

Title: UP ϕ phages, a new group of filamentous phages found in several members of *Enterobacteriales*

Jason W. Shapiro^{1*} and Catherine Putonti^{1,2,3,4}

¹Department of Biology, Loyola University Chicago, Chicago, IL, USA

²Department of Computer Science, Loyola University Chicago, Chicago, IL, USA

³Bioinformatics Program, Loyola University Chicago, Chicago, IL, USA

⁴Department of Microbiology and Immunology, Stritch School of Medicine, Loyola University Chicago, Maywood, IL, USA

*corresponding author email: jshapiro2@luc.edu

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 ϕ 901, which we originally found integrated in a clinical isolate of uropathogenic *Escherichia coli*. UP ϕ 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 ϕ viruses," also harbors two novel genes of unknown function.

Keywords: bacteriophage, inovirus, prophage, bladder

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 (Warwick-Dugdale et al 2019; Mai-Prochnow et al 2015). 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 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).

These filamentous phages can have dramatic effects on their hosts' phenotypes. In *Vibrio cholerae*, the phage CTX ϕ 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 (Gonzalez et al 2002) and agricultural pests (Yamada 2013; Kamiunten and Wakimoto 1982). In other cases, filamentous phages have been found to alter bacterial motility (Jian et al 2013) and biofilm formation (May et al 2011; Rice et al 2009). Despite their myriad effects on hosts, these phages remain understudied. As of this writing, there are only 45 recognized members of *Inoviridae* in RefSeq, and only 33 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 ϕ 901, discovered as a prophage in a clinical *E. coli* isolate from a patient urine sample. UP ϕ 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 ϕ 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 ϕ 901 in published bacterial genomes and assemblies. We found closely related prophages in over two hundred 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 infections; more distant relatives can also be found in soil, water, animal feces, and contaminated food. We refer to the collection of related filamentous phages as "UP ϕ viruses." In several cases, identical UP ϕ strains were found infecting multiple bacterial genera.

Methods

Bacterial strains in lab

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 studies on the urinary microbiota of women with and without symptoms of urinary tract infections or other urinary ailments (Garretto 2019, Price et al. 2016ab). These strains and their GenBank accessions are summarized in Table 1. *C. koseri* strains in this study have not been fully sequenced. *E. 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).

Table 1. Bacterial strains in this study

Strain ID	Species	Has UP ϕ ?	WGS Accession
UMB0731	<i>Escherichia coli</i>	No	NZ_RRWP000000000*
UMB0901	<i>E. coli</i>	No	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>Citrobacter koseri</i>	No	unsequenced**
UMB7451	<i>C. koseri</i>	Yes	unsequenced
UMB8248	<i>C. koseri</i>	No	unsequenced
JE-1	<i>E. coli</i>	No	unavailable

* UMB0901 was sequenced as "E75" in its original BioProject (PRJNA316969)

***Citrobacter* strains not sequenced at time of writing. Phage identified by PCR.

Identifying phage in bacterial genomes

UP ϕ 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). In order to identify UP ϕ 901 in other published genomes, we took advantage of its integration at its hosts' *dif* site as a tandem repeat. We defined a BLAST (Altschul et al 1990) query sequence for UP ϕ 901 as the region starting with the *dif* locus in UMB0901 and extending to the start of the first phage repeat. Throughout this paper, we will refer to the specific sequence (and homologs that are over 99% identical) as "UP ϕ 901." More distantly related phages are collectively referred to as "UP ϕ 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>). This BLAST search identified hits to UP ϕ 901 in *C. koseri*, *E. coli*, *K. pneumoniae*, *S. enterica*, and *Y. enterocolitica*. Notably, neither PHASTER nor VirSorter (Roux et al 2015) consistently predicted UP ϕ 901 homologs in bacteria that had significant BLAST hits. This high false negative rate for these two tools is likely due to UP ϕ 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 2019) with genomes carrying UP ϕ 901 homologs.

We then downloaded all available and complete assemblies on NCBI for the five bacterial species listed above (as of August 2018). Each assembly was queried locally with tblastx to identify the specific start and stop positions of the putative phage regions. The start positions of each phage were confirmed by a separate BLAST search for the *dif* site specific to the bacterial host (Carnoy and Roten 2009), and stop positions were identified by confirming the first repeat in the BLAST results. Because not all assemblies correctly included phage repeats, some sequences did not include a repeat as an obvious stopping point. In these instances, we could only rely on the end position of the last significant hit in the query.

Last, we used searchsra (Levi et al 2018; www.searchsra.org) to check for UP ϕ 901 relatives in metagenomes. We then used the pileup.sh function from BBMap (Bushnell 2014) to check the completeness of UP ϕ 901 coverage for each putative searchsra hit.

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 BLAST with individual phage IKE genes to confirm each gene's correct annotation. Full-length UP ϕ 901 and most of the putative phage regions have every gene found in IKE (as in Figure 2). We then built a phylogeny for this subset of complete genomes with FastTree (Price et al 2010), using the concatenated alignment of each UP ϕ 901 gene amino acid sequence (genes *I*, *II*, *III*, *IV*, *V*, *VI*, *VIII*, *h1*, and *h2*). Individual gene alignments were found using MAFFT (Katoh and Standley 2013). Individual gene trees were also built using FastTree. Genes *VII* and *IX* were excluded, because they consist of only about 30 amino acids and were not consistently identified as genes. Gene *X* was also excluded as it is contained entirely within gene *II*.

Trees were visualized using iTOL (Letunic and Bork 2019) and the ape package (Paradis and Schliep 2018) in R (R Core Team). 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.

We also generated a bacterial phylogeny for each of the phage hosts, using tblastx to identify each of the 24 single-copy universal genes in each genome (Lang et al 2013), MAFFT for alignments, and FastTree for producing the final tree.

Phage presence in culture medium

UP ϕ 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 16,000g for 1 minute, and the supernatants were filtered through 0.2 μ m cellulose acetate syringe filters. 80 μ l of each filtrate were then treated with OPTIZYME DNase I (Fisher BioReagents BP81071) for 30 minutes, followed by heat inactivation with EDTA at 65°C for 10 minutes. 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 minutes to degrade phage protein coats and expose the phage ssDNA. We then amplified phage genomic DNA using UP ϕ 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 ϕ 901 and 16S PCRs were performed on 1 μ l of unfiltered culture as positive controls.

Phage infection assays

We tested the potential for UP ϕ 901 to infect new hosts using standard plaque assays by spotting 5 μ l of filtrate from UMB0901 (as described above) on 0.7% 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 μ l of filtered supernatant from UMB0901. Uninfected controls of each strain were also grown. These cultures were incubated for 18 hours. Following growth, 2 μ 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: CTGGCAGAACGAACGATTAC) recognized a region early in the phage genome. This PCR can only amplify DNA if UP ϕ 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.

Data Availability

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

Results

Initial identification

We originally identified UP ϕ 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 nearly identical phages in three other patient isolates in our lab collection (UMB1220, UMB1526, and UMB5814). As with many other filamentous phages, UP ϕ 901 is integrated as a tandem repeat at the *dif* site in each of these genomes. 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 ϕ 901 DNA (Figure 1a) but negative for bacterial genomic DNA (Figure 1b).

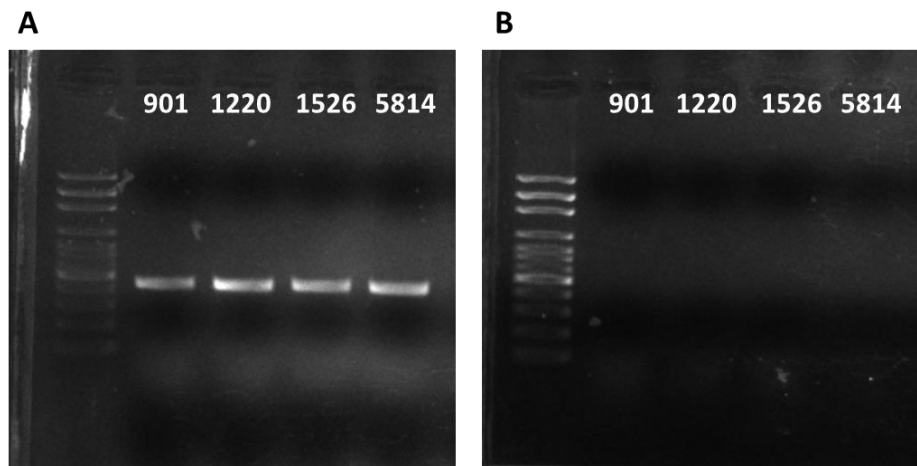


Figure 1. Phage DNA confirmation. PCR results using (A) UP ϕ or (B) 16S primers after DNase treatment to remove bacterial genomic DNA from filtrates of UMB0901, UMB1220, UMB1526, and UMB5814 cultures.

Comparative genomics

UP ϕ 901 is approximately 50% 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 also closely related to the F-specific filamentous phage, M13. Figure 2 shows how UP ϕ 901's genome compares with these, with each gene shaded according to its amino acid similarity with the homolog in UP ϕ 901. UP ϕ 901 contains each of the "core" genes characteristic of filamentous coliphages with gene order preserved.

Two notable differences set UP ϕ 901 apart from I2-2, IKe, and M13: 1) UP ϕ 901 includes a copy of the *E. coli dif* site at the start of its genome, enabling it to integrate via the XerCD recombinase

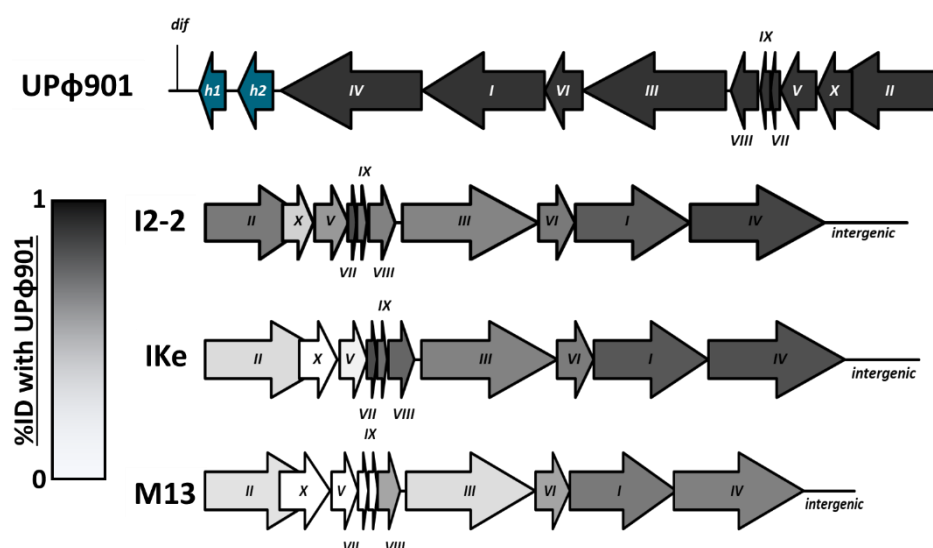


Figure 2. Comparison of UP ϕ 901 to related inoviruses. Roman numerals are the names for core inovirus genes common in M13 and the *Lineavirus* members. Genes shaded grey are darker if they are more similar to the UP ϕ 901 variant. 'h1' and 'h2' are two hypothetical proteins unique to UP ϕ phages. The UP ϕ 901 genes are in reverse orientation due to integration within the host genome at the *dif* site.

system; 2) UP ϕ 901 encodes two 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 homologs of these genes in GenBank returned hits to prophages related to UP ϕ 901.

UP ϕ 901 strains are nearly 100% identical 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 ϕ 901 has an additional repeat of a glycine-rich motif ("GGGES") than the viruses in UMB1220, UMB1526, and UMB5814.

Using the UP ϕ 901 sequence as a query, we searched GenBank for homologous prophages in other bacteria. Closely related viruses 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 related viruses integrated at *dif* sites. In all, 331 genomes harbored a similar prophage (see Table 2 for a summary by host).

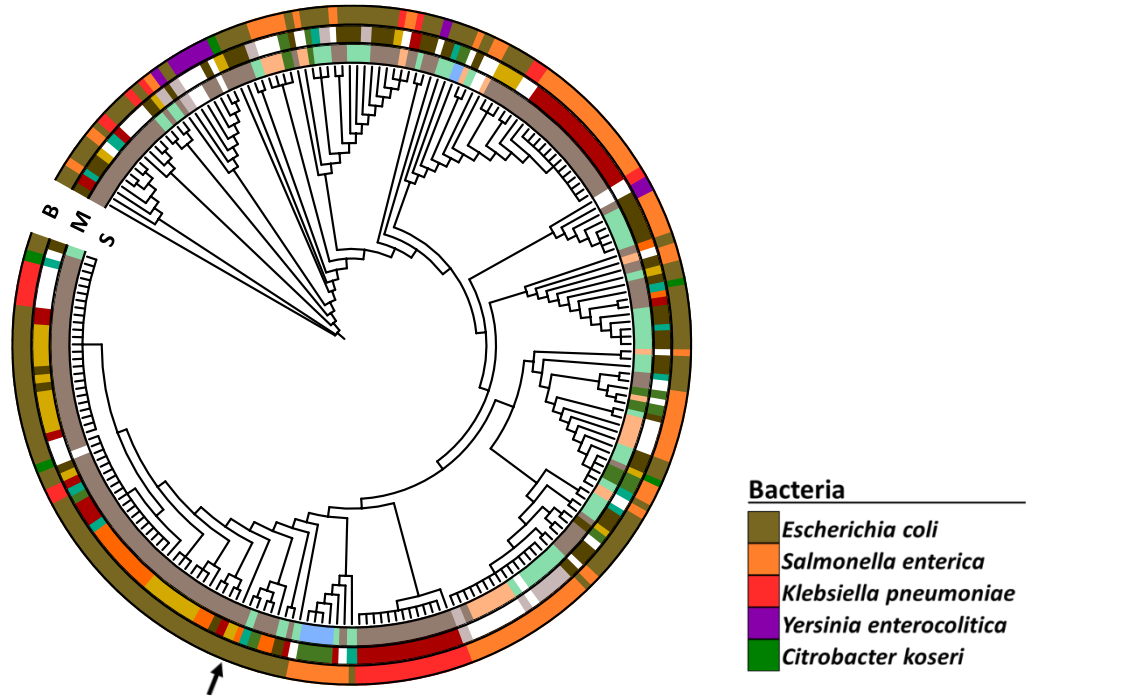
Table 2. Prevalence of UP ϕ in GenBank Assemblies

Host Species	GenBank Assemblies	Assemblies with UP ϕ
<i>Escherichia coli</i>	12,401	189
<i>Salmonella enterica</i>	9,144	95
<i>Klebsiella pneumoniae</i>	4,647	29
<i>Yersinia enterocolitica</i>	177	10
<i>Citrobacter koseri</i>	28	8

229 of the 331 putative phages in GenBank assemblies contained each gene found in UP ϕ 901. For the remainder of this paper, we will refer to this collection of strains as “UP ϕ viruses.” The 102 remaining genomes contained a significant hit to UP ϕ 901, but assembly or sequencing quality was inadequate to predict full phage genomes reliably. We then built a phylogeny for the 229 complete phages (Figure 3a). 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 ϕ 901 is indicated by an arrow in Figure 3a and is part of a clade of closely related phages found predominantly in *E. coli* from urine. Two large polytomies make up the majority of this group, with one consisting mostly of phages found in urine, and the other found in the blood or urine of patients with a urinary tract infection (UTI). Additionally, while most viruses in this urine-associated clade were found in *E. coli*, there are several cases where identical phages were also found within *C. koseri* and *K. pneumoniae*. This group also contains three samples from animals. The rest of the tree is dominated by phages found in *S. enterica* from both animals and humans, though phages infecting other bacteria are also present. Most of these remaining strains were found in feces or contaminated food. This portion of the tree also includes more distantly related phages found in *Y. enterocolitica*.

(A) Phage phylogeny



(B) Host phylogeny

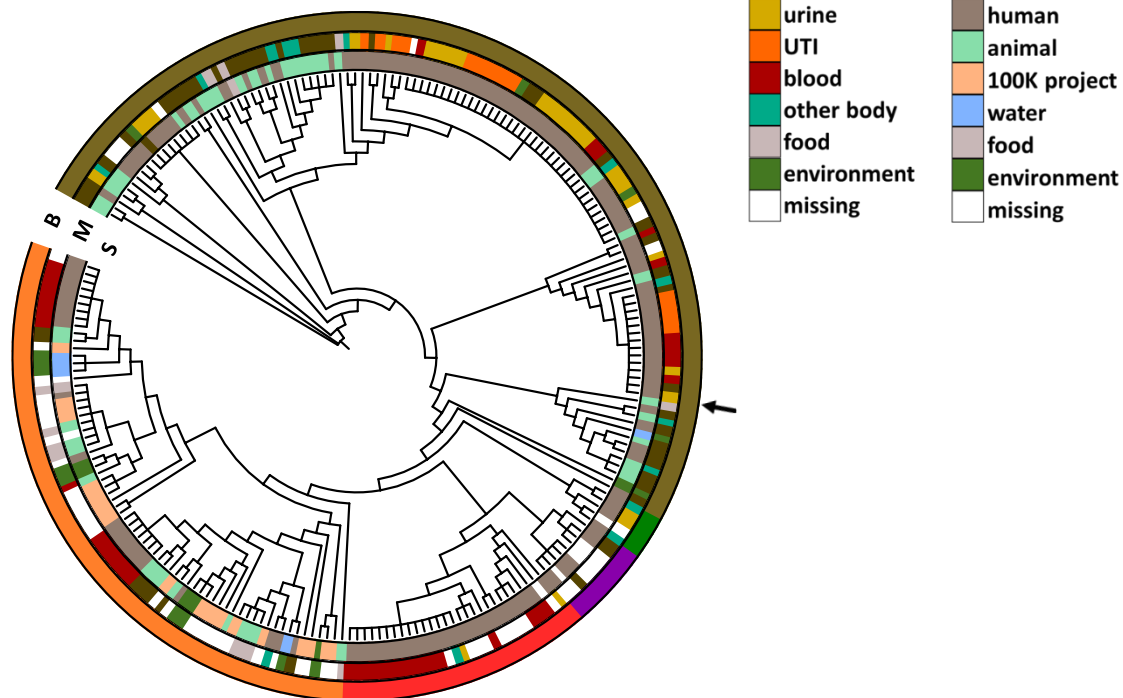


Figure 3. Phylogenies of UP ϕ -related prophages in GenBank assemblies (A) and their hosts (B). The outer ring is the host bacterial species, the middle ring is the source material, and the innermost ring is the source environment. The black arrows indicate UP ϕ 901 (A) and UMB0901 (B).

We also constructed a phylogeny for the host bacteria, using a common set of conserved single-copy genes (Lang et al 2013). 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.

Individual phage gene trees (Supplemental Figure 1) show relatively low variation across the phages, with most genes composed of a small number of variants (Figure 4). Gene *III* stands out as having the most variation, with 95 distinct 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.

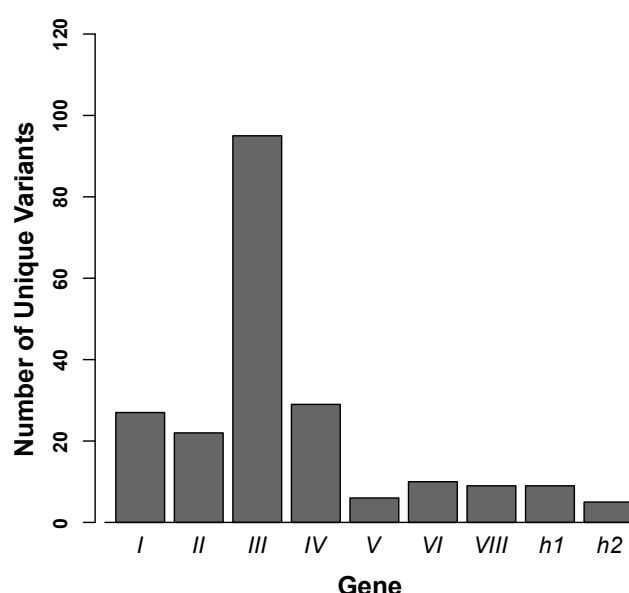


Figure 4. Number of unique sequence variants for each UPφ gene used to build trees.

We also took advantage of the online tool searchsra (Levi et al 2018) to find relatives of UPφ901 in metagenomes. Significant hits were returned for 4819 datasets, with samples from soil, gut, and aquatic environments. Of these, at least 1766 have over 95% coverage of the UPφ901 genome.

Identifying additional infections in the lab

Though only 28 *C. koseri* genomes were available, nearly one third carried a UPφ virus. In two cases, these *C. koseri* phages were identical to prophage sequences found in urine *E. colis*. Given this high frequency of infection and similarity to UPφ901, we tested three unsequenced *C. koseri* isolates (UMB1389, UMB7451, UMB8248) using UPφ901 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 (Supplemental Figure 2).

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 ϕ 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 ϕ 901's ability to establish new infections. These strains all clustered together with UP ϕ 901-infected strains (UMB1220, UMB1526, UMB5814) in a 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 ϕ 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 ϕ 901's closest known relative, I2-2 (Bradley et al 1983). Last, we used PCR to test for phage integration in these candidate hosts but were again unable to confirm new infections.

Discussion

We have introduced a new filamentous phage, UP ϕ 901, found in multiple environments, with nearly identical strains found in clinical isolates of *E. coli* from urine. 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 genera in similar environments. These phages are distinguished from characterized filamentous phages by the addition of two novel genes of unknown function. We have tentatively termed this collection of phages carrying these genes as "UP ϕ viruses."

These UP ϕ viruses are most similar to the characterized 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 ϕ phages. Given the prevalence of *dif* site integration among filamentous phages and their high sequence similarity to much of the UP ϕ 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 ϕ viruses deserve to be identified as a new phage genus. For the time being, we propose that the UP ϕ 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.

A good deal of additional work remains to understand the prevalence and role of UP ϕ 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 ϕ , 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.

Future work that includes long-read resequencing of infected hosts and phages released into culture medium would also help to correct any errors in inferring the phage end positions. While we have identified 331 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 pangenome of these phages, as some might carry additional accessory genes not found in UP ϕ 901.

We also attempted to determine the host requirements for establishing new infections of UP ϕ 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 ϕ 901's closest relatives, I2-2 and IKe, each infect *E. coli* harboring IncI conjugative plasmids, and each *E. coli* strain infected by UP ϕ 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 ϕ viruses rely on different conjugative pili to initiate new infections. The phage attachment protein, g3p, is highly variable 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.

Despite these technical challenges, UP ϕ 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 ϕ viruses, though frequently found in urine, affect the frequency of urinary tract infections or other aspects of urinary health. It is possible that UP ϕ phages might affect bacterial ecology within the urinary environment without changing bacterial virulence. Similarly, it is unknown if these phages play any role in *S. enterica* food contamination. 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.

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. This work is supported by NSF 1661357 to C. Putonti.

References

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J., 1990. Basic local alignment search tool. *Journal of molecular biology*, 215(3): 403-410.
- Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y. and Wishart, D.S., 2016. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic acids research*, 44(W1): W16-W21.
- Bradley, D.E., Coetsee, J.N. and Hedges, R.W., 1983. IncI2 plasmids specify sensitivity to filamentous bacteriophage IKe. *Journal of bacteriology*, 154(1): 505-507.
- Bruno, R. and Bradbury, A., 1997. A natural longer glycine-rich region in IKe filamentous phage confers no selective advantage. *Gene*, 184(1): 121-123.
- Bushnell, B., 2014. *BBMap: a fast, accurate, splice-aware aligner* (No. LBNL-7065E). Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States).
- Carnoy, C. and Roten, C.A., 2009. The dif/Xer recombination systems in proteobacteria. *PloS one*, 4(9): e6531.

Eren, A.M., Esen, Ö.C., Quince, C., Vineis, J.H., Morrison, H.G., Sogin, M.L. and Delmont, T.O. 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.

Gonzalez, M. D., C. A. Lichtensteiger, R. Caughlan, and E. R. Vimr. 2002. Conserved filamentous prophage in *Escherichia coli* O18: K1: H7 and *Yersinia pestis* biovar *orientalis*. *J. Bacteriol.* 184: 6050– 6055.

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. *Appl. Environ. Microbiol.*, 79(22): 7101-7109.

Kamiunten, H., and S. Wakimoto. 1982. Effect of infection with filamentous phage Xf on the growth, ultrastructure and virulence of *Xanthomonas campestris* pv. *oryzae* N5850. *Annal. Phytopathol. Soc. Japan* 48: 642– 647.

Katoh, K. and Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, 30(4): 772-780.

Kortright, K.E., Chan, B.K., Koff, J.L. and Turner: E., 2019. Phage therapy: a renewed approach to combat antibiotic-resistant bacteria. *Cell host & microbe*, 25(2): 219-232.

Lang, J.M., Darling, A.E. and Eisen, J.A., 2013. Phylogeny of bacterial and archaeal genomes using conserved genes: supertrees and supermatrices. *PloS one*, 8(4): e62510.

Letunic, I. and Bork: , 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic acids research*.

Levi, K., Rynge, M., Abeysinghe, E. and Edwards, R.A., 2018, July. Searching the sequence read archive using Jetstream and Wrangler. In *Proceedings of the Practice and Experience on Advanced Research Computing* (p. 50). ACM.

Mai-Prochnow, A., Hui, J.G.K., Kjelleberg, S., Rakonjac, J., McDougald, D. and Rice, S.A., 2015. Big things in small packages: the genetics of filamentous phage and effects on fitness of their host. *FEMS Microbiology Reviews*, 39(4): 465-487.

May, T., K. Tsurata, and S. Okabe. 2011. Exposure of conjugative plasmid carrying *Escherichia coli* biofilms to male-specific bacteriophages. *ISME Journal* 5: 771-775.

Miller-Ensminger, T., Garretto, A., Brenner, J., Thomas-White, K., Zambom, A., Wolfe, A.J. and Putonti, C., 2018. Bacteriophages of the urinary microbiome. *Journal of bacteriology*, 200(7): e00738-17.

Paradis, E. and Schliep, K. 2018. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35: 526-528.

Price, M.N., Dehal, P.S. and Arkin, A.P. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. *PloS one*, 5(3): e9490.

Price, T.K. et al., 2016. Genome sequences and annotation of two urinary isolates of *E. coli*. *Standards in genomic sciences*, 11(1): 79.

Price, T.K. et al., 2016. The clinical urine culture: enhanced techniques improve detection of clinically relevant microorganisms. *Journal of clinical microbiology*, 54(5): 1216-1222.

R Core Team, 2013. R: A language for and environment for statistical computing.

Rice, S.A., C.H. Tan, J. Mikkelsen, V. Kung, J. Woo, M. Tay, A. Hauser, D. McDougald, J. S. Webb, and S. Kjelleberg., 2009. The biofilm life cycle and virulence of *Pseudomonas aeruginosa* are dependent on a filamentous prophage. *ISME Journal* 3: 271-282.

Riechmann, L. and Holliger, P., 1997. The C-terminal domain of TolA is the coreceptor for filamentous phage infection of *E. coli*. *Cell*, 90(2): 351-360.

Roux, S., Enault, F., Hurwitz, B.L. and Sullivan, M.B., 2015. VirSorter: mining viral signal from microbial genomic data. *PeerJ*, 3: e985.

Roux, S., Krupovic, M., Daly, R.A., Borges, A.L., Nayfach, S., Schulz, F., Sharrar, A., Carnevali, B.M., Cheng, J.F., Ivanova, N.N. and Bondy-Denomy, J., 2019. Cryptic inoviruses revealed as pervasive in bacteria and archaea across Earth's biomes. *Nature microbiology*: 1-12.

Val, M.E., Bouvier, M., Campos, J., Sherratt, D., Cornet, F., Mazel, D. and Barre, F.X.. 2005. The single-stranded genome of phage CTX is the form used for integration into the genome of *Vibrio cholerae*. *Molecular cell*, 19(4): 559-566.

Waldor, M. K., and J. J. Mekalanos. 1996. Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* 272: 1910-1914.

Warwick-Dugdale, J., Buchholz, H.H., Allen, M.J. and Temperton, B., 2019. Host-hijacking and planktonic piracy: how phages command the microbial high seas. *Virology journal*, 16(1): 15.

Weimer, B.C., 2017. 100K Pathogen genome project. *Genome Announc.*, 5(28): e00594-17.

Yamada, T. 2013. Filamentous phages of *Ralstonia solanacearum*: double-edged swords for pathogenic bacteria. *Front. Microbiol.* 4: 325.