



1-2020

## UP $\Phi$ phages, a new group of filamentous phages found in several members of Enterobacteriales

Jason Shapiro

Loyola University Chicago, [jshapiro2@luc.edu](mailto:jshapiro2@luc.edu)

Catherine Putonti

Loyola University Chicago, [cputonti@luc.edu](mailto:cputonti@luc.edu)

Follow this and additional works at: [https://ecommons.luc.edu/bioinformatics\\_facpub](https://ecommons.luc.edu/bioinformatics_facpub)

---

### Recommended Citation

Shapiro, Jason and Putonti, Catherine. UP $\Phi$  phages, a new group of filamentous phages found in several members of Enterobacteriales. *Virus Evolution*, 6, 1: , 2020. Retrieved from Loyola eCommons, Bioinformatics Faculty Publications, <http://dx.doi.org/10.1093/ve/veaa030>

This Article is brought to you for free and open access by the Faculty Publications and Other Works by Department at Loyola eCommons. It has been accepted for inclusion in Bioinformatics Faculty Publications by an authorized administrator of Loyola eCommons. For more information, please contact [ecommons@luc.edu](mailto:ecommons@luc.edu).



This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

© The Authors, 2020.

# UP $\phi$ phages, a new group of filamentous phages found in several members of *Enterobacteriales*

Jason W. Shapiro<sup>1,\*†</sup> and Catherine Putonti<sup>1,2,3,4</sup>

<sup>1</sup>Department of Biology, Loyola University Chicago, 1032 W Sheridan Rd, Chicago, IL 60660, USA, <sup>2</sup>Department of Computer Science, Loyola University Chicago, 1052 W Loyola Ave, Chicago, IL, 60626, USA, <sup>3</sup>Bioinformatics Program, Loyola University Chicago, 1052 W Loyola Ave, Chicago, IL 60626, USA and <sup>4</sup>Department of Microbiology and Immunology, Stritch School of Medicine, Loyola University Chicago, 2160 S First Ave, Maywood, IL 60153, USA

\*Corresponding author: E-mail: jshapiro2@luc.edu

†<https://orcid.org/0000-0002-1321-8418>

## Abstract

Filamentous phages establish chronic infections in their bacterial hosts, and new phages are secreted by infected bacteria for multiple generations, typically without causing host death. Often, these viruses integrate in their host's genome by co-opting the host's XerCD recombinase system. In several cases, these viruses also encode genes that increase bacterial virulence in plants and animals. Here, we describe a new filamentous phage, UP $\phi$ 901, which we originally found integrated in a clinical isolate of *Escherichia coli* from urine. UP $\phi$ 901 and closely related phages can be found in published genomes of over 200 other bacteria, including strains of *Citrobacter koseri*, *Salmonella enterica*, *Yersinia enterocolitica*, and *Klebsiella pneumoniae*. Its closest relatives are consistently found in urine or in the blood and feces of patients with urinary tract infections. More distant relatives can be found in isolates from other environments, including sewage, water, soil, and contaminated food. Each of these phages, which we collectively call 'UP $\phi$  viruses', also harbors two or more novel genes of unknown function.

**Key words:** bacteriophage; inovirus; prophage; bladder.

## 1. Introduction

Phages are often described by their potential to kill their hosts. Obligately lytic phages kill their hosts following infection, whereas temperate phages may lie dormant as prophages within lysogenized bacteria for several generations before entering a lytic cycle. While killing by phages has immediate consequences for bacterial ecology and has led to the revival of phage therapy for treating bacterial infections (Kortright et al. 2019), many phages also carry genes that alter bacterial behavior (Mai-Prochnow et al. 2015; Warwick-Dugdale et al. 2019). The effects of these phage-encoded genes range from modifying photosynthesis in cyanobacteria (Sieradzki et al. 2019) to producing toxins in potential pathogens (Waldor and Mekalanos 1996).

Filamentous phages in the family *Inoviridae* go beyond the standard dichotomy of viral lysis and lysogeny. Instead, the majority of characterized inoviruses maintain productive infections over multiple bacterial generations, without killing their hosts. In many cases, these phages integrate as tandem repeats (e.g. Derbise et al. 2007) into their hosts' genomes at a locus called the *dif* site (Mai-Prochnow et al. 2015). The *dif* site is a 28bp region at the bacterial terminus that includes two short palindromic regions recognized separately by the site-specific recombinase pair XerC and XerD. In bacteria with this system, XerCD is responsible for resolving chromosome dimers during DNA replication and cell division (Carnoy and Roten 2009). Filamentous phages co-opt this system by carrying their own copy of the *dif* site within their genomes, causing XerCD to

confuse phage DNA for bacterial DNA. After integration, new phages are produced at relatively low rates and may be maintained indefinitely within the bacterial population (Val et al. 2005). Other filamentous phages that lack a *dif* site can also integrate elsewhere in the bacterial genome by encoding their own integrases, as in the Pf bacteriophages of *Pseudomonas aeruginosa* (Mai-Prochnow et al. 2015).

Filamentous phages can have dramatic effects on their hosts' phenotypes. In *Vibrio cholerae*, the phage CTX $\phi$  encodes the toxin genes responsible for bacterial virulence in humans (Waldor and Mekalanos 1996). Similarly, filamentous phages have been associated with the virulence of other human pathogens including hemorrhagic *Escherichia coli*, *Yersinia pestis* (the cause of plague), and *P. aeruginosa* in both patients with cystic fibrosis and skin wounds (Gonzalez et al. 2002; Derbise et al. 2007; Sweere et al. 2019). Inoviruses have also been correlated with the increased virulence of agricultural pests like tomato wilt and potato blight (Kamiunten and Wakimoto 1982; Yamada 2013). Across these various infections, filamentous phages may carry toxin genes, alter bacterial motility (Jian, Xiao, and Wang 2013) and biofilm formation (Rice et al. 2009; May, Tsuruta, and Okabe 2011), or indirectly interfere with the immune system's ability to clear the bacterial infection (Sweere et al. 2019). Despite their myriad effects on a wide range of clinically and agriculturally important hosts, these phages remain understudied. As of this writing, there are only forty-six recognized members of *Inoviridae* in RefSeq, and only thirty-three with established taxonomy by the International Committee on Taxonomy of Viruses (ICTV). Nonetheless, recent work has shown that thousands of potential inoviruses can be found as prophages in published bacterial genomes and metagenomes (Roux et al. 2019).

Here, we describe a novel filamentous phage, UP $\phi$ 901, discovered as a prophage in a clinical *E. coli* isolate from a patient urine sample. UP $\phi$ 901 is most closely related to the non-integrating phages I2-2 and IKe in the *Lineavirus* genus of the *Inoviridae*. Like many filamentous phages, UP $\phi$ 901 integrates as a tandem repeat at the *dif* site in its hosts. Using the phage sequence from our isolate as a query, we searched for close relatives of UP $\phi$ 901 in published bacterial genomes and assemblies. We found homologous prophages in over 200 strains of *E. coli*, *Citrobacter koseri*, *Klebsiella pneumoniae*, *Salmonella enterica*, and *Yersinia enterocolitica*. The most similar phages tended to come from patient urine samples or from the blood or feces of patients with urinary tract infection (UTI); more distant relatives could be found in soil, water, animal feces, and contaminated food. We refer to the collection of related filamentous phages as 'UP $\phi$  viruses'. In several cases, identical UP $\phi$  strains were found infecting multiple bacterial genera, suggesting recent host switching events. Last, we describe the prevalence of putative UP $\phi$  viruses in metagenomes.

## 2. Methods

### 2.1 Bacterial strains

Clinical isolates of *E. coli* and *C. koseri* (see Table 1) were provided by the Wolfe Lab at Loyola University Chicago and were originally isolated as part of separate IRB-approved studies on the urinary microbiota of women with and without symptoms of UTIs or other urinary ailments (Price et al. 2016a,b; Garretto 2019; Penckofer et al. 2020). These strains and their GenBank accessions (where available) are summarized in Table 1. *Citrobacter koseri* strains in this study have not been fully

**Table 1.** Bacterial strains in this study.

Strain ID	Species	Has UP $\phi$ ?	WGS accession
UMB0731	<i>E. coli</i>	No	NZ_RRWP000000000 <sup>a</sup>
UMB0901	<i>E. coli</i>	Yes	NZ_PKHH000000000
UMB1220	<i>E. coli</i>	Yes	NZ_RRVZ000000000
UMB1526	<i>E. coli</i>	Yes	NZ_RRVH000000000
UMB5814	<i>E. coli</i>	Yes	NZ_RRUX000000000
UMB6655	<i>E. coli</i>	No	NZ_RRUL000000000
UMB6890	<i>E. coli</i>	No	NZ_RRUI000000000
UMB1389	<i>C. koseri</i>	No	Unsequenced <sup>b</sup>
UMB7451	<i>C. koseri</i>	Yes	Unsequenced
UMB8248	<i>C. koseri</i>	No	Unsequenced
JE-1	<i>E. coli</i>	No	Unavailable

<sup>a</sup>UMB0901 was sequenced as 'E75' in its original BioProject (PRJNA316969).

<sup>b</sup>*Citrobacter* strains not sequenced at time of writing. Phage identified by PCR.

sequenced. *Escherichia coli* JE-1 is an IncI plasmid-bearing strain typically used to propagate phage I2-2 and was obtained from the Felix d'Herelle Reference Center for Bacterial Viruses (Université Laval, QC, Canada).

### 2.2 Identifying phage in bacterial genomes

UP $\phi$ 901 was originally discovered as part of predicting phage sequences using PHASTER (Arndt et al. 2016) in UMB0901 and other *E. coli* isolates in a previous study (Miller-Ensminger et al. 2018; Garretto 2019). (The 'UP' in its name originated uncreatively from 'urine project', though not all related viruses are necessarily found in urine, as described in Section 3.) To identify UP $\phi$ 901 in other published genomes, we took advantage of its integration at its hosts' *dif* sites as a tandem repeat. We defined a preliminary BLAST (Altschul et al. 1990) query sequence for UP $\phi$ 901 as the region starting with the *dif* locus in UMB0901 and extending to the first phage repeat (7,560 bp). Throughout this paper, we will refer to the specific sequence (and homologs that are over 99% identical by amino acid sequence) as 'UP $\phi$ 901'. Remaining phages that share a common set of core genes (identified below) and with less than 99 per cent amino acid identity are collectively referred to as 'UP $\phi$  viruses'. The *dif* site itself is not repeated in full within the tandem duplications. We then used this single copy of the phage as a query for a blastn search using the NCBI web tool (<https://blast.ncbi.nlm.nih.gov>) with default parameters. This BLAST search identified prophages with over 99 per cent query coverage and over 98 per cent nucleotide identity to UP $\phi$ 901 in strains of *C. koseri*, *E. coli*, *K. pneumoniae*, *S. enterica*, and *Y. enterocolitica*. Notably, neither PHASTER nor VirSorter (Roux et al. 2015) consistently predicted UP $\phi$ 901 homologs in bacteria that had significant BLAST hits. This high false-negative rate for these two tools is likely due to UP $\phi$ 901's short genome and the reliance of these tools on information about known viruses. We have not tested the newer tool Inovirus\_detector (Roux et al. 2019) with genomes carrying UP $\phi$ 901 homologs.

We next downloaded all available draft and complete assemblies on NCBI for the five bacterial species listed above (as of August 2018). Each assembly was queried locally with tblastx with default parameters to identify contigs containing putative phage regions. We retained any contigs containing a hit covering at least 1,000 nucleotides of the UP $\phi$ 901 query with over 75 per cent nucleotide identity. In no case did we observe a putative UP $\phi$ 901 homolog on two contigs in the same assembly. The start positions of each phage were then confirmed by a separate

local blastn search for the *dif* site (as identified in Carnoy and Roten 2009), and stop positions were identified by confirming the first repeat in the BLAST results. These repeats typically matched just the first fifteen bases in the *dif* site, and we accepted putative hits that covered at least fourteen of the twenty-eight *dif* site nucleotides with at least 85 per cent identity. Due to varying assembly quality, not all assemblies included phage repeats, and some sequences did not include a *dif* repeat as an obvious stopping point. In these instances, we could only rely on the end position of the last significant hit in the initial tblastx query.

To confirm the sensitivity of our UP $\phi$ 901 detection, we then blasted each of the 229 putative UP $\phi$ 901-infected genomes against a query for the related phage, Ypf $\phi$  (from *Y. pestis* CO92, GenBank: CP009973), which has been found at *dif* sites in other studies (Derbise et al. 2007). We liberally accepted all putative hits, regardless of identity, and found a single case (GCA\_001519645) with a cumulative query coverage over 1,000 bp. We queried both UP $\phi$ 901 and Ypf $\phi$  in SRA BLAST searches against the original reads for this assembly (SRX1528813).

### 2.3 Phylogenetic analyses

We used Anvi'o (Eren et al. 2015) to facilitate gene clustering and annotation for the putative phage regions found in GenBank assemblies. Standard annotation tools fail to identify inovirus genes accurately, so we then used blastp through the NCBI web tool with individual phage IKe (GenBank: NC\_002014) and I2-2 (GenBank: NC\_001332) genes to confirm each gene's correct annotation. These were identified by visually comparing the aligned genes rather than relying on a significance threshold, as homologous sequences may share less than 30 per cent amino acid identity in these phages. Full-length UP $\phi$ 901 and most of the putative phage regions have every gene found in IKe and I2-2 (as in Fig. 2).

We identified nine UP $\phi$ 901 core genes to use for phylogenetic inference (genes I, II, III, IV, V, VI, VIII, h1, and h2). Genes VII and IX were excluded from analyses, because they consist of only about thirty amino acids and were not consistently identified as genes. Gene X was also excluded as it is contained entirely within gene II. Next, we used MAFFT (Katoh and Standley 2013) to align each of the core genes. In reviewing individual gene alignments, we identified cases in genes I, II, III, and IV where the original gene calls by Anvi'o split the open reading frame (ORF) into two pieces because of early stop codons. It is unknown if these genes are truly pseudogenized or if they might still encode functional proteins. For the purpose of building the trees, we concatenated the two halves of these genes, since their complete sequences reflect evolutionary relationships, even if they might not be translated in full. Both the original and concatenated versions of these sequences are provided in the data repository for this work.

We then built a phylogeny of the 229 complete prophages with IQ-TREE version 1.6.12 (Nguyen et al. 2015) using the concatenated alignment of each core UP $\phi$ 901 gene amino acid sequence with 1,000 bootstraps. We took advantage of IQ-TREE's integrated ModelFinder (Kalyaanamoorthy et al. 2017) option to perform model selection. This step identified VT+F+R2 (variable time with empirical state frequencies and two rate categories) as the best substitution model according to Bayesian Information Criterion (BIC). Individual gene trees shown in Supplementary Fig. S1 were built using FastTree (Price, Dehal, and Arkin 2010).

We also generated a bacterial phylogeny for the phage hosts, using Anvi'o to identify a set of 400 single-copy core genes

shared among each host that also contained a prophage in the virus phylogeny. These genes were aligned within Anvi'o using MUSCLE (Edgar 2004) and concatenated. We used IQ-TREE as above to build the final phylogeny. For the host tree, ModelFinder identified JTT+F+R10 (Jones-Taylor-Thornton (Jones, Taylor, and Thornton 1992) with empirical state frequencies and ten rate categories) as the best substitution model by BIC.

Trees were visualized using iTOL (Letunic and Bork 2019) and the ape package (Paradis and Schliep 2019) in R (R Core Team 2013). Additional metadata for tree visualization were obtained from NCBI, EBI, or from the literature. Key metadata included the type of material sampled for isolating bacteria (e.g. soil, blood, urine, feces) and the more general source of that material (e.g. environmental, animal, human, food). The full metadata associated with each sample is provided in the data repository (see Data availability). Nine strains had no available metadata from a publication or database. Twenty-two other strains without available metadata were from the '100K Pathogen Genomes Project' (BioProject PRJNA186441; Weimer 2017) and are identified as '100K Project' in figures.

### 2.4 Phage in SRA metagenomes

We used searchsra (Levi et al. 2018; www.searchsra.org) to check for UP $\phi$ 901 relatives in metagenomes. We used a truncated 6,703 bp UP $\phi$ 901 reference that included only the core UP $\phi$  virus genes used in phylogenetic analyses. The results from the search contained over 100,000 potential hits, most of which are false positives returned from a small number of erroneous reads aligning to the reference genome. Following the searchsra authors' GitHub repository (https://github.com/linsalrob/SearchSRA), we first identified the depth of coverage across our reference genome for each potential hit. We used the pileup.sh function from BMAP (Bushnell 2014) to perform this step. We filtered results by both the query coverage and smoothness of coverage (Aziz et al. 2015) to identify true positives. A detailed description of these filtering steps with examples of true and false-positive coverage plots is provided in the Supplementary Methods. Metadata for the likeliest true-positive results are available in Supplementary File S1 and as part of the data repository. Statistical analysis related to these metadata was done in R.

### 2.5 Phage presence in culture medium

UP $\phi$ 901-infected strains were grown in lysogeny broth (LB) at 37°C with moderate shaking overnight. The following morning, 1 ml of each culture was removed, centrifuged at 16,000 g for 1 min, and the supernatants were filtered through 0.2- $\mu$ m cellulose acetate syringe filters. An 80  $\mu$ l sample of each filtrate was then treated with OPTIZYME DNase I (Fisher BioReagents BP81071) for 30 min, followed by heat inactivation with EDTA at 65°C for 10 min. This DNase step was included to remove any bacterial genomic DNA from the supernatant, which could result from cell death due to lysis from other prophages or shearing forces. DNase-treated samples were incubated at 95°C for 10 min to denature phage protein coats and expose the phage ssDNA. We then amplified phage genomic DNA using UP $\phi$ 901-specific primers, UPphi\_shortFW (GGGTTTATCAGAGGGGTCAG) and UPphi\_shortRV (AGGATGGCTCTAAGTCAACG). 16S PCR (63F/1387R primers) was used to confirm the absence of bacterial genomic DNA in the DNase-treated filtrates. The UP $\phi$ 901 and 16S PCRs were performed on 1  $\mu$ l of unfiltered culture as

positive controls. Positive UP $\phi$  PCR products were verified by Sanger sequencing using the UPphi\_shortFW primer.

## 2.6 Phage infection assays

We tested the potential for UP $\phi$ 901 to infect new hosts using standard plaque assays by spotting 5  $\mu$ l of filtrate from UMB0901 (as described above) on 0.7 per cent agar LB overlays containing candidate bacterial hosts (UMB0731, UMB6655, UMB6890, JE-1).

We also tested for new infections using PCR. Colonies of prospective host strains were added to 5 ml of LB supplemented with 50  $\mu$ l of filtered supernatant from UMB0901. Uninfected controls of each strain were also grown. These cultures were incubated for 18 h. Following growth, 2  $\mu$ l of each culture were used in a PCR designed to test for integration into the new host's chromosome. The forward primer (IntCheck\_FW: GTGTGTGGATGTGAATGGTG) in the assay was based on a conserved sequence in all tested hosts just upstream of the *dif* site, and the reverse primer (IntCheck\_RV: CTGGCAGAACG AACGATTAC) recognized a region early in the phage genome. This PCR can only amplify DNA if UP $\phi$ 901 integrates into a new genome. An overnight culture of UMB0901 was used as a positive control for each reaction, and the uninfected cultures were used as negative controls.

## 3. Results

### 3.1 Initial identification

We originally identified UP $\phi$ 901 in a clinical isolate of *E. coli* (UMB0901) as part of separate work examining prophages present in the urinary microbiome (Garretto 2019). We then found prophages with over 99 per cent nucleotide identity in three other patient isolates in our lab collection (UMB1220, UMB1526, and UMB5814). In each case, we confirmed by PCR that the phage is shed from infected bacteria, as centrifuged culture supernatant treated with DNase is positive for UP $\phi$ 901 DNA (Fig. 1A) but negative for bacterial genomic DNA (Fig. 1B). Sanger sequencing confirmed that each PCR product was a true match to the UP $\phi$ 901 genome.

### 3.2 Comparison to characterized inoviruses

UP $\phi$ 901 is approximately 50 per cent identical to phages IKe and I2-2 (70% coverage and ~70% amino acid identity for covered genes), each a non-integrating member of the *Lineavirus* genus of filamentous phages. These phages are closely related to the

F-specific filamentous phage, M13. We also compared UP $\phi$ 901 to Ypf $\phi$ , a *dif*-integrating filamentous phage that infects pathogenic strains of *E. coli* and *Y. pestis*. Overall, Ypf $\phi$  shares less than 30 per cent amino acid identity with UP $\phi$ 901.

Figure 2 shows how UP $\phi$ 901's genome compares with these other inoviruses, with each gene shaded according to its amino acid similarity with the homolog in UP $\phi$ 901. UP $\phi$ 901 contains each of the 'core' genes characteristic of filamentous coliphages with gene order preserved. UP $\phi$ 901 and Ypf $\phi$  are less than 20 per cent identical in all genes except for genes II and V, which are each over 85 per cent identical between these phages. Genes II and V are both involved in regulating phage DNA replication, and this conserved module is less than 15 per cent similar in IKe and M13.

Two other differences set UP $\phi$ 901 apart from the non-integrating phages, I2-2, IKe, and M13: 1, UP $\phi$ 901 includes a copy of the *E. coli dif* site at the start of its genome, enabling it to integrate via the XerCD recombinase system; 2, UP $\phi$ 901 encodes two putative genes of unknown function within a region that is intergenic in I2-2, IKe, and M13. These genes are not homologous to any known protein, and all attempts to find annotated homologs of these genes in GenBank returned hits to hypothetical proteins in prophages related to UP $\phi$ 901. In addition to these hypothetical proteins, the UP $\phi$ 901 prophage region also contains three ORFs in reverse orientation just downstream of the *dif* site. This genome organization is mirrored by Ypf $\phi$  and other integrating filamentous phages, including CTX $\phi$ , where accessory genes often surround the core functions (Derbise et al. 2007; Mai-Prochnow et al. 2015).

### 3.3 Comparative genomics

UP $\phi$ 901 strains share nearly 100 per cent amino acid identity across infected clinical isolates of *E. coli* in our lab collection. The only distinction is that the attachment protein, g3p (encoded by gene III), in UP $\phi$ 901 has an additional repeat of a glycine-rich motif ('GGGES') than the viruses found in UMB1220, UMB1526, and UMB5814. As with many other filamentous phages, UP $\phi$ 901 is integrated as a tandem repeat at the *dif* site in each of these genomes. In the case of UMB0901, we reassessed our own assembly by aligning the raw reads to the contig with UP $\phi$ 901 and estimating the depth of coverage (Supplementary Fig. S2). On average, the UP $\phi$ 901 prophage region has three times the coverage of the surrounding loci, indicating a tandem triple in UMB0901.

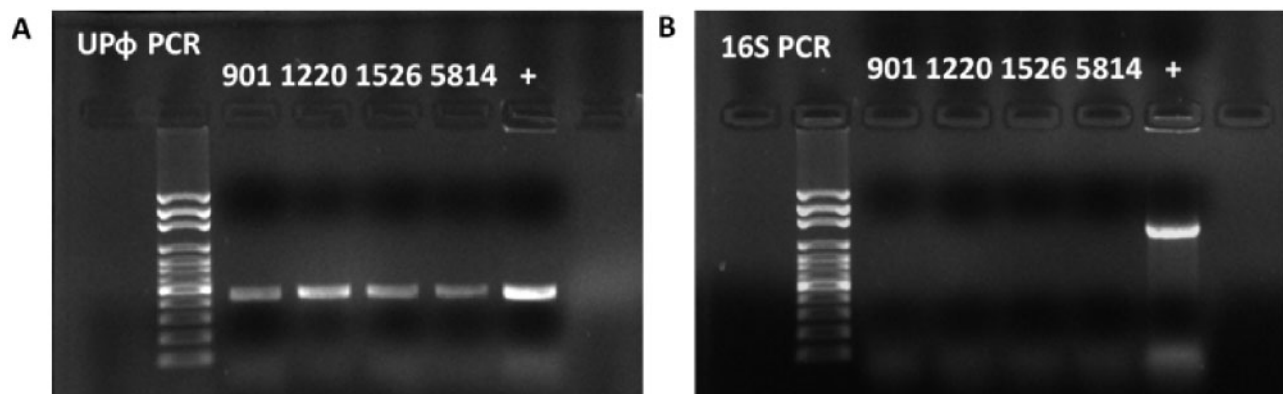
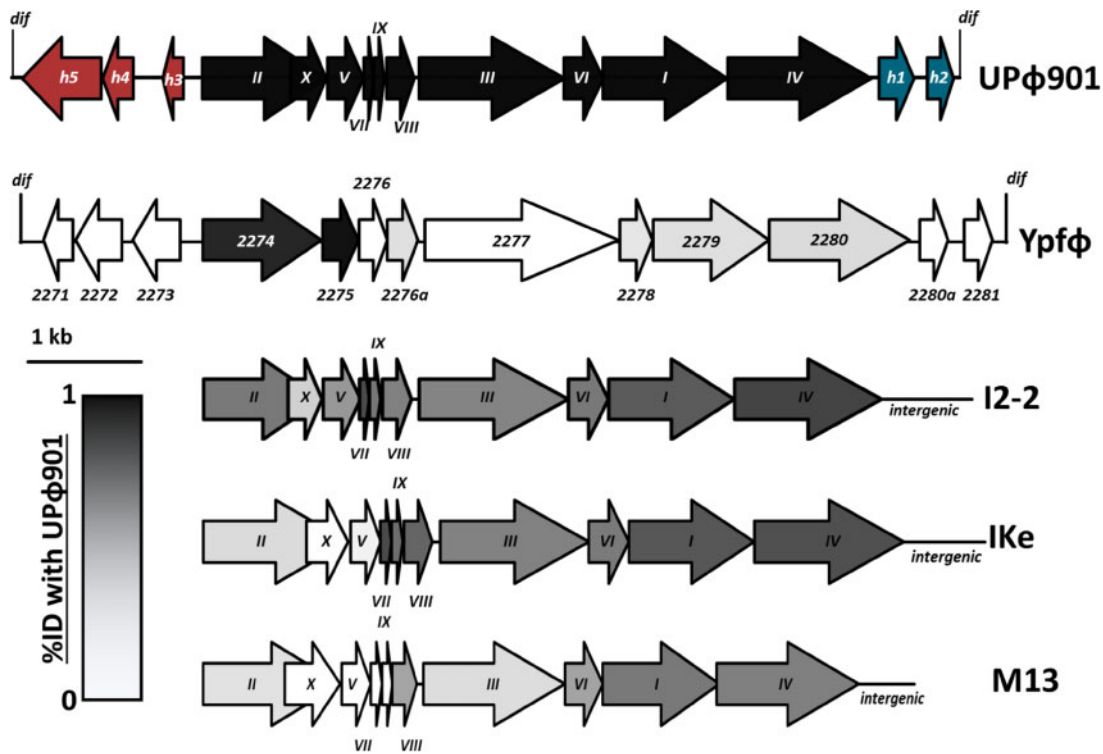


Figure 1. Phage DNA confirmation. PCR results using (A) UP $\phi$  or (B) 16S primers after DNase treatment to remove bacterial genomic DNA from filtrates of UMB0901, UMB1220, UMB1526, and UMB5814 cultures. The + control lane was from UMB0901 colony PCR.



**Figure 2.** Comparison of UP $\phi$ 901 to related inoviruses. Roman numerals are the names for core inovirus genes common in M13 and the *Lineavirus* members. Ypf $\phi$  genes labeled as in [Derbise et al. \(2007\)](#). Genes shaded gray are darker if they are more similar to the sequence from the UP $\phi$ 901 prophage in UMB0901.

**Table 2.** Prevalence of UP $\phi$  in GenBank assemblies.

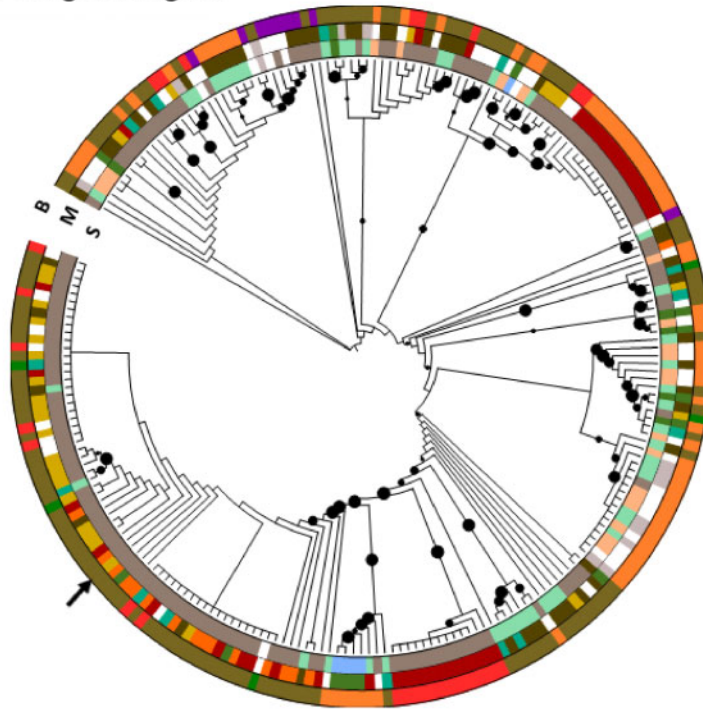
Host species	GenBank assemblies	Assemblies with UP $\phi$ hit
<i>E. coli</i>	12,401	189
<i>S. enterica</i>	9,144	95
<i>K. pneumoniae</i>	4,647	29
<i>Y. enterocolitica</i>	177	10
<i>C. koseri</i>	28	8

Using the UP $\phi$ 901 sequence as a BLAST query, we searched GenBank for homologous prophages in other bacteria. Prophages with over 98 per cent nucleotide identity across the UP $\phi$ 901 query sequence could be found in published genomes of *E. coli*, *S. enterica*, *C. koseri*, *K. pneumoniae*, and *Y. enterocolitica*. We then downloaded all available assemblies from GenBank of these five species (26,397 in total) and used tblastx to identify viruses integrated at *dif* sites. In all, 331 genomes harbored a potential UP $\phi$ 901 prophage (see [Table 2](#) for a summary by host). Of these 331 putative prophages, 229 contained each gene found in UP $\phi$ 901, excluding the three ORFs neighboring the *dif* site. The third of these ORFs (labeled h3 in [Fig. 2](#)) is present in 221 of these 229 putative prophages, whereas h4 and h5 are found in only 56 and 57 genomes, respectively. For the remainder of this paper, we will refer to this collection of 229 strains as ‘UP $\phi$  viruses’. The 102 remaining genomes contained a significant hit to UP $\phi$ 901, but assembly or sequencing quality was inadequate to predict full phage genomes reliably. We also checked each of the 229 UP $\phi$ -infected genomes for Ypf $\phi$  to rule out possible false-positive results. We found one instance (*E. coli* assembly GCA\_001519645) of coinfection between a *dif*-integrated UP $\phi$  prophage and a putative Ypf $\phi$  prophage integrated at a different locus. None of the remaining 228 genomes harbored a prophage resembling Ypf $\phi$ .

We then built a phylogeny for the 229 complete UP $\phi$  prophages. Overall, the phylogeny reflects low genetic variation across the UP $\phi$  viruses, and most clades are separated by short branches and include polytomies. Even the more distant phage relatives found in *Y. enterocolitica* are 98 per cent identical in amino acid sequence to the original UP $\phi$ 901 core genes. As a result, the tree has multiple splits with bootstrap support below 50 per cent. For this reason, we show the tree as a cladogram in [Fig. 3A](#) and will not attempt to over-interpret the tree topology. (A high-resolution phylogeny with complete bootstrap supports is shown in [Supplementary Fig. S3](#).) For each of the genomes in the phylogeny, we identified metadata, where available, for the sample source material (e.g. urine, feces, blood) and source environment (e.g. human, animal, environmental) and added these data to the tree visualization.

UP $\phi$ 901 is indicated by an arrow in [Fig. 3A](#) and is part of a clade of sixty-five phages. Fifty-five of these prophages come from *E. coli*, of which thirty-six came from urine (including fourteen UTI samples). Two large polytomies make up the majority of this group, and at least one sample (assembly GCA\_000457405) was from the blood of a patient with UTI-induced bacteremia. In addition, each polytomy includes cases where identical phages (i.e. 100% amino acid identity across the genes used in the phylogeny) were also found within *C. koseri* and *K. pneumoniae*. This group contains three samples from animals. The rest of the tree includes seventy prophages found in *S. enterica* (from both animals and humans), though phages infecting each of the other host species are also present. None of these *S. enterica* samples were isolated from urine, and only ten of the remaining ninety-four sequences came from urine or UTI samples. The total counts for each host species, sample material, and source are indicated in the legend for [Fig. 3](#).

## A Phage cladogram



## Bacteria

<span style="color: brown;">■</span>	<i>Escherichia coli</i> (118)
<span style="color: purple;">■</span>	<i>Yersinia enterocolitica</i> (8)
<span style="color: orange;">■</span>	<i>Salmonella enterica</i> (70)
<span style="color: red;">■</span>	<i>Klebsiella pneumoniae</i> (28)
<span style="color: green;">■</span>	<i>Citrobacter koseri</i> (5)

## Material

<span style="color: brown;">■</span>	feces (48)
<span style="color: yellow;">■</span>	urine (30)
<span style="color: orange;">■</span>	UTI (16)
<span style="color: red;">■</span>	blood (42)
<span style="color: teal;">■</span>	other body (14)
<span style="color: grey;">■</span>	food (12)
<span style="color: green;">■</span>	environment (14)
<span style="color: white;">■</span>	missing (53)

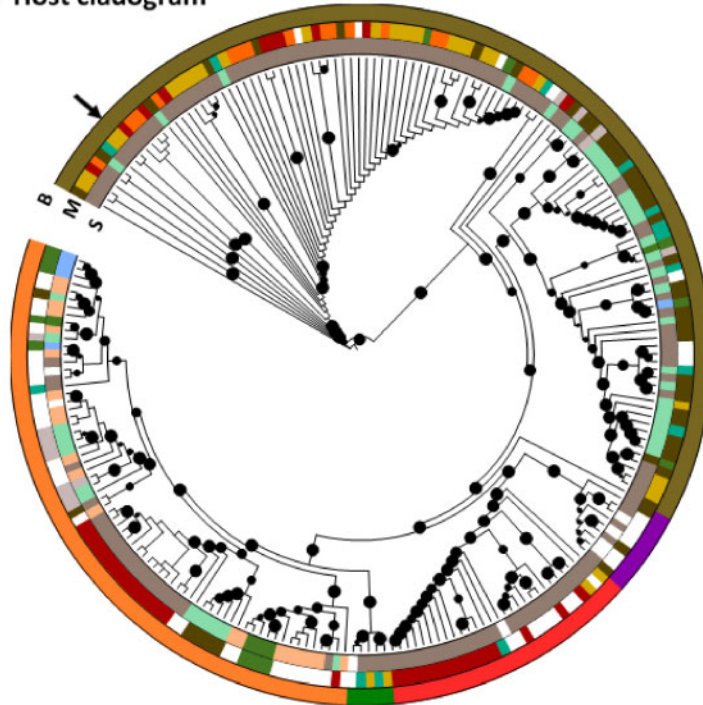
## Source

<span style="color: brown;">■</span>	sewage (1)
<span style="color: grey;">■</span>	human (141)
<span style="color: lightgreen;">■</span>	animal (45)
<span style="color: orange;">■</span>	100K Project (21)
<span style="color: blue;">■</span>	water (5)
<span style="color: grey;">■</span>	food (2)
<span style="color: green;">■</span>	environment (6)
<span style="color: white;">■</span>	missing (8)

## Bootstrap Support (%)

- 51
- 100

## B Host cladogram



**Figure 3.** Cladograms of UP $\phi$  prophages in GenBank assemblies (A) and their hosts (B). The outer ring is the host species, the middle ring is the sample material, and the inner ring is the source environment. Black arrows indicate (A) UP $\phi$ 901 and (B) UMB0901. Samples sizes shown in parentheses in the legend. Trees are based on amino acid sequences of core genes. Bootstrap support is shown by the size of black dots with at least 51 per cent support. High-resolution trees with scale bars and bootstrap supports are in [Supplementary Fig. S3](#).

We next constructed a phylogeny for the host bacteria, using a set of conserved single-copy genes identified with Anvi'o ([Eren et al. 2015](#)). The bacterial tree shows the expected

breakdown by host species, with additional structure corresponding to sample material and whether the strain was from human or animal sources. Branch supports are generally higher

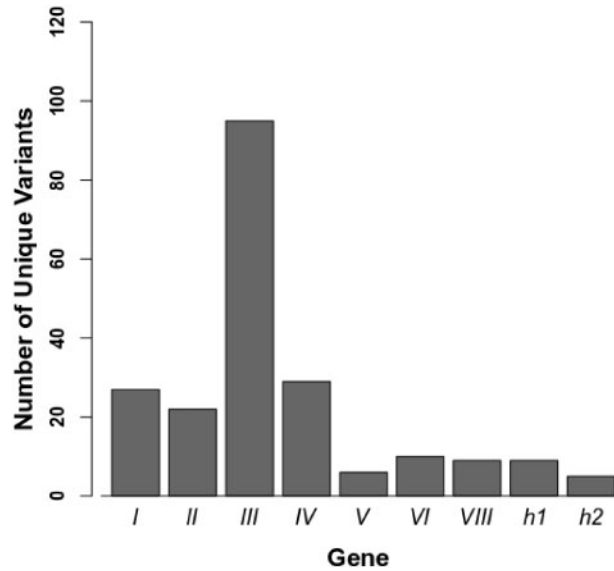


Figure 4. Number of unique amino acid sequence variants for each UP $\phi$  gene used to build trees.

in the bacterial tree except for clades with very short branches. Due to the low branch support of the phage tree, it is difficult to compare the phage and host tree topologies.

As expected from the short branches of the phage tree, individual phage gene trees (Supplementary Fig. S1) show little variation in amino acid sequence across the phages, with most genes composed of a small number of unique variants (Fig. 4). Gene III stands out as having the most variation, with ninety-five distinct amino acid sequences among the 229 genomes. Examining the alignment for gene III, most of the variation is in the number of glycine-rich repeats in the region that also varied in our lab isolates.

### 3.4 Phages in metagenomes

We next used searchsra (Levi et al. 2018) to find relatives of UP $\phi$ 901 in metagenomes. The initial results included over 100,000 sequencing runs with at least one read aligned to UP $\phi$ 901. After a series of filtering steps to remove erroneous data and false-positive results (see Supplementary Methods) we identified 257 SRA runs that were most likely to have true-positive results. We then used SRA BLAST (Camacho et al. 2009) to validate these putative hits and identified any instances of redundant BioSample numbers (multiple sequencing runs from the same sample) or redundant sample source (e.g. multiple samples from the same individual). This validation step identified 78 true positive, non-redundant samples that covered over 75 per cent of the UP $\phi$ 901 reference genome with over 90 per cent sequence identity. SRA BLAST is sensitive enough to distinguish between UP $\phi$  and Ypf $\phi$  prophages (Supplementary Fig. S4).

These positive metagenomic results include sixty-four from gut, five from urine, and four from environmental samples. Two additional hits came from reference isolates of *C. koseri* and from mock communities used in validating microbiome projects. (Full metadata are provided in Supplementary File S1.) Five of the positive gut samples were from non-human mammals, including three from cats and two from pandas. Environmental samples included two from the New York subway, one from wastewater, and one from river sediment. Most notably,

twenty-four of the sixty-four unique gut samples came from nine different studies of the infant microbiome and included children from China, Estonia, Finland, Singapore, and the USA. One hypothesis for this high number of infant samples carrying UP $\phi$ -infected bacteria is that UP $\phi$  may be more common in hospital-acquired Proteobacteria.

Several filamentous phages have been associated with pathogen virulence (Mai-Prochnow et al. 2015), including one study identifying their potential involvement in cases of neonatal meningitis caused by *E. coli* O18:K1:H7 (Gonzalez et al. 2002). We were interested to see if any of the infant microbiome studies identified through searchsra might shed light on potential connections between UP $\phi$  viruses and bacterial virulence.

In one study (Ward et al. 2016), the authors sequenced the gut microbiomes of 144 pre-term and twenty-two full-term infants and identified cases where necrotizing enterocolitis (NEC) was associated with strains of uropathogenic *E. coli*. Furthermore, the anonymized patient metadata were available with the publication and could be matched to sample metadata from the SRA. This patient data included indicator variables for high (over one or ten) per cent abundance of *E. coli* and *K. pneumoniae*, as well as information on patient outcomes. We identified nine patients in this study that carried bacteria infected by a putative UP $\phi$  prophage (patient IDs: 11211, 11961, 11962, 12321, 21461, 30031, 30081, 30251, 30252). Of these, six contained *E. coli* and five contained *K. pneumoniae*; two patients had both bacteria present. The overall rate of UP $\phi$  presence (9/166) is about four times higher than across GenBank assemblies of *E. coli* and *K. pneumoniae* (218/17048). This difference represents a significant enrichment ( $\chi^2 = 170.44$ ,  $P < 0.001$ ) in this study. Moreover, two of the nine patients were ultimately diagnosed with NEC. This is not significantly different ( $\chi^2 = 0.0011$ ,  $P = 0.973$ ) from the rate of NEC without UP $\phi$  in the study (25/157). In addition, four of the patients (IDs starting with '30') were full-term and not treated in the NICU. The rate of UP $\phi$  prophages among the full-term infants (4/22) is significantly greater than among the pre-term patients (5/144) ( $\chi^2 = 5.44$ ,  $P = 0.020$ ). Thus, while it is possible that UP $\phi$  is more common in hospital-acquired bacteria, there is no evidence to suggest that it is associated with patient outcomes in this study.

### 3.5 Identifying additional infections in the lab

After identifying UP $\phi$  viruses in other hosts, we returned to the lab to test additional strains. Though only twenty-eight *C. koseri* genomes were available in GenBank, nearly one-third carried a UP $\phi$  virus. In two cases, these *C. koseri* phages were identical to prophage sequences found in urine *E. coli*s. Given this high frequency of infection and similarity to UP $\phi$ 901, we tested three unsequenced *C. koseri* isolates (UMB1389, UMB7451, UMB8248) using UP $\phi$ 901 PCR primers. The PCR results identified an integrated phage in UMB7451, but it was not actively produced, in contrast to the infected *E. coli* isolates (Supplementary Fig. S5).

We also attempted to establish new infections using filtrate from UMB0901 cultures. Filamentous phages typically rely on conjugative pili as a primary receptor for infection (Mai-Prochnow et al. 2015), and each *E. coli* carrying UP $\phi$ 901 in our lab collection also harbored an IncI conjugative plasmid. We identified three uninfected *E. coli* strains (UMB0731, UMB6655, and UMB6890) as candidates for testing UP $\phi$ 901's ability to establish new infections. These strains all clustered together with UP $\phi$ 901-infected strains (UMB1220, UMB1526, UMB5814) in a previous gene presence-absence analysis of strains in our



collection of urinary *E. coli* (Garretto 2019), contained an IncI plasmid, and were not already infected by UP $\phi$ 901.

Despite their potential as candidate hosts, UMB0901 filtrates that were confirmed to contain the phage did not produce plaques on any of these strains. We were also unable to observe plaques on *E. coli* JE-1, an IncI-bearing strain that is the standard host for UP $\phi$ 901's closest known relative, I2-2 (Bradley, Coetzee, and Hedges 1983). Last, we used PCR to test for phage integration in these candidate hosts but were again unable to confirm new infections.

#### 4. Discussion

We have introduced a new filamentous phage, UP $\phi$ 901, found in multiple environments, with the largest clade found predominantly in urine *E. coli*. In addition to infecting *E. coli*, related viruses can be found in *S. enterica*, *K. pneumoniae*, *C. koseri*, and *Y. enterocolitica*. In several cases, identical phage genomes were found in multiple host genera. These phages all share two putative genes of unknown function. We have tentatively named this collection of phages carrying these genes as 'UP $\phi$  viruses'. Three additional novel ORFs were found in UP $\phi$ 901 and fifty-five other UP $\phi$  viruses, and most UP $\phi$  prophages lacking these genes harbored other ORFs in these positions.

These UP $\phi$  viruses are related to phages IKE and I2-2 in the *Lineavirus* genus of *Inoviridae*, but IKE and I2-2 cannot integrate into the host genome and lack the accessory genes that appear to be unique to UP $\phi$  phages (Fig. 2). Given the prevalence of *dif* site integration among filamentous phages and their high sequence similarity to much of the UP $\phi$ 901 genome, it is likely that IKE and I2-2 evolved from an integrating ancestor.

Recent work has called into question the current taxonomy of filamentous phages and has suggested that the *Inoviridae* may require substantial revision into multiple new families of viruses (Roux et al. 2019). We are, therefore, hesitant to claim that UP $\phi$  viruses deserve to be identified as a new phage genus. For the time being, we propose that the UP $\phi$  viruses should be considered members of the existing *Lineavirus* genus, with the current members (IKE and I2-2) forming a subgenus of phages that have lost the *dif* site and nearby accessory genes. We acknowledge, though, that this taxonomy could change as the *Inoviridae* are revised and may be complicated by recombination between co-infecting phages.

##### 4.1 Relationship to CUS $\phi$ and Ypf $\phi$

Prior work identified the filamentous prophages CUS $\phi$  and Ypf $\phi$  in strains of extra-intestinal *E. coli* (ExPEC) and *Y. pestis* (Gonzalez et al. 2002; Derbise et al. 2007). While these phages have variable accessory gene regions (the same regions where UP $\phi$  has its own novel genes), CUS $\phi$  and Ypf $\phi$  otherwise share 99 per cent nucleotide identity with each other (Derbise et al. 2007). These phages are found in similar hosts as UP $\phi$  viruses and rely on *dif* site integration. In these prior studies, the authors demonstrated both the potential for these phages to increase the virulence of ExPEC (Gonzalez et al. 2002), and also the likely association between Ypf $\phi$  and the modern strains of *Y. pestis* pv *orientalis* associated with the third plague pandemic (Derbise et al. 2007).

As part of the exploration of UP $\phi$ , we compared the individual genes from our UP $\phi$ 901 reference sequence to the Ypf $\phi$  prophage region in its type host, *Y. pestis* CO92. We found that only genes II and V have significant homology (Fig. 2), whereas the rest of UP $\phi$ 901 is more closely related to phages I2-2 and IKE.

Genes II and V are both involved in regulating phage DNA replication, and this modular split in homology between the disparate viruses likely reflects a deeper history of recombination among the phages.

Because they infect related hosts and might have a history of recombination, we also checked for any instances of co-infection between UP $\phi$  and Ypf $\phi$  in our data. We found one such case (GCA\_001519645), corresponding to *E. coli* bacteremia in a patient with sepsis. This assembly, unfortunately, consisted of 206 contigs, and Ypf $\phi$  appears to be split across at least four different contigs. Notably, Ypf $\phi$  is not included on a large contig containing both *dif* and UP $\phi$ , suggesting this Ypf $\phi$  prophage integrated at a different locus using another mechanism or is present as a plasmid in this host. SRA BLAST (Supplementary Fig. S6) confirmed that each phage is fully covered by the reads. It is unknown if the prophage genomes interact. Sachs and Bull (2005) demonstrated that co-infecting filamentous phages (in their case, f1 and IKE) could coevolve to package their genomes together using shared coat proteins. It is feasible that UP $\phi$  and Ypf $\phi$  could also produce chimaeric virions. In fact, the UP $\phi$  region in GCA\_001519645 has an early stop codon in its gene III and would likely be unviable unless it is able to incorporate Ypf $\phi$ 's version of the attachment protein during assembly.

##### 4.2 Tandem repeats and prophage diversity

*dif* site integration typically results in tandem duplication of the phage genome (Mai-Prochnow et al. 2015). It has been observed in CTX $\phi$  and Ypf $\phi$  that hosts harboring a *dif*-integrated phage can be super-infected by other such viruses, with the new phage potentially supplanting the original (Midonet et al. 2019) or integrating in the same region (Derbise and Carniel 2014). In the latter cases, the phage can integrate between the original copies or downstream (Chouikha et al. 2010). It is, therefore, possible to observe tandem triples or even quadruples of inovirus prophages at the *dif* site of infected bacteria, and these downstream homologs are not always identical. Furthermore, this series of integration events can promote hybridization between co-infecting inoviruses when new phages are produced by infected hosts (Davis and Waldor 2000; Derbise and Carniel 2014), as well as co-packaging of phages into shared coats as described above (Sachs and Bull 2005). In the case of UP $\phi$  prophages, we did not always know the exact number of tandem repeats within each host genome, as most sequencing efforts relied on only short Illumina reads that were not resolved into tandem copies during assembly by the original data providers. For UMB0901, we used depth of coverage to infer that UP $\phi$ 901 is integrated as a tandem triple.

Given the propensity for tandem prophage recombination, there is additional uncertainty in interpreting and comparing the genomes of UP $\phi$  viruses that presents an additional challenge for taxonomy. The *E. coli* strain ECONIH2 (GenBank: CP014667) was originally sequenced with both Illumina and PacBio technologies and offers a valuable case study. ECONIH2 contains a tandem triple of UP $\phi$  prophages. The first two copies of the phage are identical by nucleotide sequence and contain each of the genes also found in UP $\phi$ 901, as well as three additional ORFs. The third copy of the phage in ECONIH2, however, contains only the first of these additional genes followed by a fourth novel ORF. This one bacterial strain demonstrates the potential heterogeneity among co-infecting prophages.

### 4.3 Future work

Additional work remains to understand the prevalence and role of UP $\phi$  viruses in different microbial communities. First, future research will need to characterize the two novel genes, *h1* and *h2*. These genes are found in a region of the genome that often includes genes that alter host behavior or virulence (Mai-Prochnow et al. 2015). In the prototypical integrating inovirus, CTX $\phi$ , this region encodes the cholera toxin genes and their regulators (Waldor and Mekalanos 1996). It appears likely that these two new genes interact with one another, but they share no homology to any known gene, and there is little to hint at their possible functions.

While we have identified putative phages from genome assemblies, it was not always possible to predict phage end positions or the number of tandem repeats. These issues make it difficult to assess the pan-genome of UP $\phi$  viruses, as some might carry additional accessory genes not found in UP $\phi$ 901. Future research that includes long-read resequencing of infected hosts and phages released into culture medium would help to resolve these ambiguities.

We also attempted to determine the host requirements for establishing new infections of UP $\phi$ 901. We were able to confirm by PCR that new virions are released into the culture medium by infected bacteria, but we were unsuccessful in our attempts to infect new hosts. Our failure to observe new infections may be the result of lacking appropriate host strains for testing or due to low efficiency of establishing new infections. At the same time, UP $\phi$ 901's closest relatives, I2-2 and IKE, each infect *E. coli* harboring IncI conjugative plasmids, and each *E. coli* strain infected by UP $\phi$ 901 in our collection also contained an IncI plasmid. It, therefore, appears likely that at least some of these phages rely on IncI-encoded pili, and future work with additional strains and different growth conditions will test this possibility.

It is also possible that different UP $\phi$  viruses rely on different conjugative pili to initiate new infections. The phage attachment protein, g3p, showed the greatest variation across strains, and each of these variants might correspond to different host specificity. Most of this variation was observed in a glycine-rich repeat region that links N- and C-terminal domains of g3p. Previous work showed that this region can diversify in phage IKE in a few generations in the lab, but no effects on phage fitness were observed (Bruno and Bradbury 1997). Those experiments, however, did not test for changes in host range.

## 5. Conclusion

UP $\phi$  viruses represent an exciting new group of filamentous phages. Many inoviruses play important roles in modifying bacterial pathogens within eukaryotic hosts. It remains to be seen if UP $\phi$  viruses, though frequently found as prophages in bacteria from urine, affect the frequency of UTI or other aspects of urinary health. In the case of pre-term infant microbiomes, we observed no correlation between UP $\phi$  prophage presence and patient outcomes. It is possible that UP $\phi$  phages affect bacterial ecology within the urinary environment without changing bacterial virulence. Perhaps more important will be understanding how frequently the phages shift hosts within different environments and whether these new infections alter the phenotypes of multiple genera within a community.

## Data availability

Data from this work is available at figshare (<https://figshare.com/s/235048b35ae0617dac68>).

## Supplementary data

Supplementary data are available at *Virus Evolution* online.

## Acknowledgments

We are grateful to the Wolfe Lab for generously providing strains of *E. coli* and *C. koseri* used in this work and to Putonti Lab members for helpful comments on the manuscript. We also thank O. Pybus, M. Marston, and two anonymous reviewers for constructive feedback and suggesting a deeper analysis of the metagenomic data and comparison to Ypf $\phi$ .

## Funding

This work is supported by NSF (1661357 to C.P.).

**Conflict of interest:** None declared.

## References

- Altschul, S. F. et al. (1990) 'Basic Local Alignment Search Tool', *Journal of Molecular Biology*, 215: 403–10.
- Arndt, D. et al. (2016) 'PHASTER: A Better, Faster Version of the PHAST Phage Search Tool', *Nucleic Acids Research*, 44: W16–21.
- Aziz, R. K. et al. (2015) 'Multidimensional Metrics for Estimating Phage Abundance, Distribution, Gene Density, and Sequence Coverage in Metagenomes', *Frontiers in Microbiology*, 6: 381.
- Bradley, D. E., Coetzee, J. N., and Hedges, R. W. (1983) 'IncI2 Plasmids Specify Sensitivity to Filamentous Bacteriophage IKE', *Journal of Bacteriology*, 154: 505–7.
- Bruno, R., and Bradbury, A. (1997) 'A Natural Longer Glycine-Rich Region in IKE Filamentous Phage Confers No Selective Advantage', *Gene*, 184: 121–3.
- Bushnell, B. (2014) *BBMap: A Fast, Accurate, Splice-Aware Aligner* (No. LBNL-7065E). Berkeley, CA: Lawrence Berkeley National Lab.
- Camacho, C. et al. (2009) 'BLAST+: Architecture and Applications', *BMC Bioinformatics*, 10: 421.
- Carnoy, C., and Roten, C. A. (2009) 'The Dif/Xer Recombination Systems in Proteobacteria', *PLoS One*, 4: e6531.
- Chouikha, I. et al. (2010) 'Insights into the Infective Properties of Ypf $\phi$ , the *Yersinia pestis* Filamentous Phage', *Virology*, 407: 43–52.
- Davis, B. M., and Waldor, M. K. (2000) 'CTX $\phi$  Contains a Hybrid Genome Derived from Tandemly Integrated Elements', *Proceedings of the National Academy of Sciences*, 97: 8572–7.
- Derbise, A. et al. (2007) 'A Horizontally Acquired Filamentous Phage Contributes to the Pathogenicity of the Plague Bacillus', *Molecular Microbiology*, 63: 1145–57.
- , and Carniel, E. (2014) 'Ypf $\phi$ : A Filamentous Phage Acquired by *Yersinia pestis*', *Frontiers in Microbiology*, 5: 701.
- Edgar, R. C. (2004) 'MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput', *Nucleic Acids Research*, 32: 1792–7.
- Eren, A. M. et al. (2015) 'Anvi'o: An Advanced Analysis and Visualization Platform for 'Omics Data'', *PeerJ*, 3: e1319.

- Garretto, A. (2019) 'Exploring Associations between Lysogeny and Host Abundance', Master's thesis, Loyola University Chicago, Chicago.
- Gonzalez, M. D. et al. (2002) 'Conserved Filamentous Prophage in *Escherichia coli* O18: K1: H7 and *Yersinia pestis* Biovar Orientalis', *Journal of Bacteriology*, 184: 6050–5.
- Jian, H., Xiao, X., and Wang, F. (2013) 'Role of Filamentous Phage SW1 in Regulating the Lateral Flagella of *Shewanella piezotolerans* Strain WP3 at Low Temperatures', *Applied and Environmental Microbiology*, 79: 7101–9.
- Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992) 'The Rapid Generation of Mutation Data Matrices from Protein Sequences', *Bioinformatics*, 8: 275–82.
- Kalyaanamoorthy, S. et al. (2017) 'ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates', *Nature Methods*, 14: 587–9.
- Kamiunten, H., and Wakimoto, S. (1982) 'Effect of Infection with Filamentous Phage Xf on the Growth, Ultrastructure and Virulence of *Xanthomonas campestris* pv. *oryzae* N5850', *Japanese Journal of Phytopathology*, 48: 642–7.
- Katoh, K., and Standley, D. M. (2013) 'MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability', *Molecular Biology and Evolution*, 30: 772–80.
- Kortright, K. E. et al. (2019) 'Phage Therapy: A Renewed Approach to Combat Antibiotic-Resistant Bacteria', *Cell Host & Microbe*, 25: 219–32.
- Letunic, I., and Bork, P. (2019) 'Interactive Tree of Life (iTOL) v4: Recent Updates and New Developments', *Nucleic Acids Research*, 47: W256–9.
- Levi, K. et al. (2018) *Searching the Sequence Read Archive Using Jetstream and Wrangler. Proceedings of the Practice and Experience on Advanced Research Computing*, p. 50.
- Mai-Prochnow, A. et al. (2015) 'Big Things in Small Packages: The Genetics of Filamentous Phage and Effects on Fitness of Their Host', *FEMS Microbiology Reviews*, 39: 465–87.
- May, T., Tsuruta, K., and Okabe, S. (2011) 'Exposure of Conjugative Plasmid Carrying *Escherichia coli* Biofilms to Male-Specific Bacteriophages', *The ISME Journal*, 5: 771–5.
- Midonet, C. et al. (2019) 'The TLCΦ Satellite Phage Harbors a Xer Recombination Activation Factor', *Proceedings of the National Academy of Sciences*, 116: 18391–6.
- Miller-Ensminger, T. et al. (2018) 'Bacteriophages of the Urinary Microbiome', *Journal of Bacteriology*, 200: e00738–17.
- Nguyen, L. T. et al. (2015) 'IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies', *Molecular Biology and Evolution*, 32: 268–74.
- Paradis, E., and Schliep, K. (2019) 'ape 5.0: An Environment for Modern Phylogenetics and Evolutionary Analyses in R', *Bioinformatics*, 35: 526–8.
- Penckofer, S. et al. (2020) 'Characteristics of the Microbiota in the Urine of Women with Type 2 Diabetes', *Journal of Diabetes and Its Complications*, 107561.
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2010) 'FastTree 2—Approximately Maximum-Likelihood Trees for Large Alignments', *PLoS One*, 5: e9490.
- Price, T. K. et al. (2016a) 'Genome Sequences and Annotation of Two Urinary Isolates of *E.coli*', *Standards in Genomic Sciences*, 11: 79.
- et al. (2016b) 'The Clinical Urine Culture: Enhanced Techniques Improve Detection of Clinically Relevant Microorganisms', *Journal of Clinical Microbiology*, 54: 1216–22.
- R Core Team. (2013) *R: A Language for and Environment for Statistical Computing*.
- Rice, S. A. et al. (2009) 'The Biofilm Life Cycle and Virulence of *Pseudomonas aeruginosa* are Dependent on a Filamentous Prophage', *The ISME Journal*, 3: 271–82.
- Roux, S. et al. (2015) 'VirSorter: Mining Viral Signal from Microbial Genomic Data', *PeerJ*, 3: e985.
- et al. (2019) 'Cryptic Inoviruses Revealed as Pervasive in Bacteria and Archaea across Earth's Biomes', *Nature Microbiology*, 4: 1–12.
- Sachs, J. L., and Bull, J. J. (2005) 'Experimental Evolution of Conflict Mediation between Genomes', *Proceedings of the National Academy of Sciences*, 102: 390–5.
- Sieradzki, E. T. et al. (2019) 'Dynamic Marine Viral Infections and Major Contribution to Photosynthetic Processes Shown by Spatiotemporal Picoplankton Metatranscriptomes', *Nature Communications*, 10: 1–9.
- Sweere, J. M. et al. (2019) 'Bacteriophage Trigger Antiviral Immunity and Prevent Clearance of Bacterial Infection', *Science*, 363: eaat9691.
- Val, M. E. et al. (2005) 'The Single-Stranded Genome of Phage CTX is the Form Used for Integration into the Genome of *Vibrio cholerae*', *Molecular Cell*, 19: 559–66.
- Waldor, M. K., and Mekalanos, J. J. (1996) 'Lysogenic Conversion by a Filamentous Phage Encoding Cholera Toxin', *Science*, 272: 1910–4.
- Ward, D. V. et al. (2016) 'Metagenomic Sequencing with Strain-Level Resolution Implicates Uropathogenic *E.coli* in Necrotizing Enterocolitis and Mortality in Preterm Infants', *Cell Reports*, 14: 2912–24.
- Warwick-Dugdale, J. et al. (2019) 'Host-Hijacking and Planktonic Piracy: How Phages Command the Microbial High Seas', *Virology Journal*, 16: 15.
- Weimer, B. C. (2017) '100K Pathogen Genome Project', *Genome Announcements*, 5: e00594–17.
- Yamada, T. (2013) 'Filamentous Phages of *Ralstonia solanacearum*: Double-Edged Swords for Pathogenic Bacteria', *Frontiers in Microbiology*, 4: 325.