relationship between Joad and EL, indicative of divergent evolution and lateral gene transfer that has occurred.

A dot plot of the Major Capsid Protein (MCP) and terminase amino acid sequences (Fig 2B & 2C) and percent identity data (Table 2B) support this distant relationship between Joad and EL. Table 2B shows a 57% identity between the MCP amino acid sequences of the two phages, whereas *Vibrio* phages pTD1 and VP4B MCP amino acid sequences are highly similar at nearly 90% amino acid identity. The terminase dot plot, Fig 2C, and percent identity data in Table 2C continue to support a distant relationship between Joad and EL.

Both the dot plot and ANI indicate that the Joad cluster is markedly different from the *Vibrio* cluster and EL phage consistent with the weak similarity of the MCP and terminase amino acid sequences. Comparing Joad to OBP and phiKZ, which have similar terminase proteins, shows little relation and confirms the distinctiveness of Joad. Since terminase conservation has been shown to reflect phage packaging mechanisms [46] the terminase similarity to phiKZ, although distant (~28% identity) suggests these phages package DNA by a headful mechanism [47].

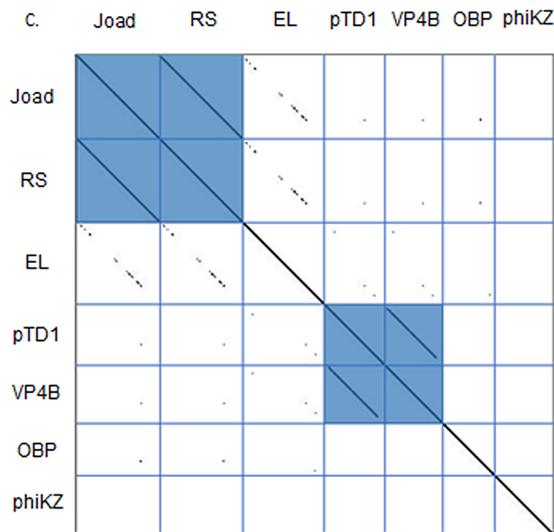
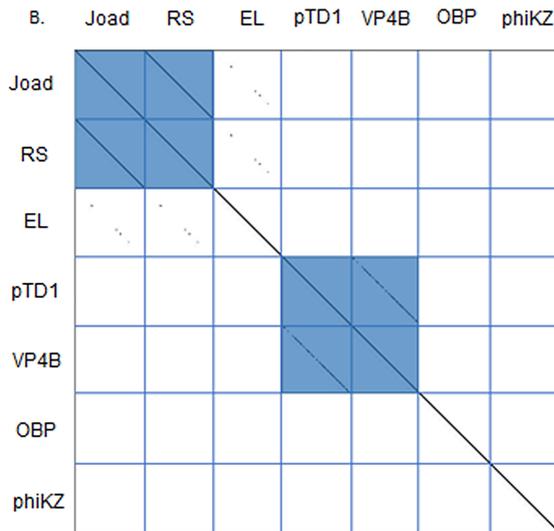
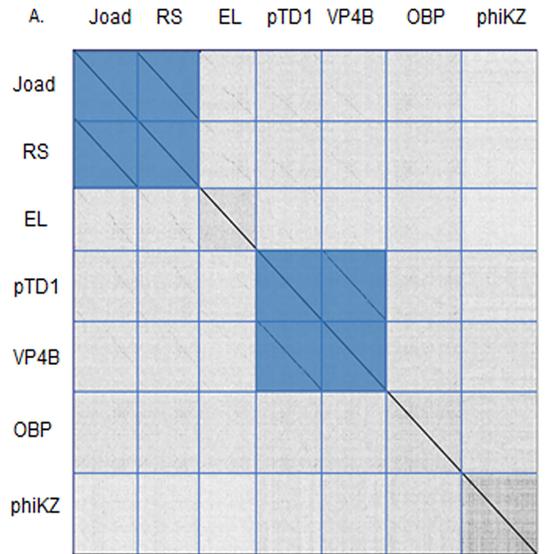


Fig 2. Dot plot comparisons for whole genome nucleotide sequences, MCP amino acid sequences, and terminase amino acid sequences of seven phages. Blue horizontal and vertical lines were added to show genome boundaries. *Erwinia* phages: Joad, RisingSun (RS). *Vibrio* phages: pTD1, VP4B. *Pseudomonas* phages: EL, OBP, phiKZ. A) Nucleotide dot plot shows genome similarity between seven phage genomes. Two distinct clusters are shown, the Joad and RS cluster and the *Vibrio* phage cluster. B) Dot plot comparison for MCP amino acid sequences of seven phages. The clusters are shown between Joad and RS and between *Vibrio* phages pTD1 and VP4B. C) Dot plot comparison for terminase amino acid sequences of seven phages. Two clusters are formed between Joad and RS and between *Vibrio* phages pTD1 and VP4B.

<https://doi.org/10.1371/journal.pone.0200202.g002>

This analysis is consistent with analysis of our whole-genome phage sequencing raw results for RisingSun analyzed by Phageterm [48] which also suggested a headful packaging mechanism.

Whole-proteome comparison of Joad and RisingSun

Phamerator [42] was used to produce full-genome comparison maps for both Joad and RisingSun which were modified for simplicity (Fig 3). Consistent with the mild differences seen in ANI analysis, the genomes encode nearly identical gene products (colored with similar coloring based on the Phamerator default values of greater than 32.5% identity by BLASTP and less than 1e-50 e-value from ClustalO), with many having 100% identity. The most obvious difference is the two genes present in Joad that are not present in RisingSun, of which gene product

Table 2. Joad and RisingSun are a unique cluster of phages when compared to others using whole genomes, MCP amino acid sequences, and terminase amino acid sequences. 2A) Similarity of seven phage genomes according to Average Nucleotide Identity (ANI). 2B) Percent identity of major capsid proteins (MCP) amongst seven phages. 2C) Percent identity of terminase proteins amongst seven phages. Percent identity was determined by BLASTP analysis.

2A		ANI of Whole genome						
	Joad	Rising Sun	EL	pTD1	VP4B	OBP	phiKZ	
Joad	100							
RisingSun	96.61	100						
EL	45.83	45.73	100					
pTD1	38.24	37.86	35.21	100				
VP4B	38.8	38.41	35.67	75.14	100			
OBP	36.53	36.17	33.7	37.08	41.11	100		
phiKZ	36.25	35.94	33.5	37.04	41.2	46.98	100	
2B		Identity of MCP gene products						
	Joad	Rising Sun	EL	pTD1	VP4B	OBP	phiKZ	
Joad	100							
RisingSun	100	100						
EL	57.22	57.22	100					
pTD1	40.78	40.78	21.98	100				
VP4B	40.48	40.48	38.27	89.74	100			
OBP	30.71	30.71	31.18	33.61	34.90	100		
phiKZ	23.58	23.58	21.98	21.60	22.19	25.29	100	
2C		Terminase gene products						
	Joad	Rising Sun	EL	pTD1	VP4B	OBP	phiKZ	
Joad	100							
RisingSun	99.63	100						
EL	56.43	56.43	100					
pTD1	53.63	53.63	50.42	100				
VP4B	53.96	53.96	50.08	95.75	100			
OBP	48.49	48.49	48.31	47.59	47.97	100		
phiKZ	27.71	27.71	27.29	31.54	32.24	28.73	100	

<https://doi.org/10.1371/journal.pone.0200202.t002>

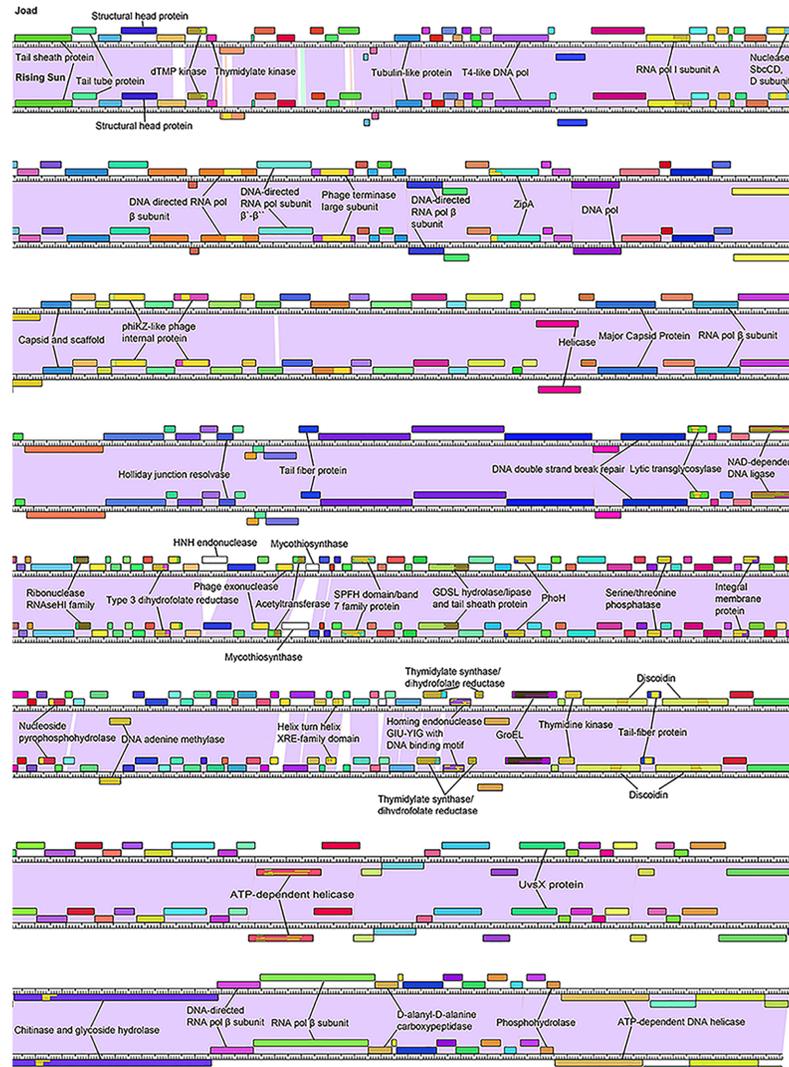


Fig 3. Whole genome comparison map between the phages Joad and RisingSun adapted from Phamerator [42]. Boxes on top of the genome ruler are genes expressed on the forward strand, while boxes under the genome ruler are genes expressed on the reverse strand. The colored boxes categorize homologous proteins. The purple between the two genomes represents high nucleotide similarity, while the white between the two nucleotides represents nucleotide variation. Annotated functions were collected through BLAST and Phamerator searches. Abbreviations include: Discoidin (Coagulation factor also known as F5/8 type C domain known as the discoidin (DS) domain family); GroEL (GroEL-like type 1 chaperonin protein); PhoH (Phosphate starvation protein PhoH); ZipA (cell division protein ZipA).

<https://doi.org/10.1371/journal.pone.0200202.g003>

122 had a BLASTP hit to an HNH endonuclease and the other encodes a protein with unknown function. Not every gene with a BLASTP or Phamerator hit is labeled on this map as many have an unknown function or are indiscriminant structural proteins. Proteins of significance are discussed in the following section “Interesting proteins”, for which Fig 3 will serve as a reference.

Interesting proteins and host range

Joad and RisingSun appear to be two very similar Jumbo phages [43] with only distant relationship to other phages. Due to the fact that these phages are so similar, we will be specifically

referring to the genes, gene functions, and gene products of RisingSun. Of the 243 putative gene products in the RisingSun genome ~43% have no known function and do not have any significant BLASTP hits (e-value of 10^{-7} or less). This large proportion of proteins with no BLASTP hit (~43%) represents proteins that have not been previously found in nature [49, 50]. This finding, combined with their nucleotide dissimilarity discussed above, further sets RisingSun and Joad apart. The remaining gene products are represented as those with BLASTP hits but no known function (NKF) (~24%) and those with BLASTP hits and putative or known functions (~33%) (Fig 4A). Of the ~33% of gene products with known functions, ~36% of them are unspecified structural proteins and another ~12% represent major capsid and tail fiber proteins. The remaining gene products with putative function are primarily putative enzymes, namely those involved with DNA and RNA synthesis including an NAD-dependent DNA ligase (gp108), RNA polymerase beta subunit (gp29), and a helix-turn-helix XRE-family domain among others (gp180) (Fig 4B).

The RisingSun gene products that have BLASTP hits are mainly homologous to other phage gene products. Specifically, of the 140 gene products with BLASTP hits, 81% correspond to *Pseudomonas* phage gene products. Gene products from phages EL, OBP, and phiKZ match 74%, 53%, and 24% of NKF/putative function gene products, respectively, while other *Pseudomonas* phages match 37%. *Vibrio* phages also showed a marked similarity to the RisingSun proteome with 63% of gene products with a BLASTP hit from a *Vibrio* phage gene product. Phages VP4B and pTD1 match 58% and 60%, respectively, and other *Vibrio* phages match 24% of NKF/putative function gene products. Due to the high similarity of gene products with a BLASTP hit, we further analyzed the entire proteome of RisingSun in comparison to the *Pseudomonas* and *Vibrio* phages EL, OBP, VP4B and pTD1.

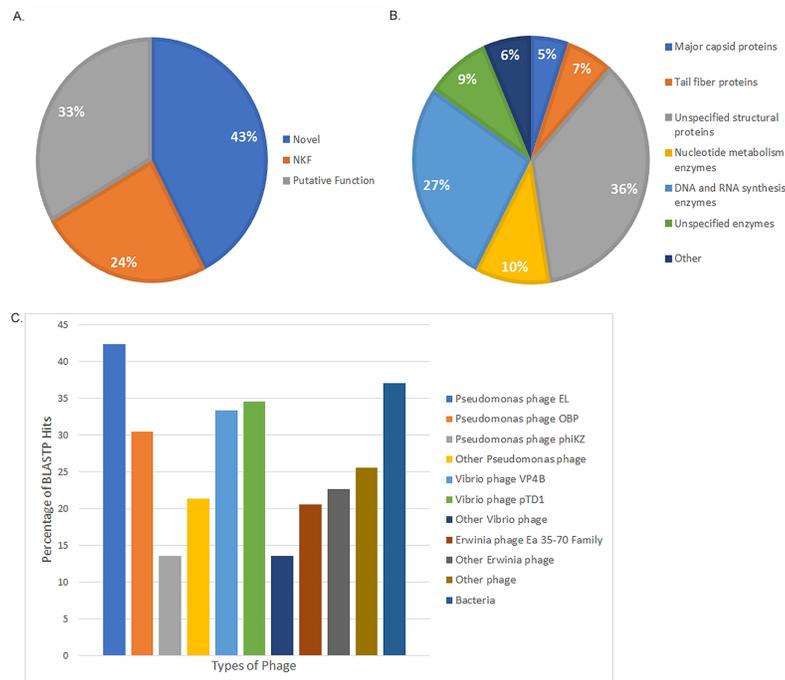


Fig 4. RisingSun and Joad are unique phages whose proteomes contain novel proteins. A) Distribution of proteins in RisingSun based upon BLASTP hits that are novel, have no known function, and putative function. B) Putative gene ontology in RisingSun. C) Percentage of RisingSun gene products with BLASTP hits to proteins in other phages/organisms.

<https://doi.org/10.1371/journal.pone.0200202.g004>

Despite the lack of strong nucleotide similarity (see Table 2), an analysis of total RisingSun gene homologs reveals 42% of RisingSun genes have homologs in *Pseudomonas* phage EL, 30% have homologs in *Pseudomonas* phage OBP, 33% in *Vibrio* phage VP4B and 35% in *Vibrio* phage pTD1 (see S1 Table for specific gene product homologs). In contrast RisingSun has only 14% of its gene products in common with phiKZ (Fig 4C). Note that these numbers are based off of annotation of gene products and could be different based on annotation. Given that other phage classifications systems have grouped related phage by 40% or greater proteome conservation, *Pseudomonas* phages EL and OBP as well as VP4B and pTD1 are distant members of a more evolutionarily diverse supercluster, with EL being the closest member to the Joad Cluster. The conserved gene products of this supercluster (totaling 63 gene products) are primarily structural genes (24 gene products) and include the MCP, portal, terminase and tail proteins. Nineteen of the remaining conserved gene products have putative functions in DNA replication and recombination, one appears to be involved in cell lysis (a phage related lysozyme). This leaves 19 of the 63 conserved gene products (~30%) that have no putative function. Due to this relationship, we tested the ability of phage Joad to infect *Pseudomonas aeruginosa* as well as several *Enterobacteriaceae* strains (Table 3). Although clear spots could be seen on *Erwinia amylovora*, *Pantoea vegans* and *Pseudomonas aeruginosa* Boston 41501 by spot test, no plaques were observed on *Pseudomonas aeruginosa* when assayed by plaque assay suggesting the spot test resulted from a toxin product in the Joad lysate. Joad did not appear to infect several other *Enterobacteriaceae* tested (including a *Pantoea agglomerans* strain, a *Vibrio cholerae* strain, an *E. coli* strain, a *Dickeya* strain, a *Salmonella* strain and an *Enterobacter* strain). As noted by the infectivity *Pantoea vegans* but not *Pantoea agglomerans* strain, this host range is no wise comprehensive since several other species or even strains within a species may be a host of Joad.

The RisingSun proteome also showed similarity to a range of other phage and bacteria. The most prominent group being a highly related *Erwinia* phage family, which consists of phages Deimos-Minion, Simmy50, SpecialG, and Ea35-70. This family matched to 21% of all RisingSun gene products, while gene products from other *Erwinia* phages corresponded to a total of 23% of gene products. Other phages infecting *Cronobacter* and *Ralstonia* bacteria among others matched 26% of the gene products, while 37% of total RisingSun gene products had a BLASTP hit to various bacteria including *Pseudomonas aeruginosa* (Fig 4C). While several of these less common matches had the lowest e-value from a particular BLAST, in a handful of

Table 3. Host range of phage Joad suggests narrow specificity towards *Erwinia amylovora* and the closely related *Pantoea vegans* bacteria. The bacteria tested is provided along with the plaque forming units from three independent assays.

Bacteria tested	Plaque forming units/mL (Pfu/mL)
<i>Erwinia amylovora</i> ATCC 29780	10 ⁸
<i>Erwinia amylovora</i> EA110	10 ⁸
<i>Pantoea vegans</i> C9-1	10 ⁸
<i>Pantoea agglomerans</i> E325	0
<i>Dickeya chrysanthemi</i> ATCC 11663	0
<i>Salmonella enterica</i> serovar typhimurium LT2	0
<i>E. coli</i> BW25113	0
<i>Enterobacter cloacae</i> ATCC 13047	0
<i>Pseudomonas aeruginosa</i> Boston 41501	0
<i>Pseudomonas chlororaphis</i> ATCC13985	0
<i>Vibrio cholerae</i> ATCC 14035	0

<https://doi.org/10.1371/journal.pone.0200202.t003>

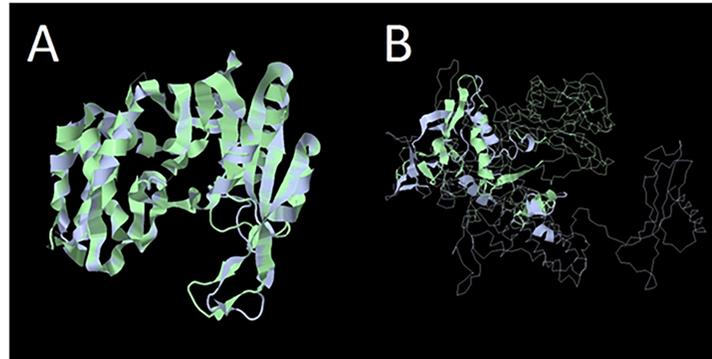


Fig 5. Predicted gp164 RisingSun functional protein modeling using RaptorX [51, 52]. A) This model reflects RisingSun gene product 164 aligned to the protein DNA adenine methylase [*Bilophila* sp. 4_1_30]: WP_009733305.1. Both proteins have similar folds, indicating that they may share the same function. B) This model reflects RisingSun gene product 85 aligned to the protein RNA polymerase beta subunit [*Erwinia* phage vB_EamM_RAY]: ANH51783.1. The protein folds diverge drastically, indicating that the two proteins may not share a similar function.

<https://doi.org/10.1371/journal.pone.0200202.g005>

cases they were for the most part orders of magnitude higher than the *Pseudomonas* and *Vibrio* phages discussed above.

Due to the distant nature of many of the BLASTP hits, we analyzed several protein putative functions by comparing their predicted folds to those of their BLASTP hits using RaptorX (Fig 5). Of the 50 RisingSun gene products with a putative function of interest, 26 shared similar folds to at least one of their respective BLASTP hits, suggesting that they share that putative function (S2 Table). Ten of these 26 proteins matched those from bacteria other than *Erwinia* while nine matched those from non-*Erwinia* phages. This leaves over half of these proteins originating from a source other than *Erwinia* or its phages. The remaining 24 proteins do not have enough sequence homology to suggest further conserved function.

Mass spectrometry

As shown in Table 4, we were able to confirm the accuracy of our genome annotation via mass spectrometry identification of protein fragments within the viral lysate of RisingSun. Mass spectrometry was able to identify nine structural proteins, three proteins with other putative functions, five novel proteins, and 14 hypothetical proteins within the phage lysate. The two proteins that had the most significant coverage and retrieval were gp68 and gp83. The gp83 product is the likely major capsid protein from BLASTP analysis, whereas gp68 has homology (E-value: $3.10e^{-07}$) to pfam12699 phiKZ-like internal head proteins. Besides phage structural proteins, three proteins with putative function were identified; a putative transglycosylase (gp231) that may aid in breaking down the *Erwinia* cell wall, a putative UvsX-like protein that is likely to function in DNA recombination (gp218), and a tubulin-like protein (gp17). Although a majority (84%) of the peptides identified by mass spectrometry belong to putative hypothetical proteins having homologs in NCBI, five completely novel proteins were identified that have never before been reported (that lack a BLASTP hit with an E-value $<10^{-7}$). These proteins demonstrate the distant nature of this phage when compared to known phages, as 43% of the proteins annotated were novel. Proteins that are under-represented in the mass spectrometry data may indicate either lower expression levels, or proteins whose expression occur in the host but are not present in the virion.

In work by Lecoutere et. al. the team performed ESI-MS/MS on *Pseudomonas* phages phiKZ and EL [53]. Upon comparing the gene products, they and we have retrieved several different gp have been in common. Of the structural proteins retrieved in RisingSun, gene

products 1, 2, 63, 66, and 68 had corresponding proteins retrieved by Lecoutere. Gene products 200 and 204 do not have matches to phage EL according to BLASTP while gp83 and gp196 do have matches but were not retrieved in the other analysis. Gene product 231 with a putative transglycosylase function was also retrieved by both labs while gp17 and 218 were retrieved in our lab. In the hypothetical protein grouping gp131 and 185 did not have matches to phage EL while gp28, 212, and 216 were only retrieved by our group. The rest of the proteins in this group namely gp5, 59, 64, 65, 73, 189, 207, 233, and 243 were obtained by both labs.

Table 4. RisingSun gene products with peptides detected by LC/MS/MS of a crude phage lysate.

RisingSun gp #	Putative Protein Function	Retrieval #*	%Cov/%Cov(95)	Peptides(95%)
Phage Structural Proteins				
gp83	major capsid protein	1	75.1/54.88	48
gp68	internal head protein	2	62.65/40.96	80
gp200	virion structural protein	12	36.36/26.22	5
gp204	virion structural protein	13	35.56/25.7	6
gp2	tail tube protein	17	61,41/18.79	4
gp63	phage capsid and scaffold	26	34.33/6.54	2
gp1	tail sheath protein	43	24.12/1.26	1
gp196	virion structural protein	57	52.94/6.23	1
gp66	internal head protein	73	27.61*	0
Other Putative Functions				
gp231	transglycosylase	44	53.06/4.76	1
gp218	UsvX recombination protein-like	84	20*	0
gp17	tubulin-like protein	62	41.43/4.36	1
Novel Hypothetical Proteins (no BLASTP hit E-value of $<10^{-7}$)				
gp163		9	64.81/37.96	7
gp206		60	29.38/9.00	1
gp71		54	47.54/3.93	1
gp150		72	11.45*	0
gp72		11	37.77/20.11	10
Hypothetical Phage Proteins				
gp28		14	34.19/11.65	5
gp5		15	51.69/13.84	6
gp189		16	40.26/14.29	5
gp59		18	50/8.59	3
gp243		20	46.08/6.14	2
gp185		25	38.59/9.00	2
gp233		27	29.56/4.56	2
gp65		32	18.5/4.62*	1
gp216		33	29.08/1.59*	1
gp73		40	37.26/3.79	1
gp207		51	18.21/5.84	1
gp131	SPFH domain containing protein	55	41.87/5.19	1
gp64		67	43.49/4.75	1
gp212		71	21.25*	0

The gene product number (RisingSun gp #), putative protein function (from BLASTP homolog or conserved domain CD), retrieval number, percent of the protein covered by peptide matches is provided (the percentage is given for both total and 95 percent confidence peptides), and number of 95% peptides is given. *The retrieval numbers missing correspond to bacterial or human (such as keratin) proteins present in the crude lysate. No reverse peptides of RisingSun were detected. The asterisks (*) indicates low confidence proteins due to total coverage of less than 30%.

<https://doi.org/10.1371/journal.pone.0200202.t004>

Motif analysis shows putative structural protein conservation. Finding conserved motifs in phage genomes may provide insight into transcriptional regulons, and therefore a deeper understanding of phage lifecycle and proteomics due to the frequent co-regulation of related genes [54]. RisingSun was analyzed for motifs using The Center for Phage Technology Galaxy Server [55], a software that allows you to search an entire genome at once. After covering the entire genome only one motif passed our e-value significance threshold of 10^{-7} . We then ran this motif through FIMO [40], a program that allows the search of an entire genome for a specific motif. We looked for occurrences of the motif represented by a logogram in (Fig 6A), which passed a q-value significance threshold of 0.01. Only 5 repetitions of this motif occurred and all in a relatively small area near the end of the genome (Fig 6B).

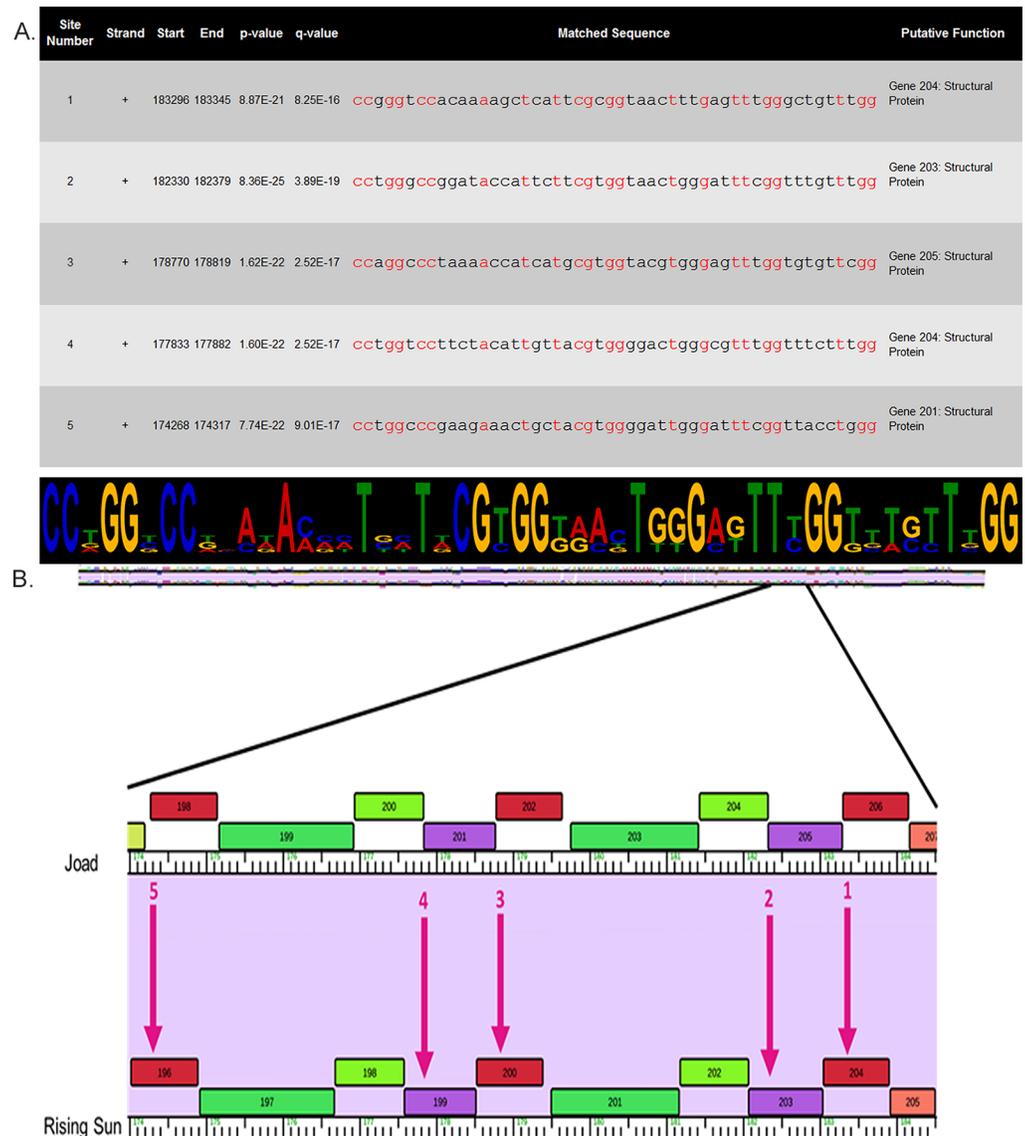


Fig 6. Locations of one motif discovered in five different structural genes. A) Exact location of motif sites, p and q-values, sequence of motifs (conserved nucleotides shown in red), and putative gene function. Logogram shows motif sequence and larger letters represent frequent conservation. B) Arrows show approximate location of motif site. Genes were analyzed from the RisingSun annotation and thus locations are shown in the RisingSun phamerator map.

<https://doi.org/10.1371/journal.pone.0200202.g006>

Identifying genes neighboring these motifs may identify co-regulated genes with related function. Fig 6. is a portion of the Joad and RisingSun genome map showing the motif locations within a span of 10 genes. Five gene products, namely gp204, gp203, gp200, gp199 and gp196 contain the motifs suggesting gene products 196–205 may be co-regulated. One of the gene products (gp205) in this region is completely novel (no BLASTP hit), and nine (gp204, gp203, gp202, gp201, gp200, gp199, gp198, gp197 and gp196) are putative virion/structural proteins based on annotation of related phage BLASTP hits. Despite the novelty of one of the genes, the evidence that this conserved motif brings suggests that genes gp196–205 are indeed structural proteins or are involved in virion assembly and that all are transcribed together. These results, however, are hypothetical and wet lab experiments are necessary for confirmation.

Conclusions

Herein we report the discovery and characterization of two newly isolated *Erwinia* phages, Joad and RisingSun, bringing the apparent total number of *Erwinia* phages on NCBI as of writing of this article to 45 reported full genome sequences. Joad and RisingSun, at over 200kbp, are two highly related Jumbo phages of the *Myoviridae* family, only distantly related to *Pseudomonas* phages EL and OBP and even more distantly related to *Vibrio* phages VP4B and pTD1 (Fig 2A). All six of these phages are most likely distant phiKZ-like phages as previously determined for EL and OBP [56, 57]. Dot Plot, ANI and BLASTP analysis all suggest the unique nature of Joad and RisingSun phages but also brings to attention that these phages have closest similarity to phages that infect bacteria outside of the *Enterobacteriales* order of bacteria. A BLASTP analysis of the putative proteome reveals BLASTP hits for ~57% of the proteins, however 33% are proteins of unknown function. The remaining putative gene products with no BLASTP hits (43%) represent proteins that have not been previously found in nature. These join the bounty of phage gene products [49, 50] with unknown structural folds and function.

Both *Pseudomonas* and *Vibrio* phages had higher similarity to Joad and RisingSun than any of the >800 *Enterobacteriaceae* phages on NCBI, despite all three of their respective hosts phylogenetically only having the class *Gammaproteobacteria* in common [58]. Of the 140 gene products with a BLASTP hit, 81% and 63% of the closest hits corresponded to gene products from *Pseudomonas* and *Vibrio* phages, respectively. Given that *Pseudomonas* bacteria are found on apple blossoms [59–64] it is reasonable to posit that extensive exposure to this strain has allowed for lateral gene transfer between phages, with possible tail fiber exchange for host recognition [65–69]. However, recent work by Adeolu and colleagues has indicated that the *Enterobacteriales* order is more nuanced than previously thought and that classification methods may need to be rethought [10]. In addition to a standard 16S rRNA based phylogenetic tree for the order *Enterobacteriales*, Adeolu and colleagues constructed three more trees based on 1548 core proteins, 53 ribosomal proteins, or four multi-locus sequence analysis proteins. After analyzing 179 species within this order they propose forming seven families, one of which is an *Erwinia-Pantoea* clade [10]. This classification is supported by our findings that these phages are quite distinct from other *Enterobacteriaceae* phages and can infect a *Pantoea vagans* strain. A reclassification of this magnitude highlights the complex evolution, and hence classification, of bacteria and the phage that infect them. Because phages are a major source of bacterial evolution, through the control of bacterial number as well as lateral gene transfer, understanding their complexity is vital. Joad and RisingSun join the 43 other *Erwinia* phages available on NCBI and provide insight into this evolutionary complexity, highlighting the similarity between *Erwinia* and the *Pseudomonads*. The relationship of these phages with more distant bacteria may reflect both an ecological niche as well as true diversity within the *Enterobacteriales*.

Supporting information

S1 Table. A list of RisingSun (MF459646) gene products that are in common with Pseudomonas phages EL (NC_007623) and OBP (NC_016571) as well as Vibrio phages pTD1 (AP017972) and VP4B(KC131130). BLASTP was used to determine the gene products in common, using the e-value cutoff of $1e-7$. Putative functions and the gene product number are provided from the annotation of phage RisingSun (from both BLASTP hits and conserved domains retrieved). The symbol “x” indicates no homolog was found. Bolded gp# indicates that the putative function was taken from the phage of that column.

(DOCX)

S2 Table. Predicted protein fold of putative RisingSun functional proteins as compared to BLASTP functional hits. Putative gene products identified for RisingSun were compared to their respective BLASTP hit with known function using RaptorX. Each alignment was given an average template modeling score (TM) between 0 and 1. According to Sheng Wang et al., if the TM is greater than 0.6, there is a 90% chance that two proteins share a similar fold. However, if the TM is less than 0.4, then the two proteins are not similar. When considering 50 genes with a putative protein function of interest, 26 proteins demonstrated similar folding to at least one of their respective BLASTP hit. Three proteins demonstrated similar folding to two BLASTP hits, for a given pair the structures were compared using RaptorX. In each of the three cases the protein structures matched each other having TM values above 0.6.

(DOCX)

Acknowledgments

The authors extend a special thanks to the Howard Hughes Medical Institute Science Education Alliance–Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) for training on phage analysis. We appreciate the help of Edward Wilcox (BYU DNA Sequencing Center), Michael Standing (BYU Microscopy Lab), and Ruchira Sharma (who aided in phageterm analysis).

Author Contributions

Conceptualization: Daniel K. Arens, Steven M. Johnson, Julianne H. Grose.

Data curation: Daniel K. Arens.

Formal analysis: Daniel K. Arens.

Funding acquisition: Julianne H. Grose.

Investigation: Daniel K. Arens, T. Scott Brady, John L. Carter, Jenny A. Pape, David M. Robinson, Kerri A. Russell, Lyndsay A. Staley, Jason M. Stettler, Olivia B. Tateoka, Michelle H. Townsend, Kiara V. Whitley, Trevor M. Wienclaw, Taryn L. Williamson.

Methodology: Daniel K. Arens.

Project administration: Julianne H. Grose.

Supervision: Julianne H. Grose.

Validation: Daniel K. Arens.

Visualization: Jenny A. Pape, Lyndsay A. Staley.

Writing – original draft: Daniel K. Arens.

Writing – review & editing: Daniel K. Arens, Steven M. Johnson, Julianne H. Grose.

References

1. Duckworth DH. WHO DISCOVERED BACTERIOPHAGE. *Bacteriological Reviews*. 1976; 40(4):793–802. PubMed PMID: WOS:A1976CT29300001. PMID: [795414](#)
2. Grose JH, Casjens SR. Understanding the enormous diversity of bacteriophages: The tailed phages that infect the bacterial family Enterobacteriaceae. *Virology*. 2014; 468:421–43. <https://doi.org/10.1016/j.virol.2014.08.024> PubMed PMID: WOS:000344434400046. PMID: [25240328](#)
3. Cheetham BF, Katz ME. A ROLE FOR BACTERIOPHAGES IN THE EVOLUTION AND TRANSFER OF BACTERIAL VIRULENCE DETERMINANTS. *Mol Microbiol*. 1995; 18(2):201–8. https://doi.org/10.1111/j.1365-2958.1995.mmi_18020201.x PubMed PMID: WOS:A1995TG43300003. PMID: [8709840](#)
4. Schicklmaier P, Schmieger H. FREQUENCY OF GENERALIZED TRANSDUCING PHAGES IN NATURAL ISOLATES OF THE SALMONELLA-TYPHIMURIUM COMPLEX. *Appl Environ Microbiol*. 1995; 61(4):1637–40. PubMed PMID: WOS:A1995QQ36200074. PMID: [7747978](#)
5. Faruque SM, Mekalanos JJ. Phage-bacterial interactions in the evolution of toxigenic *Vibrio cholerae*. *Virulence*. 2012; 3(7):556–65. <https://doi.org/10.4161/viru.22351> PubMed PMID: WOS:000313285700003. PMID: [23076327](#)
6. Trost E, Blom J, Soares SD, Huang IH, Al-Dilaimi A, Schroder J, et al. Pangenomic Study of *Corynebacterium diphtheriae* That Provides Insights into the Genomic Diversity of Pathogenic Isolates from Cases of Classical Diphtheria, Endocarditis, and Pneumonia. *J Bacteriol*. 2012; 194(12):3199–215. <https://doi.org/10.1128/JB.00183-12> PubMed PMID: WOS:000304978400018. PMID: [22505676](#)
7. Rasko DA, Rosovitz MJ, Myers GSA, Mongodin EF, Fricke WF, Gajer P, et al. The pangenome structure of *Escherichia coli*: Comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J Bacteriol*. 2008; 190(20):6881–93. <https://doi.org/10.1128/JB.00619-08> PubMed PMID: WOS:000259631900036. PMID: [18676672](#)
8. Doss J, Culbertson K, Hahn D, Camacho J, Barekzi N. A Review of Phage Therapy against Bacterial Pathogens of Aquatic and Terrestrial Organisms. *Viruses-Basel*. 2017; 9(3):10. <https://doi.org/10.3390/v9030050> PubMed PMID: WOS:000397261500013. PMID: [28335451](#)
9. D'Herelle F. On an invisible microbe antagonistic dysenteric bacilli: brief note by Mr. F. D'Herelle presented by Mr. Roux. *Res Microbiol*. 2007; 158(7):553–4. <https://doi.org/10.1016/j.resmic.2007.07.005> PubMed PMID: WOS:000250885900002. PMID: [17855060](#)
10. Adeolu M, Alnajjar S, Naushad S, Gupta RS. Genome-based phylogeny and taxonomy of the 'Enterobacteriales': proposal for Enterobacteriales ord. nov divided into the families Enterobacteriaceae, Erwiniaceae fam. nov., Pectobacteriaceae fam. nov., Yersiniaceae fam. nov., Hafniaceae fam. nov., Morganellaceae fam. nov., and Budviciaceae fam. nov. *Int J Syst Evol Microbiol*. 2016; 66:5575–99. <https://doi.org/10.1099/ijsem.0.001485> PubMed PMID: WOS:000393357900103. PMID: [27620848](#)
11. Prevention CfDCa. Antibiotic/Antimicrobial Resistance Biggest Threats 2013 [updated April 14th, 2017; cited 2017 May 26th]. Available from: https://www.cdc.gov/drugresistance/biggest_threats.html.
12. Norelli JL, Jones AL, Aldwinckle HS. Fire blight management in the twenty-first century—Using new technologies that enhance host resistance in apple. *Plant Dis*. 2003; 87(7):756–65. <https://doi.org/10.1094/pdis.2003.87.7.756> PubMed PMID: WOS:000183693400001.
13. Vrancken K, Holtappels M, Schoofs H, Deckers T, Valcke R. Pathogenicity and infection strategies of the fire blight pathogen *Erwinia amylovora* in Rosaceae: State of the art. *Microbiology-(UK)*. 2013; 159:823–32. <https://doi.org/10.1099/mic.0.064881-0> PubMed PMID: WOS:000319850600001.
14. Shapiro LR, Scully ED, Straub TJ, Park J, Stephenson AG, Beattie GA, et al. Horizontal Gene Acquisitions, Mobile Element Proliferation, and Genome Decay in the Host-Restricted Plant Pathogen *Erwinia Tracheiphila*. *Genome Biol Evol*. 2016; 8(3):649–64. <https://doi.org/10.1093/gbe/evw016> PubMed PMID: WOS:000373839200016. PMID: [26992913](#)
15. Esplin IND, Berg JA, Sharma R, Allen RC, Arens DK, Ashcroft CR, et al. Genome Sequences of 19 Novel *Erwinia amylovora* Bacteriophages. *Genome announcements*. 2017; 5(46). Epub 2017/11/18. <https://doi.org/10.1128/genomeA.00931-17> PMID: [29146842](#); PubMed Central PMCID: PMC5690319.
16. Ritchie DF, Klos EJ. ISOLATION OF ERWINIA AMYLOVORA BACTERIOPHAGE FROM AERIAL PARTS OF APPLE-TREES. *Phytopathology*. 1977; 67(1):101–4. PubMed PMID: WOS:A1977CU27300018.
17. Pusey PL. Crab apple blossoms as a model for research on biological control of fire blight. *Phytopathology*. 1997; 87(11):1096–102. <https://doi.org/10.1094/PHYTO.1997.87.11.1096> PubMed PMID: WOS:A1997YD86700002. PMID: [18945005](#)

18. Ishimaru CA, Klos EJ, Brubaker RR. MULTIPLE ANTIBIOTIC PRODUCTION BY ERWINIA-HERBICOLA. *Phytopathology*. 1988; 78(6):746–50. <https://doi.org/10.1094/Phyto-78-746> PubMed PMID: WOS:A1988P382900026.
19. Burkholder WH, McFadden LA, Dimock AW. A BACTERIAL BLIGHT OF CHRYSANTHEMUMS. *Phytopathology*. 1953; 43(9):522–6. PubMed PMID: WOS:A1953UM96100182.
20. Medeiros AA, O'Brien TF, Wacker WE, Yulug NF. Effect of salt concentration on the apparent in-vitro susceptibility of *Pseudomonas* and other gram-negative bacilli to gentamicin. *The Journal of infectious diseases*. 1971;124 Suppl:S59–64. Epub 1971/12/01. PMID: 5001630.
21. Haynes WC, Stodola FH, Locke JM, Pridham TG, Conway HF, Sohns VE, et al. PSEUDOMONAS-AUREOFACIENS KLUYVER AND PHENAZINE ALPHA-CARBOXYLIC ACID, ITS CHARACTERISTIC PIGMENT. *J Bacteriol*. 1956; 72(3):412–7. PubMed PMID: WOS:A1956WD31800024. PMID: 13366938
22. Sneath PHA, Cowan ST. AN ELECTRO-TAXONOMIC SURVEY OF BACTERIA. *J Gen Microbiol*. 1958; 19(3):551–65. <https://doi.org/10.1099/00221287-19-3-551> PubMed PMID: WOS:A1958WM15500014. PMID: 13611198
23. Skerman VBD, McGowan V, Sneath PHA. APPROVED LISTS OF BACTERIAL NAMES. *Int J Syst Bacteriol*. 1980; 30(1):225–420. <https://doi.org/10.1099/00207713-30-1-225> PubMed PMID: WOS:A1980JG46200030.
24. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A*. 2000; 97(12):6640–5. <https://doi.org/10.1073/pnas.120163297> PubMed PMID: WOS:000087526300074. PMID: 10829079
25. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods*. 2012; 9(7):671–5. <https://doi.org/10.1038/nmeth.2089> PubMed PMID: WOS:000305942200020. PMID: 22930834
26. Guttman M, Betts GN, Barnes H, Ghassemian M, van der Geer P, Komives EA. Interactions of the NPXY microdomains of the low density lipoprotein receptor-related protein 1. *Proteomics*. 2009; 9(22):5016–28. <https://doi.org/10.1002/pmic.200900457> PubMed PMID: WOS:000272391300002. PMID: 19771558
27. McCormack AL, Schieltz DM, Goode B, Yang S, Barnes G, Drubin D, et al. Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level. *Anal Chem*. 1997; 69(4):767–76. <https://doi.org/10.1021/ac960799q> PubMed PMID: WOS:A1997WH35700036. PMID: 9043199
28. Krumsiek J, Arnold R, Rattei T. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics*. 2007; 23(8):1026–8. <https://doi.org/10.1093/bioinformatics/btm039> PubMed PMID: WOS:000246293000015. PMID: 17309896
29. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. BASIC LOCAL ALIGNMENT SEARCH TOOL. *J Mol Biol*. 1990; 215(3):403–10. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) PubMed PMID: WOS:A1990ED16700008. PMID: 2231712
30. Gish W, States DJ. IDENTIFICATION OF PROTEIN CODING REGIONS BY DATABASE SIMILARITY SEARCH. *Nature Genet*. 1993; 3(3):266–72. <https://doi.org/10.1038/ng0393-266> PubMed PMID: WOS:A1993KP79000019. PMID: 8485583
31. Madden TL, Tatusov RL, Zhang JH. Applications of network BLAST server. *Methods Enzymol*. 1996; 266:131–41. PubMed PMID: WOS:A1996BF77C00009. PMID: 8743682
32. Hatfull GF, Jacobs-Sera D, Lawrence JG, Pope WH, Russell DA, Ko CC, et al. Comparative Genomic Analysis of 60 Mycobacteriophage Genomes: Genome Clustering, Gene Acquisition, and Gene Size. *J Mol Biol*. 2010; 397(1):119–43. <https://doi.org/10.1016/j.jmb.2010.01.011> PubMed PMID: WOS:000275785600009. PMID: 20064525
33. Grose JH, Jensen GL, Burnett SH, Breakwell DP. Genomic comparison of 93 *Bacillus* phages reveals 12 clusters, 14 singletons and remarkable diversity (vol 15, 855, 2014). *BMC Genomics*. 2014; 15:20. <https://doi.org/10.1186/1471-2164-15-20> PubMed PMID: WOS:000349050500001.
34. Goujon M, McWilliam H, Li WZ, Valentin F, Squizzato S, Paern J, et al. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Research*. 2010; 38:W695–W9. <https://doi.org/10.1093/nar/gkq313> PubMed PMID: WOS:000284148900113. PMID: 20439314
35. Lassmann T, Frings O, Sonnhammer ELL. Kalign2: high-performance multiple alignment of protein and nucleotide sequences allowing external features. *Nucleic Acids Research*. 2009; 37(3):858–65. <https://doi.org/10.1093/nar/gkn1006> PubMed PMID: WOS:000263831400029. PMID: 19103665
36. Lassmann T, Sonnhammer ELL. Kalign—an accurate and fast multiple sequence alignment algorithm. *BMC Bioinformatics*. 2005; 6:9. <https://doi.org/10.1186/1471-2105-6-9> PubMed PMID: WOS:000234486000001.

37. Li WZ, Cowley A, Uludag M, Gur T, McWilliam H, Squizzato S, et al. The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Research*. 2015; 43(W1):W580–W4. <https://doi.org/10.1093/nar/gkv279> PubMed PMID: WOS:000359772700093. PMID: 25845596
38. McWilliam H, Li WZ, Uludag M, Squizzato S, Park YM, Buso N, et al. Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Research*. 2013; 41(W1):W597–W600. <https://doi.org/10.1093/nar/gkt376> PubMed PMID: WOS:000323603200095. PMID: 23671338
39. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*. 2009; 37:W202–W8. <https://doi.org/10.1093/nar/gkp335> PubMed PMID: WOS:000267889100037. PMID: 19458158
40. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. *Bioinformatics*. 2011; 27(7):1017–8. <https://doi.org/10.1093/bioinformatics/btr064> PubMed PMID: WOS:000289162000022. PMID: 21330290
41. Lawrence J. DNA Master. 2012.
42. Cresawn SG, Bogel M, Day N, Jacobs-Sera D, Hendrix RW, Hatfull GF. Phamerator: a bioinformatic tool for comparative bacteriophage genomics. *BMC Bioinformatics*. 2011; 12:14. <https://doi.org/10.1186/1471-2105-12-14> PubMed PMID: WOS:000297784500001.
43. Yuan YH, Gao MY. Jumbo Bacteriophages: An Overview. *Front Microbiol*. 2017; 8:9. <https://doi.org/10.3389/fmicb.2017.00009> PubMed PMID: WOS:000396093900002.
44. Lowe TM, Eddy SR. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*. 1997; 25(5):955–64. <https://doi.org/10.1093/nar/25.5.955> PubMed PMID: WOS:A1997WM3000005. PMID: 9023104
45. Casjens SR, Grose JH. Contributions of P2- and P22-like prophages to understanding the enormous diversity and abundance of tailed bacteriophages. *Virology*. 2016; 496:255–76. <https://doi.org/10.1016/j.virol.2016.05.022> PubMed PMID: WOS:000381235700029. PMID: 27372181
46. Kuebler D, Rao VB. Functional analysis of the DNA-packaging/terminase protein gp17 from bacteriophage T4. *J Mol Biol*. 1998; 281(5):803–14. <https://doi.org/10.1006/jmbi.1998.1952> PubMed PMID: WOS:000075771200005. PMID: 9719636
47. Merrill BD, Ward AT, Grose JH, Hope S. Software-based analysis of bacteriophage genomes, physical ends, and packaging strategies. *BMC Genomics*. 2016; 17:16. <https://doi.org/10.1186/s12864-015-2264-z> PubMed PMID: WOS:000384980300001.
48. Garneau JRD, F.; Fortier L.; Bikard D.; Monot M. PhageTerm: a Fast and User-friendly Software to Determine Bacteriophage Termini and Packaging Mode using randomly fragmented NGS data. *bioRxiv*. 2017. <https://doi.org/10.1101/108100>.
49. Hatfull GF. Dark Matter of the Biosphere: the Amazing World of Bacteriophage Diversity. *J Virol*. 2015; 89(16):8107–10. <https://doi.org/10.1128/JVI.01340-15> PubMed PMID: WOS:000358278200002. PMID: 26018169
50. Mokili JL, Rohwer F, Dutilh BE. Metagenomics and future perspectives in virus discovery. *Curr Opin Virol*. 2012; 2(1):63–77. <https://doi.org/10.1016/j.coviro.2011.12.004> PubMed PMID: WOS:000312112600010. PMID: 22440968
51. Wang S, Ma JZ, Peng J, Xu JB. Protein structure alignment beyond spatial proximity. *Sci Rep*. 2013; 3:7. <https://doi.org/10.1038/srep01448> PubMed PMID: WOS:000316102200002. PMID: 23486213
52. Wang S, Peng J, Xu JB. Alignment of distantly related protein structures: algorithm, bound and implications to homology modeling. *Bioinformatics*. 2011; 27(18):2537–45. <https://doi.org/10.1093/bioinformatics/btr432> PubMed PMID: WOS:000294755400009. PMID: 21791532
53. Lecoutere E, Ceyssens PJ, Miroshnikov KA, Mesyanzhinov VV, Krylov VN, Noben JP, et al. Identification and comparative analysis of the structural proteomes of phi KZ and EL, two giant *Pseudomonas aeruginosa* bacteriophages. *Proteomics*. 2009; 9(11):3215–9. <https://doi.org/10.1002/pmic.200800727> PubMed PMID: WOS:000267414500025. PMID: 19526553
54. Berdygulova Z, Westblade LF, Florens L, Koonin EV, Chait BT, Ramanculov E, et al. Temporal Regulation of Gene Expression of the *Thermus thermophilus* Bacteriophage P23-45. *J Mol Biol*. 2011; 405(1):125–42. <https://doi.org/10.1016/j.jmb.2010.10.049> PubMed PMID: WOS:000286700800013. PMID: 21050864
55. Blankenberg D. In preparation.
56. Sokolova OS, Shaburova OV, Pechnikova EV, Shaytan AK, Krylov SV, Kiselev NA, et al. Genome packaging in EL and Lin68, two giant phiKZ-like bacteriophages of *P. aeruginosa*. *Virology*. 2014; 468:472–8. <https://doi.org/10.1016/j.virol.2014.09.002> PubMed PMID: WOS:000344434400050. PMID: 25254945
57. Cornelissen A, Hardies SC, Shaburova OV, Krylov VN, Mattheus W, Kropinski AM, et al. Complete Genome Sequence of the Giant Virus OBP and Comparative Genome Analysis of the Diverse phi KZ-

- Related Phages. *J Virol.* 2012; 86(3):1844–52. <https://doi.org/10.1128/JVI.06330-11> PubMed PMID: WOS:000299308000050. PMID: 22130535
58. Euzéby JP. List of bacterial names with standing in nomenclature: A folder available on the Internet. *Int J Syst Bacteriol.* 1997; 47(2):590–2. PubMed PMID: WOS:A1997WT56800057. <https://doi.org/10.1099/00207713-47-2-590> PMID: 9103655
 59. Pusey PL, Stockwell VO, Mazzola M. Epiphytic bacteria and yeasts on apple blossoms and their potential as antagonists of *Erwinia amylovora*. *Phytopathology.* 2009; 99(6):S105–S. PubMed PMID: WOS:000266213300626.
 60. Bedford KE, Macneill BH, Bonn WG, Dirks VA. POPULATION-DYNAMICS OF PSEUDOMONAS-SYRINGAE PV PAPULANS ON MUTSU APPLE. *Can J Plant Pathol-Rev Can Phytopathol.* 1988; 10(1):23–9. PubMed PMID: WOS:A1988P663200005.
 61. Kennelly MM, Cazorla FM, de Vicente A, Ramos C, Sundin GW. *Pseudomonas syringae* diseases of fruit trees—Progress toward understanding and control. *Plant Dis.* 2007; 91(1):4–17. <https://doi.org/10.1094/pd-91-0004> PubMed PMID: WOS:000243011800001.
 62. Mansvelt EL, Hattingh MJ. PSEUDOMONAS-SYRINGAE PV SYRINGAE ASSOCIATED WITH APPLE AND PEAR BUDS IN SOUTH-AFRICA. *Plant Dis.* 1987; 71(9):789–92. <https://doi.org/10.1094/pd-71-0789> PubMed PMID: WOS:A1987J809000005.
 63. Kotan R, Sahin F, Ala A. Identification and pathogenicity of bacteria isolated from pome fruit trees in the Eastern Anatolia region of Turkey. *J Plant Dis Prot.* 2006; 113(1):8–13. PubMed PMID: WOS:000238127200003.
 64. Miliute I, Buzaitė O, Gelvonauskienė D, Sasnauskas A, Stanys V, Baniulis D. Plant growth promoting and antagonistic properties of endophytic bacteria isolated from domestic apple. *Zemdirbyste.* 2016; 103(1):77–82. <https://doi.org/10.13080/z-a.2016.103.010> PubMed PMID: WOS:000375815600010.
 65. Haggardjungquist E, Halling C, Calendar R. DNA-SEQUENCES OF THE TAIL FIBER GENES OF BACTERIOPHAGE P2—EVIDENCE FOR HORIZONTAL TRANSFER OF TAIL FIBER GENES AMONG UNRELATED BACTERIOPHAGES. *J Bacteriol.* 1992; 174(5):1462–77. PubMed PMID: WOS:A1992HF27900005. PMID: 1531648
 66. Sandmeier H, Iida S, Arber W. DNA INVERSION REGIONS MIN OF PLASMID P15B AND CIN OF BACTERIOPHAGE-P1—EVOLUTION OF BACTERIOPHAGE TAIL FIBER GENES. *J Bacteriol.* 1992; 174(12):3936–44. PubMed PMID: WOS:A1992HY09300013. PMID: 1534556
 67. Pickard D, Thomson NR, Baker S, Wain J, Pardo M, Goulding D, et al. Molecular characterization of the *Salmonella enterica* serovar Typhi Vi-typing bacteriophage E1. *J Bacteriol.* 2008; 190(7):2580–7. <https://doi.org/10.1128/JB.01654-07> PubMed PMID: WOS:000254323300037. PMID: 18192390
 68. Casjens SR, Thuman-Commike PA. Evolution of mosaically related tailed bacteriophage genomes seen through the lens of phage P22 virion assembly. *Virology.* 2011; 411(2):393–415. <https://doi.org/10.1016/j.virol.2010.12.046> PubMed PMID: WOS:000288343000023. PMID: 21310457
 69. Jacobs-Sera D, Marinelli LJ, Bowman C, Broussard GW, Bustamante CG, Boyle MM, et al. On the nature of mycobacteriophage diversity and host preference. *Virology.* 2012; 434(2):187–201. <https://doi.org/10.1016/j.virol.2012.09.026> PubMed PMID: WOS:000312509300008. PMID: 23084079