Biology: Faculty Publications and Other Works

Faculty Publications and Other Works by Department

4-1-2023

Coliphages of the human urinary microbiota

Elias Crum
Loyola University Chicago

Zubia Merchant
Loyola University Chicago

Adriana Ene
Loyola University Chicago

Taylor Miller-Ensminger
Loyola University Chicago

Genevieve Johnson
Loyola University Chicago

See next page for additional authors

Follow this and additional works at: https://ecommons.luc.edu/biology_facpubs

Authors

Elias Crum, Zubia Merchant, Adriana Ene, Taylor Miller-Ensminger, Genevieve Johnson, Alan J. Wolfe, and Catherine Putonti

RESEARCH ARTICLE

Coliphages of the human urinary microbiota

Elias Crum[1,2☉], Zubia Merchant[2☉], Adriana Ene[1,2], Taylor Miller-Ensminger[1],
Genevieve Johnson[1], Alan J. Wolfe[3], Catherine Putonti[1,2,3]*

1 Bioinformatics Program, Loyola University Chicago, Chicago, Illinois, United States of America,
2 Department of Biology, Loyola University Chicago, Chicago, Illinois, United States of America,
3 Department of Microbiology and Immunology, Stritch School of Medicine, Loyola University Chicago,
Maywood, Illinois, United States of America

☉ These authors contributed equally to this work.
* cputonti@luc.edu

Abstract

Due to its frequent association with urinary tract infections (UTIs), *Escherichia coli* is the
best characterized constituent of the urinary microbiota (urobiome). However, uropatho-
genic *E. coli* is just one member of the urobiome. In addition to bacterial constituents, the
urobiome of both healthy and symptomatic individuals is home to a diverse population of
bacterial viruses (bacteriophages). A prior investigation found that most bacterial species in
the urobiome are lysogens, harboring one or more phages integrated into their genome (pro-
phages). Many of these prophages are temperate phages, capable of entering the lytic
cycle and thus lysing their bacterial host. This transition from the lysogenic to lytic life cycle
can impact the bacterial diversity of the urobiome. While many phages that infect *E. coli*
(coliphages) have been studied for decades in the laboratory setting, the coliphages within
the urobiome have yet to be cataloged. Here, we investigated the diversity of urinary coli-
phages by first identifying prophages in all publicly available urinary *E. coli* genomes. We
detected 3,038 intact prophage sequences, representative of 1,542 unique phages. These
phages include both novel species as well as species also found within the gut microbiota.
Ten temperate phages were isolated from urinary *E. coli* strains included in our analysis,
and we assessed their ability to infect and lyse urinary *E. coli* strains. We also included in
these host range assays other urinary coliphages and laboratory coliphages. The temperate
phages and other urinary coliphages were successful in lysing urinary *E. coli* strains. We
also observed that coliphages from non-urinary sources were most efficient in killing urinary
E. coli strains. The two phages, T2 and N4, were capable of lysing 83.5% (n = 86) of strains
isolated from females with UTI symptoms. In conclusion, our study finds a diverse commu-
nity of coliphages in the urobiome, many of which are predicted to be temperate phages, ten
of which were confirmed here. Their ability to infect and lyse urinary *E. coli* strains suggests
that urinary coliphages may play a role in modulating the *E. coli* strain diversity of the
urobiome.

Introduction

While several different bacterial species are known to cause urinary tract infections (UTIs), uropathogenic *Escherichia coli* (UPEC) is estimated to account for up to 74.4% of all community-acquired infections [1]. The presence of *E. coli* is often taken as evidence of UTI. While UTIs can be the result of colonization of the urinary tract by *E. coli* strains from the gut [2–5], *E. coli* can also be a member of the urinary microbiota (urobiome) of individuals without lower urinary tract symptoms (LUTS) [6–8]. In fact, recent studies found that *E. coli* can be the predominant taxon in female without LUTS, particularly in some older females [8, 9]. Furthermore, human host genetic make-up can contribute to the presence of *E. coli* in the urine of females without UTI symptoms [9].

The urobiome of individuals with or without LUTS is home to a wide variety of other bacterial taxa (see reviews [10–13]), fungi [14], and viruses [15]. In fact, viruses are the most abundant members of the urobiome and recently have been associated with LUTS [16]. Viruses that infect bacteria (bacteriophages or phages) far outnumber human viruses in the urobiome [15, 17, 18]. Similar observations have been made in other organs' microbiota [19]. Phages can drive bacterial diversity within a community through predation [20–22]. Phages that integrate into their bacterial host's genome (prophages) can increase the virulence of their host [23, 24], e.g., by encoding for toxins [25, 26]. While studies of phage-bacteria dynamics have yet to be conducted for the urobiome, investigations in the gut microbiota are ongoing [27].

Prior studies of urinary bacterial genomes found that most strains harbor one or more prophage sequences [28–30]. This includes bacterial species associated with urinary health, e.g., *Lactobacillus* species [28, 29], as well as bacterial species associated with UTIs, e.g., *E. coli* [7, 28] and *Proteus mirabilis* [28]. These phages replicate with their bacterial host and are either integrated in the bacterial host genome or persist as an extrachromosomal plasmid; this is referred to as the lysogenic life cycle of the phage. Our previous work with urinary strains harboring prophages found that many of these prophages could switch from the lysogenic life cycle to the lytic (predatory) life cycle [28, 30–32]. This switch, a process called induction, is mediated by intrinsic and/or extrinsic factors that cause the prophage to excise itself from the bacterial genome, replicate to produce mature (lytic) phages, and burst or kill the host cell (see review [33]). Bladder-relevant stressors, e.g., changes in pH, have been shown to be effective in inducing prophages [31].

Prophages and phage genes have been routinely identified in urinary *E. coli* genome sequences [28, 34–37]. While phages and prophages that infect *E. coli* (coliphages) in the lab have been studied for decades [38] and coliphages from urine samples have been isolated and characterized [39–42], coliphage diversity within the urobiome has yet to be thoroughly investigated. Here, we present the first catalog of urinary *E. coli* prophages. All publicly available urinary *E. coli* genomes were examined for prophage sequences and the diversity and genic content of these prophages was explored. Ten urinary prophages were then induced from urinary *E. coli* strains, and we assessed their ability to lyse laboratory and urinary *E. coli* strains isolated from females with UTI or overactive bladder (OAB) symptoms. Additionally, we tested several urinary coliphages and laboratory coliphages against these same set of urinary *E. coli* strains. We find that coliphages are ubiquitous in the urobiome, including temperate coliphages capable of lysing other urinary *E. coli* strains.

Methods

Prophage identification

All publicly available *E. coli* complete and draft genome assemblies in NCBI that were documented as being collected from urine or the urinary tract in the genome metadata were downloaded (February 2021). 906 genomes were obtained from NCBI meeting this criterion; 2 sequences were removed due to quality concerns. The final set of 904 *E. coli* sequences (**S1 Table**) were examined using PHASTER [43]. A Python script was written to pull results from the PHASTER API and to separate PHASTER predicted "intact," "questionable," and "incomplete" prophage sequences. This script is available at https://github.com/putonti/phaster_commands.

All intact prophage sequences were compared via local blastn queries to a database of all complete and partial phage genome sequences in GenBank as of February 2021 (Advance Query Fields—Organism: "Virus" and Division: "PHG"). This database includes 26,381 sequences. Results with a query coverage greater than 50% and a percent identity greater than 70% were considered high confidence hits and the associated taxonomies of the GenBank phage records were used to predict the taxonomies of the predicted urinary phages.

Furthermore, all intact prophage sequence were screened for antibiotic resistance genes using the RGI tool v.5.2.1, which uses CARD [44], exploring perfect, strict, and loose hits, and virulence factors via the Virulence Factor Database (VFDB) [45]. RGI was installed locally using conda, and the CARD database (v.3.1.4) was downloaded. The RGI program was run using default parameters. For virulence factor screening, the full dataset of known bacterial virulence factor gene sequences was downloaded from VFDB on March 2021 (http://www.mgc.ac.cn/VFs/download.htm). These sequences were made into a local database via the makeblastdb command (BLAST+ v.2.9.0). Prophage sequences were queried against this database using BLASTn with the parameter -evalue 0.001. Hits of <90% sequence identity and query length <90% of the virulence factor gene sequence were removed from further consideration. The virulence factor gene descriptions were determined using VFDB.

The PATRIC online tool was used to annotate the intact prophage sequences with the "Annotation Recipe" parameter set to "Bacteriophage" [46]. All annotated genes containing "integrase" in the description were added to a multi-FASTA file, and Kalign v.2.04 was used to produce a multiple sequence alignment with default parameters [47]. The multiple sequence alignment was manually inspected using Geneious Prime 2021.2.2 (Dotmatics, Auckland, NZ). Partial integrase gene sequences were identified through this manual inspection and were removed from the data set. The remaining integrase gene sequences were aligned again with Kalign. This alignment aided in confirming the identity of integrase genes within the predicted prophage sequences.

Prophage network construction

Using Anvi'o v.6.2 [48], the intact prophage sequences were annotated, and a coliphage pangenome was constructed. Prophage sequences were made into an Anvi'o database annotated using the 'anvi-run-hmms' command. The annotated prophage sequences were then used to produce a coliphage pan-genome using the Anvi'o 'anvi-pan-genome' command with an mclinflation of 2 and a minbit of 0.35 to identify homologous genes among the prophage sequences.

An R script (www.R-project.org) was written to derive a network of prophages. The methods described here for phage network construction were adapted from our prior work [49]. Using the output result (mcl-cluster.txt) of the Anvi'o mediated clustering using the Markov

Clustering Algorithm (MCL), the MCL results were translated into a genome-gene presence/absence matrix, P, in which each entry {i,j} was 1 if virus genome i contained a homolog found in gene cluster j. This matrix is equivalent to the adjacency matrix for a bipartite network of phage genomes and genes. Adjacency matrices for the genome and gene level networks were then created as $A_{\text{genome}} = \text{sign}(P \times P^T)$ and $A_{\text{gene}} = \text{sign}(P^T \times P)$, where T indicates the matrix transpose. The sign() function replaced all nonzero entries resulting from the original matrix products with a 1, converting the matrices from weighted to unweighted adjacency matrices. These matrices were then transformed into undirected graphs and corresponding edge lists using igraph (https://igraph.org/). Thus, for the genome-level network, two genomes are considered connected if they share any genes. The connections were filtered using a normalization calculation: $w = (\# \text{ of shared genes between genomes 1 \& 2})/\sqrt{(l_1 \times l_2)}$, where l_i is the size of the genome i. By designating a minimum value of w ($minw$) that allows for an edge to be drawn between two genomes only if $w > minw$, the edges were filtered to construct networks of differing connectivity.

The edge-lists constructed from the edge-drawing Rscript were then visualized using Cytoscape [50]. Different values of $minw$ were considered for the PHASTER predicted prophages.

Prophage sequence clustering

All intact prophage sequences were clustered using cd-hit-est v.4.6 [51, 52]. The following parameters were used: sequence identity threshold = 80% (0.8), length of difference cutoff = 80% (0.8), word length = 4. The "accurate but slow mode" algorithm was used. For each cluster, cd-hit-est selected one sequence as a representative sequence (indicated by "*" in the output .clstr file). For more details regarding the process implemented by CD-HIT algorithms for selecting representative sequences, see Huang et al. [53]. Sequences of representatives of each cluster are available upon request.

Comparison of urinary coliphages to gut phageome

All phage sequences from metagenomic data sets from the gut microbiome were retrieved from the mMGE database [54] (accessed July 2021). Intact urinary prophage sequences were compared against this database using a local blastn query. For these results, our thresholds were query coverage $\geq 50\%$ and sequence identity $\geq 70\%$, although all results meeting the query coverage threshold had a sequence identity $\geq 80\%$.

Prophage induction

Nine strains of *E. coli* isolated from urine samples were used for induction experiments. The urinary strains were selected based upon their genome analysis by PHASTER [43]; all 9 were predicted to contain at least one intact prophage sequence. For pH-based induction, we used the previously described protocol [31]. Briefly, urinary *E. coli* strains were grown overnight in LB at 37˚C with shaking. These strains include *E. coli* UMB0527, UMB6653, UMB6721, UMB9006, UMB9105, UMB9208, UMB9344, UMB9346, and UMB9930. While UMB9006 was obtained from a "clean-catch" voided urine sample, the remaining 8 samples were isolated from catheterized urine samples. These strains were obtained through prior IRB-approved studies (Loyola University Chicago: LU204195 and LU209545, and University of California San Diego: 170077AW) as part of prior separate studies [55, 56]. The overnight culture was then subcultured into 3 mL of LB adjusted to different pH values, pH = 4, 7, and 9 and grown overnight at 37˚C with shaking. These pH values were selected informed by our prior work [31]. These pH-adjusted cultures were filtered using a 0.22um CA syringe filter. Filtrate was

next spotted onto lawns of naïve laboratory strains of *E. coli* B, *E. coli* C, and *E. coli* K-12. For each lawn, 500 uL of turbid (overnight) *E. coli* culture + 3 mL of soft (0.7%) LB agar were mixed and spread atop a 1.7% LB agar plate. The spot plates were then incubated overnight at 37˚C. Plaques were harvested and used to reinfect the laboratory strain, i.e., the bacteria of the lawn from which the plaque was harvested, and incubated again overnight at 37˚C. These cultures were filtered as previously described and plated using the pour plate technique (100 uL of phage lysate + 500 uL of turbid *E. coli* culture + 3 mL of soft LB agar); a single plaque was picked, suspended in LB, vortexed, filtered, and added to turbid *E. coli* overnight cultures. This process was repeated at least 3 times to plaque purify the phage and subsequently to create higher titer stocks of the induced urinary prophage.

Prophage identification was performed by PCR. Primers were designed using Primer-BLAST [57] to amplify the PHASTER predicted intact prophage sequences. A primer pair was designed for each individual predicted intact prophage sequence; thus, a strain harboring multiple predicted intact prophage sequences would have multiple pairs (**S2 Table**). Preference was given to primers that amplified coding regions with a predicted protein function, i.e., not hypothetical proteins. All predicted prophage sequences were annotated using the RAST server [58] and visualized using Geneious Prime. If more than one intact prophage was predicted for a given strain, primers were designed for each predicted prophage. Primers were synthesized by Eurofins Genomics LLC (Louisville, KY USA). The original bacterial strain was used as a positive control for PCR reactions of its respective induced prophages. Amplification was confirmed via agarose gel. The PHASTER predicted sequences were queried online against the nr/nt viruses (taxid:10239) to identify the closest related sequenced phage.

Phage host range

Urinary and laboratory *E. coli* strains were grown in LB overnight at 37˚C with shaking. Each was lawned, following the protocol listed previously, and 10 uL of phage lysate (at titer 10^9 phage per mL) was spotted on each lawn. Each phage lysate was spotted onto each *E. coli* strain a minimum of 4 times (technical replicates). In addition to assessing the host range for the induced urinary phages, two other urinary coliphages previously isolated by our group [41] as well as non-urinary coliphages were spotted on urinary and laboratory *E. coli* strains. These coliphages include *Escherichia* phage Greed and *Escherichia* phage Lust, both isolated from urobiome samples, and non-urinary coliphages K30, P22, T2, T3, T6, T7 and N4 obtained from the Félix d'Hérelle Reference Center for Bacterial Viruses (Quebec City, Quebec Canada). Phages T2 and N4 were further tested on urinary *E. coli* strains collected from UTI-positive females (**S3 Table**). Plaques of T2 and N4 spots were confirmed via PCR of the plaque. T2 primers: 5'-aaacaggtgcctttggtgtc-3' and 5'-ccacaatacccgcttcagtt-3'; N4 primers: 5'-tgctcttgatac-cagaggcaatg-3' and 5'-tacgttggttcaacttcttggtt-3'. Primers were synthesized by Eurofins Genomics LLC. For all phages producing plaques for the host range spot assays, infection was again tested by harvesting the plaque and replating using the pour plates technique previously described.

Results

Prophages in urinary *E. coli* genomes

904 urinary *E. coli* draft and complete genome sequences were obtained from NCBI and screened for the presence of prophage via PHASTER [43]. Sixty-nine of these genome assemblies are of strains from our own collection. Of the 8,452 predicted prophage sequences found, 3,038 prophages were identified as intact, 1,508 as questionable, and 3,906 as incomplete. We focused our analysis on those identified as intact. 45 of the 904 genomes examined did not

contain any intact prophages, while the majority (95%) of the urinary *E. coli* genomes harbored at least one intact prophage (**S1 Table**). 1,807 (59.48%) of the intact prophage sequences contained an identifiable integrase gene sequence (**S1 Table**). Each intact prophage sequence was compared to all annotated phage genome sequences to predict their taxonomy (**Table 1**).

Virulence factor genes were identified in 25% (n = 765) of the intact coliphage sequences. These virulence-carrying intact prophage sequences were harbored by 710 of the 904 *E. coli* genomes examined. The most frequently identified virulence factor gene was *aaiQ* (n = 137), a pseudogene that has been linked to enteroaggregative *E. coli* pathogenesis [59]. The iron uptake *sit* operon was the next most prevalent virulence factor identified in 61 of the intact prophage sequences. When screened for antibiotic resistance associated genes, 299 genes were identified. These genes were found in 178 different intact prophage sequences (5.86%) from 165 different *E. coli* genomes. The most frequently identified antibiotic resistance gene was the ethidium multidrug resistance protein E (*emrE*), found in 47 of the prophage sequences. The transcription factor *marA* was the second most frequent antibiotic resistance associated gene found; 37 intact prophage sequences encoded for *marA*. While most strains had just one or two antibiotic resistance genes, *E. coli* Combat11I9 (GCF_002952095.1) includes one prophage sequence with 14 antibiotic resistance genes. **S1 Table** lists information about the virulence factors and antibiotic resistance associated genes found within the prophage sequences.

Intact prophages were annotated and the number of homologous genes between each prophage was calculated. Prophage similarity was assessed using a network approach in which nodes in the network represent a single prophage. Two nodes are connected by an edge in the network if they share a homologous gene sequence. We introduced a threshold for these edges such that only edges in which the two prophages (nodes) shared at least 30% of their genic content were visualized (see Methods). 3,025 of the 3,038 predicted prophages met this threshold, connected by 414,291 edges. **Fig 1** displays the network representation of the urinary prophages.

Nodes, representative of prophages, have been colored in this network according to their predicted taxonomic family. Over half of the identified prophages most closely resembled tailed phages (order Caudovirales), including the families Myoviridae, Podoviridae, and Siphoviridae. However, 43.9% of the predicted intact prophage sequences in the urinary *E. coli* genomes did not have significant similarity to previously isolated and sequenced phages. Our analysis of the shared genetic content of the predicted urinary prophages reveals nine connected components (**Fig 1**, labeled A-I). The individual connected components do not generally share homologous genes (see Methods). While the largest connected component, labeled A in **Fig 1**, includes the majority of Caudovirales, our network contains two distinct connected components of siphoviruses (**Fig 1D and 1G**, orange); these are separate from the siphoviruses within the primary (largest) connected component. There also are six distinct connected

Table 1. Taxonomic classification of predicted prophages in urinary *E. coli* based on blastn sequence similarity to sequenced phages.

Taxonomy	# Predicted Prophages
Myoviruses	788
Podoviruses	113
Siphoviruses	738
Unclassified Caudovirales	44
Unclassified bacterial viruses	20
Unknown	1,335

https://doi.org/10.1371/journal.pone.0283930.t001
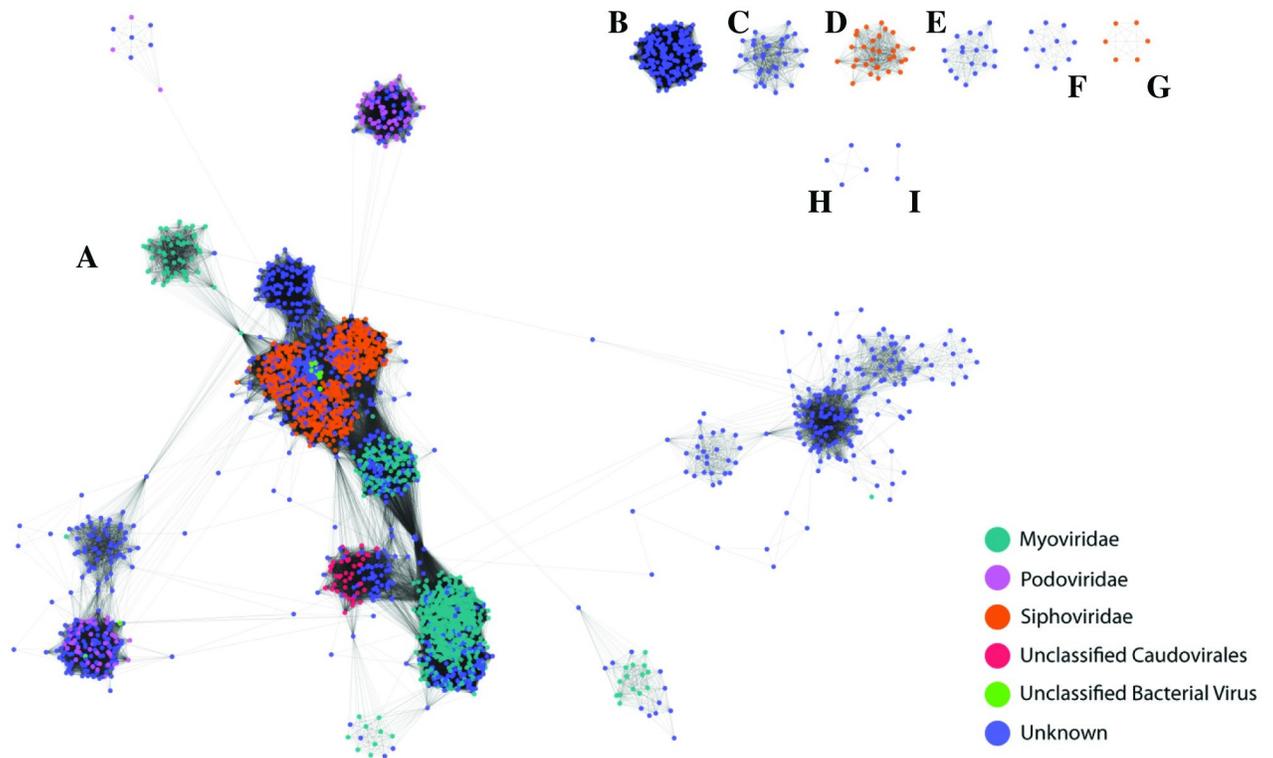
Fig 1. Diversity of urinary *E. coli* prophages. Each node corresponds with a predicted intact coliphage. Edges connecting nodes represent shared gene content. Individual connected components, clusters of prophages that share gene content, are indicated by letters A through I.

https://doi.org/10.1371/journal.pone.0283930.g001

components of prophages for which we were unable to predict their taxonomic lineage (shown in **Fig 1B, 1C, 1E, 1F, 1H and 1I** in blue for "Unknown").

Next, we investigated the network of prophages to identify clusters of similar prophage sequences, i.e., closely related prophages likely representative of the same species/strain. In total, 1,542 unique clusters of prophage sequences were identified (**S4** and **S5 Tables**). The size of these clusters varied from singletons, i.e., single representative prophage sequences (n = 1,115, 72.31%) to a single cluster with 127 representatives (length ~43–50 Kbp; 87.73% nucleotide sequence identity between members). The consensus sequence of this large cluster was queried against the nr/nt Viruses database, identifying the top hit as a MAG sequence from the human metagenome (GenBank Accession No.: BK034715.1; 76% query coverage and 99.98% identity). Most of the clusters were small; 95% of the clusters had 5 or less prophage sequences. Four of the clusters had two prophage sequences identified from the same genome sequence. While two of these were the only instances of these prophage sequences in the data set, the other two were found in other genomes, including the cluster of 127 representatives. There is no association between cluster size and prophage sequence length (**S5 Table**). Sequences for the cluster representative are provided in **S1 File**.

Comparison to gut *E. coli* prophages

Given prior evidence of UTIs by colonization of *E. coli* from the gut, we investigated whether the same or similar phage populations were harbored by urinary *E. coli* strains and phages from the gut. The 3,038 intact urinary prophages were screened against the phage database of the gut microbiome. These gut microbiome phage sequences include phages identified from

metagenomic studies and, as such, include phages infectious of bacteria other than *E. coli*; furthermore, the host genus or species for most of these phage sequences is not known. As a result of this comparison, we found 2,006 urinary coliphages that exhibited $\geq 50\%$ query coverage and $\geq 80\%$ nucleotide sequence identity to phage sequences from the gut. This is representative of 863 of the 1,542 unique clusters of urinary prophages, and 560 of these clusters only had one representative sequence from the urinary coliphages. Furthermore, 117 intact urinary prophages, representative of 49 unique clusters, had a query coverage of 100% to a gut phage sequences ($\geq 96\%$ sequence identity) and 21 were identical (100% query coverage and sequence identity) to gut phage sequences (**S6 Table**). Thirteen of these 21 prophage sequences shared no significant sequence similarity to an annotated phage genome sequence. Thus, no taxonomic classification could be assigned. These 21 prophages sequences were also screened for antibiotic resistance genes and virulence factors, although none were found. The 21 identical prophages represent 12 unique clusters, including small clusters with just 2 members as well as the largest observed cluster with 127 representatives.

Inducing urinary *E. coli* prophages

Nine urinary *E. coli* strains in our collection were selected for induction assays. These strains were selected as they were predicted to contain prophage sequences from singleton clusters, i.e., the urinary strain was the only genome predicted to contain this prophage, as well as prophage sequences from larger clusters, including clusters with prophages from *E. coli* strains from collections other than our own. For those prophages in singleton clusters, induction would lay the groundwork for future characterization of this phage strain. For those prophages in larger clusters, induction would provide insight into the putative temperance of the phage across many different urinary *E. coli* strains.

Using changes in pH, 10 induced prophages were identified as they were efficient in completely lysing one or more of the naïve laboratory strains tested–*E. coli* B, *E. coli* C, and *E. coli* K-12. We were able to identify the prophage that was induced via PCR (see Methods). Eight of the induced prophage sequences included recognizable integrase genes; i527 and i9930-2 did not. i6721 and i6653 were induced at all three pH conditions tested, pH = 4, 7 and

Table 2. pH-induced prophages from urinary *E. coli* strains and their closest characterized and sequenced phage. Phages are named as i plus the *E. coli* strain number from which they were induced. As two phages were isolated from *E. coli* UMB9930, they are signified as "-1" and "-2".

Phage ID	pH Condition (*E. coli* host)	Predicted Product Amplified by PCR	Closest Blast Hit			
			Description	Accession No.	Query Coverage	Sequence Identity
i527	4 (*E. coli* C)	DUF2560 family protein	Enterobacteria phage CUS-3	CP000711.1	60%	96.04%
i6653	4, 7 and 9 (*E. coli* C)	Phage terminase, endonuclease subunit GpM	Bacteriophage L-413C	AY251033.1	77%	97.80%
i6721	4, 7 and 9 (*E. coli* C)	Phage terminase, endonuclease subunit GpM	Enterobacteria phage fiAA91-ss	NC_022750.1	75%	93.48%
i9006	4 (*E. coli* C)	Phage replication protein GpA, endonuclease	*Escherichia* virus P2_4E6b	NC_049389.1	79%	96.52%
i9105	7 (*E. coli* C)	Phage tail fiber protein GpH	Bacteriophage L-413C	AY251033.1	88%	99.97%
i9208	4 (*E. coli* C)	Phage tail fiber protein GpH	*Escherichia* phage pro147	KR073660.1	77%	96.87%
i9344	4 (*E. coli* C)	Phage major tail tube protein GpFII	*Escherichia* virus P2_4C9	NC_049388.1	73%	96.24
i9346	7 (*E. coli* C)	Phage lysis regulatory protein, LysA	Bacteriophage R18C	NC_049461.1	68%	97.31%
i9930-1	7 (*E. coli* C)	Phage baseplate assembly protein GpJ	*Escherichia* virus P2_4C9	NC_049388.1	73%	96.24
i9930-2	7 (*E. coli* K-12)	Phage head, portal protein B	Stx-1a-converting phage Stx1_499	LC567825.1	96%	94.88%

9; the phages isolated from pH = 7 were used for subsequent testing. The closest characterized and sequenced phage record in GenBank was identified for each of the induced prophage sequences (**Table 2**). Upon further inspection, it was found that the induced prophage from UMB9344 (i9344) is identical–query coverage and sequence identity = 100%–to one of the two induced prophages from UMB9930, i9930-1. The induced prophages from UMB5563 (i6653) and UMB6721 (i6721) also were similar, belonging to the same cluster (**S7 Table**). A visualization of the induced prophage genomes is provided in **S1 Fig**.

Phage host range

Each induced urinary coliphage was tested for its ability to completely lyse laboratory strains of *E. coli*, as well as urinary *E. coli* strains representative of the diversity of *E. coli* phylotypes found within the female bladder [7]. In parallel, we tested two lytic urinary siphoviruses previously isolated by our group [41], *Escherichia* phage Greed and *Escherichia* phage Lust, alongside several well-studied lytic coliphages routinely used in the laboratory—*Escherichia* phage K30, Enterobacteria T3 and Enterobacteria T7 (family Autographiviridae), *Salmonella* virus P22 (family Podoviridae), Enterobacteria phage T2 and T6 (family Myoviridae), and *Escherichia* virus N4 (family Schitoviridae). The results of these host range assays are shown in **Table 3**.

The well-studied phages T2 and N4 were most efficient in completely lysing *E. coli* strains isolated from UTI patients. Each was capable of completely lysing eight of the ten urinary

Table 3. Phage efficacy of lysing laboratory and urinary bacteria. If a phage completely lysed the *E. coli* strain, the table lists "+". The phylotype for each urinary *E. coli* strain is listed [7]. *E. coli* UMB0103 was isolated from a female with OAB symptoms; the other urinary *E. coli* strains were isolated from females with UTI symptoms.

	LABORATORY *E. COLI* STRAINS			URINARY *E. COLI* STRAINS									
				A	B1	B2				D			F
	B	C	K-12	1358	1180	1195	5924	1220	1162	1225	1337	7431	103
Induced Phages													
i527		+					+	+		+			
i6653	+	+	+				+			+			
i6721		+		+				+	+	+			+
i9006		+											
i9105		+								+			
i9208		+	+				+			+			
i9344		+	+				+			+			
i9346		+								+			
i9930-1		+	+				+			+			
i9930-2		+	+				+						
Urinary Phages													
Lust	+	+	+				+	+	+				
Greed	+	+	+				+	+	+			+	
Laboratory Phages													
K30	+	+						+	+	+			
P22	+					+		+				+	
T2	+	+	+	+	+		+	+	+	+		+	+
T3	+	+	+			+	+	+	+			+	+
T6	+	+	+		+	+	+	+	+			+	+
T7	+	+	+			+	+	+				+	
N4		+	+	+	+	+		+	+	+	+		+

strains tested. Given this success, we further tested T2 and N4 against an additional 103 *E. coli* strains isolated from urine samples from females with UTI symptoms. A list of the strains tested, including serotype and phylotype, are listed (**S3 Table**). Sixty-three strains were completely lysed by T2, 71 strains were completely lysed by N4, 48 strains were completely lysed by both T2 and N4, and 17 strains were not completely lysed by either phage (**Table 4**). Thus, 83.5% of the *E. coli* strains tested from females with UTI symptoms would be susceptible to a phage cocktail containing T2 and N4.

Discussion

While our prior catalog of prophages in urinary bacteria considered just four *E. coli* strains [28], here we have evaluated the incidence of lysogeny among a significantly larger (n = 906) sampling of urinary *E. coli* strains. While all four of the genomes examined in our prior study were found to harbor prophage sequences, only one was predicted to include an intact prophage [28]. Here, the majority (95%) of the urinary *E. coli* genomes harbored at least one intact prophage sequence indicating that lysogeny is prevalent among *E. coli* strains of the urobiome. While some (~25%) of these prophage sequences carried virulence factor genes, very few encoded antibiotic resistance genes (**S1 Table**). One of the lysogen-associated proteins, the integrase, was identified in many (59.48%) of these intact prophage sequences. Those lacking the integrase necessitate further investigation to ascertain if they are temperate phages or prophage relics.

   We found that only 56.1% of the predicted intact prophage sequences resembled characterized phage sequences. The remaining prophage sequences either represent novel prophages or highly mosaic prophages infectious of *E. coli*. This concurs with our prior examination of genomes of other taxa from the urobiome, which found a high percentage of unknown, novel prophages [28]. It is important to note that the intact prophages examined here do not include any representatives of inoviruses. Current prophage prediction tools–including PHASTER– frequently fail to identify inovirus sequences [60–62]. Previously, we identified an inovirus harbored by Enterobacteriaceae including urinary *E. coli* strains [63]. Inovirus sequences could be included in the questionable and/or incomplete prophage sequence predictions excluded from our analysis. We thus hypothesize that the intact prophage sequences examined here underestimates the genetic diversity of prophages within urinary *E. coli* strains. It does, however, provides a good approximation of the diversity of tailed phages infectious of urinary *E. coli*.

   Our network-based analysis of the shared genetic content between the predicted urinary prophages identified nine connected components (**Fig 1**). The largest connected component contains the majority of the tailed phages identified. From the network, we can posit taxonomic classification of "Unknown" (**Fig 1**, blue) prophages within this connected component as members of Caudovirales. It is important to note that our network analysis includes a threshold for the minimum percentage of gene content shared for an edge to be drawn. Thus, edges represent "modules" shared between prophages. When this threshold is removed, such

Table 4. Susceptibility of *E. coli* strains from females with UTI symptoms to bacteriophages T2 and N4.

	# of *E. coli* lysed	% of *E. coli* lysed
Both T2 & N4	48	46.6%
N4 only	23	22.3%
T2 only	15	14.6%
Neither	17	16.5%

https://doi.org/10.1371/journal.pone.0283930.t004

that every homologous gene shared between two prophages (nodes) is represented by an edge, all prophages are connected producing a single connected component. Previous analyses of phage genomes have taken a similar modular approach [64, 65]. In this prior work, modules in temperate phages were found to correspond to functional modules [64]. These modules often serve as markers corresponding to the evolutionary history of related phages [66].

Of the 1,542 unique clusters of urinary prophage sequences identified, many (55.97%, n = 863) shared sequence similarity to phages identified in the gut microbiota suggesting that similar coliphages infect *E. coli* strains found in both niches. The interconnectedness of the urobiome and gut microbiota is an open question. Prior studies have shown that *E. coli* causing UTIs can come from the gastrointestinal tract [2–5]. However, a previous examination of the bacterial constituents of the gut and urinary tract did not find the two niches to be connected [67]. To date, the interconnectedness of the phage populations between these two microbiota have not been examined. The similarities observed here serve as the impetus for future studies.

Our ability to induce prophages from the urinary *E. coli* strains supports the working hypothesis that the PHASTER predictions of intact prophages are temperate phages, rather than prophage relics. Changes in pH were able to stress the bacterial cell and trigger the induction process. Changes in pH are particularly relevant to the urinary tract and its microbiota. *Lactobacillus* species are dominant members of the healthy female bladder [8], and lactobacilli reduce the pH of the urogenital environment (see review [68]). Thus, fluctuations in lactobacilli abundances within the urinary tract could lead to changes in pH, which in turn could trigger induction. While nine out of the ten induced urinary prophages completely lyse the naïve laboratory strain *E. coli* C, they have varied abilities to completely lyse urinary *E. coli* strains. Two main observations can be made from our host range assays (**Table 3**). First, no single induced urinary phage was capable of completely lysing all of the urinary *E. coli* strains tested. Second, induced urinary phages capable of completely lysing one representative of a phylotype were not necessarily able to completely lyse all of the strains tested for that same phylotype. While only a single strain from phylotype A, B1, and F were tested, multiple representatives of phylotypes B2 and D, the most common phylotypes in the bladder [7], were tested. It is well documented that phages often are infectious of some but not all strains of a given species (see review [69]). The observed lack of plaquing by these induced urinary phages could be the result of (1) the phage's inability to infect the *E. coli* strain, (2) the phage's inability to evade the host's defenses (e.g., CRISPR/Cas system), (3) the phage's inability to completely lyse the bacterial cell, (4) entrance of the temperate phage back into the lysogenic life cycle, and/or (5) bacterial resistance due to superinfection.

In contrast, the lytic laboratory coliphages were far more successful in completely lysing the urinary *E. coli* strains. They also were far more successful than the lytic urinary phages Lust and Greed. T2 and N4 were able to complete lyse at least one strain from all five phylotypes tested. The susceptibility of the urinary *E. coli* strains to these lytic laboratory coliphages may be due to the very fact that they are less likely to encounter these phages in the urinary tract than they are to encounter the induced urinary prophages. This suggests that phages that are not native to the urinary tract would be better candidates for use as phage therapies. Phage therapy is increasingly being explored for treatment of bacterial infections including *E. coli* and UTIs (see reviews [70–72]). Our further tests of T2 and N4 phages found that they were able to completely lyse most UTI-associated *E. coli* strains tested (**Table 4**), and thus should be explored for UTI treatments.

Supporting information

S1 Table. Urinary *E. coli* genomes retrieved from NCBI and screened for prophage sequences.
(XLSX)

S2 Table. PCR primers used to identify induced prophages.
(XLSX)

S3 Table. Urinary *E. coli* strains tested for susceptibility to T2 and N4. S3 Strains were either isolated from an individual with an acute urinary tract infection (UTI) or recurrent UTI (rUTI).
(XLSX)

S4 Table. Prophage clusters identified in each urinary *E. coli* genomes.
(XLSX)

S5 Table. Cluster sizes and length of the representative sequence for the cluster.
(XLSX)

S6 Table. Urinary prophages identical to gut phage sequences.
(XLSX)

S7 Table. Cluster information for induced prophages.
(XLSX)

S1 Fig. Annotations of induced prophages.
(PDF)

S1 File. Nucleotide sequences of the cluster representatives.
(FASTA)

Acknowledgments

Author Contributions

Conceptualization: Catherine Putonti.

Data curation: Elias Crum.

Formal analysis: Elias Crum, Zubia Merchant, Adriana Ene, Taylor Miller-Ensminger.

Investigation: Elias Crum, Zubia Merchant, Genevieve Johnson.

Resources: Alan J. Wolfe, Catherine Putonti.

Supervision: Alan J. Wolfe, Catherine Putonti.

Writing – original draft: Elias Crum.

Writing – review & editing: Zubia Merchant, Adriana Ene, Taylor Miller-Ensminger, Genevieve Johnson, Alan J. Wolfe, Catherine Putonti.

References

1. Foxman B. Urinary tract infection syndromes: occurrence, recurrence, bacteriology, risk factors, and disease burden. Infect Dis Clin North Am. 2014; 28: 1–13. https://doi.org/10.1016/j.idc.2013.09.003 PMID: 24484571

2. Yamamoto S, Tsukamoto T, Terai A, Kurazono H, Takeda Y, Yoshida O. Genetic evidence supporting the fecal-perineal-urethral hypothesis in cystitis caused by Escherichia coli. J Urol. 1997; 157: 1127–1129. PMID: 9072556

3. Hooton TM. Recurrent urinary tract infection in women. Int J Antimicrob Agents. 2001; 17: 259–268. https://doi.org/10.1016/s0924-8579(00)00350-2 PMID: 11295405

4. Chen SL, Wu M, Henderson JP, Hooton TM, Hibbing ME, Hultgren SJ, et al. Genomic diversity and fitness of E. coli strains recovered from the intestinal and urinary tracts of women with recurrent urinary tract infection. Sci Transl Med. 2013; 5: 184ra60. https://doi.org/10.1126/scitranslmed.3005497 PMID: 23658245

5. Thänert R, Reske KA, Hink T, Wallace MA, Wang B, Schwartz DJ, et al. Comparative Genomics of Antibiotic-Resistant Uropathogens Implicates Three Routes for Recurrence of Urinary Tract Infections. mBio. 2019;10. https://doi.org/10.1128/mBio.01977-19 PMID: 31455657

6. Thomas-White KJ, Gao X, Lin H, Fok CS, Ghanayem K, Mueller ER, et al. Urinary microbes and postoperative urinary tract infection risk in urogynecologic surgical patients. Int Urogynecol J. 2018; 29: 1797–1805. https://doi.org/10.1007/s00192-018-3767-3 PMID: 30267143

7. Garretto A, Miller-Ensminger T, Ene A, Merchant Z, Shah A, Gerodias A, et al. Genomic Survey of E. coli From the Bladders of Women With and Without Lower Urinary Tract Symptoms. Frontiers in Microbiology. 2020;11. https://doi.org/10.3389/fmicb.2020.02094 PMID: 33013764

8. Price TK, Hilt EE, Thomas-White K, Mueller ER, Wolfe AJ, Brubaker L. The urobiome of continent adult women: a cross-sectional study. BJOG. 2020; 127: 193–201. https://doi.org/10.1111/1471-0528.15920 PMID: 31469215

9. Adebayo AS, Ackermann G, Bowyer RCE, Wells PM, Humphreys G, Knight R, et al. The Urinary Tract Microbiome in Older Women Exhibits Host Genetic and Environmental Influences. Cell Host Microbe. 2020; 28: 298–305.e3. https://doi.org/10.1016/j.chom.2020.06.022 PMID: 32697939

10. Whiteside SA, Razvi H, Dave S, Reid G, Burton JP. The microbiome of the urinary tract—a role beyond infection. Nature Reviews Urology. 2015; 12: 81–90. https://doi.org/10.1038/nrurol.2014.361 PMID: 25600098

11. Brubaker L, Wolfe A. The urinary microbiota: a paradigm shift for bladder disorders? Curr Opin Obstet Gynecol. 2016; 28: 407–412. https://doi.org/10.1097/GCO.0000000000000298 PMID: 27379439

12. Mueller ER, Wolfe AJ, Brubaker L. Female urinary microbiota. Curr Opin Urol. 2017; 27: 282–286. https://doi.org/10.1097/MOU.0000000000000396 PMID: 28234750

13. Neugent ML, Hulyalkar NV, Nguyen VH, Zimmern PE, De Nisco NJ. Advances in Understanding the Human Urinary Microbiome and Its Potential Role in Urinary Tract Infection. mBio. 2020;11. https://doi.org/10.1128/mBio.00218-20 PMID: 32345639

14. Ackerman AL, Underhill DM. The mycobiome of the human urinary tract: potential roles for fungi in urology. Annals of Translational Medicine. 2017; 5: 31–31. https://doi.org/10.21037/atm.2016.12.69 PMID: 28217696

15. Garretto A, Miller-Ensminger T, Wolfe AJ, Putonti C. Bacteriophages of the lower urinary tract. Nat Rev Urol. 2019; 16: 422–432. https://doi.org/10.1038/s41585-019-0192-4 PMID: 31073244

16. Sun Q, Li L, Zhou H, Wu Y, Gao Y, Wu B, et al. The detection of urinary viruses is associated with aggravated symptoms and altered bacteriome in female with overactive bladder. Front Microbiol. 2022; 13: 984234. https://doi.org/10.3389/fmicb.2022.984234 PMID: 36212847

17. Santiago-Rodriguez TM, Ly M, Bonilla N, Pride DT. The human urine virome in association with urinary tract infections. Front Microbiol. 2015; 6: 14. https://doi.org/10.3389/fmicb.2015.00014 PMID: 25667584

18. Garretto A, Thomas-White K, Wolfe AJ, Putonti C. Detecting viral genomes in the female urinary microbiome. J Gen Virol. 2018; 99: 1141–1146. https://doi.org/10.1099/jgv.0.001097 PMID: 29889019

19. Barr JJ. A bacteriophages journey through the human body. Immunological Reviews. 2017; 279: 106–122. https://doi.org/10.1111/imr.12565 PMID: 28856733

20. Buckling A, Rainey PB. Antagonistic coevolution between a bacterium and a bacteriophage. Proc Biol Sci. 2002; 269: 931–936. https://doi.org/10.1098/rspb.2001.1945 PMID: 12028776

21. Rodriguez-Valera F, Martin-Cuadrado A-B, Rodriguez-Brito B, Pasić L, Thingstad TF, Rohwer F, et al. Explaining microbial population genomics through phage predation. Nat Rev Microbiol. 2009; 7: 828–836. https://doi.org/10.1038/nrmicro2235 PMID: 19834481

22. Koskella B, Brockhurst MA. Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. FEMS Microbiol Rev. 2014; 38: 916–931. https://doi.org/10.1111/1574-6976.12072 PMID: 24617569

23. Hosseinidoust Z, van de Ven TGM, Tufenkji N. Evolution of Pseudomonas aeruginosa virulence as a result of phage predation. Appl Environ Microbiol. 2013; 79: 6110–6116. https://doi.org/10.1128/AEM.01421-13 PMID: 23892756

24. Wendling CC, Piecyk A, Refardt D, Chibani C, Hertel R, Liesegang H, et al. Tripartite species interaction: eukaryotic hosts suffer more from phage susceptible than from phage resistant bacteria. BMC Evol Biol. 2017; 17: 98. https://doi.org/10.1186/s12862-017-0930-2 PMID: 28399796

25. Wagner PL, Waldor MK. Bacteriophage control of bacterial virulence. Infect Immun. 2002; 70: 3985–3993. https://doi.org/10.1128/IAI.70.8.3985-3993.2002 PMID: 12117903

26. Casas V, Maloy S. Role of bacteriophage-encoded exotoxins in the evolution of bacterial pathogens. Future Microbiol. 2011; 6: 1461–1473. https://doi.org/10.2217/fmb.11.124 PMID: 22122442

27. De Sordi L, Lourenço M, Debarbieux L. The Battle Within: Interactions of Bacteriophages and Bacteria in the Gastrointestinal Tract. Cell Host Microbe. 2019; 25: 210–218. https://doi.org/10.1016/j.chom.2019.01.018 PMID: 30763535

28. Miller-Ensminger T, Garretto A, Brenner J, Thomas-White K, Zambom A, Wolfe AJ, et al. Bacteriophages of the Urinary Microbiome. Zhulin IB, editor. Journal of Bacteriology. 2018; 200: e00738–17. https://doi.org/10.1128/JB.00738-17 PMID: 29378882

29. Miller-Ensminger T, Mormando R, Maskeri L, Shapiro JW, Wolfe AJ, Putonti C. Introducing Lu-1, a Novel Lactobacillus jensenii Phage Abundant in the Urogenital Tract. PLoS One. 2020; 15: e0234159. https://doi.org/10.1371/journal.pone.0234159 PMID: 32525961

30. Brassil B, Mores CR, Wolfe AJ, Putonti C. Characterization and spontaneous induction of urinary tract Streptococcus anginosus prophages. J Gen Virol. 2020; 101: 685–691. https://doi.org/10.1099/jgv.0.001407 PMID: 32310742

31. Miller-Ensminger T, Garretto A, Stark N, Putonti C. Mimicking prophage induction in the body: induction in the lab with pH gradients. PeerJ. 2020; 8: e9718. https://doi.org/10.7717/peerj.9718 PMID: 32944418

32. Johnson G, Putonti C. Genome Sequence of Pseudomonas Phage UMP151, Isolated from the Female Bladder Microbiota. Microbiol Resour Announc. 2019;8. https://doi.org/10.1128/MRA.00853-19 PMID: 31416880

33. Nanda AM, Thormann K, Frunzke J. Impact of spontaneous prophage induction on the fitness of bacterial populations and host-microbe interactions. J Bacteriol. 2015; 197: 410–419. https://doi.org/10.1128/JB.02230-14 PMID: 25404701

34. Beatson SA, Ben Zakour NL, Totsika M, Forde BM, Watts RE, Mabbett AN, et al. Molecular analysis of asymptomatic bacteriuria Escherichia coli strain VR50 reveals adaptation to the urinary tract by gene acquisition. Infect Immun. 2015; 83: 1749–1764. https://doi.org/10.1128/IAI.02810-14 PMID: 25667270

35. Price TK, Mehrtash A, Kalesinskas L, Malki K, Hilt EE, Putonti C, et al. Genome sequences and annotation of two urinary isolates of E. coli. Stand Genomic Sci. 2016; 11: 79. https://doi.org/10.1186/s40793-016-0202-6 PMID: 27777649

36. Subhadra B, Kim DH, Kim J, Woo K, Sohn KM, Kim H-J, et al. Complete genome sequence of uropathogenic Escherichia coli isolate UPEC 26–1. Genes Genomics. 2018; 40: 643–655. https://doi.org/10.1007/s13258-018-0665-5 PMID: 29892948

37. Garretto A. Exploring associations between lysogeny and host abundance. MS, Loyola University Chicago. 2019.

38. Brockhurst MA, Koskella B. Experimental coevolution of species interactions. Trends Ecol Evol. 2013; 28: 367–375. https://doi.org/10.1016/j.tree.2013.02.009 PMID: 23523051

39. Dallas SD, Kingsbery L. Bacteriophage Plaques on Primary Isolation Media of a Urine Culture Growing Escherichia coli. Clinical Microbiology Newsletter. 1997; 19: 53–56. https://doi.org/10.1016/S0196-4399(97)87667-2

40. Brown-Jaque M, Muniesa M, Navarro F. Bacteriophages in clinical samples can interfere with microbiological diagnostic tools. Scientific Reports. 2016;6. https://doi.org/10.1038/srep33000 PMID: 27609086

41. Malki K, Sible E, Cooper A, Garretto A, Bruder K, Watkins SC, et al. Seven Bacteriophages Isolated from the Female Urinary Microbiota. Genome Announc. 2016;4. https://doi.org/10.1128/genomeA.01003-16 PMID: 27881533

42. Pacífico C, Hilbert M, Sofka D, Dinhopl N, Pap I-J, Aspöck C, et al. Natural Occurrence of Escherichia coli-Infecting Bacteriophages in Clinical Samples. Front Microbiol. 2019; 10: 2484. https://doi.org/10.3389/fmicb.2019.02484 PMID: 31736918

43. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 2016; 44: W16–21. https://doi.org/10.1093/nar/gkw387 PMID: 27141966

44. Alcock BP, Huynh W, Chalil R, Smith KW, Raphenya AR, Wlodarski MA, et al. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. Nucleic Acids Res. 2023; 51: D690–D699. https://doi.org/10.1093/nar/gkac920 PMID: 36263822

45. Liu B, Zheng D, Jin Q, Chen L, Yang J. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. Nucleic Acids Res. 2019; 47: D687–D692. https://doi.org/10.1093/nar/gky1080 PMID: 30395255

46. Davis JJ, Wattam AR, Aziz RK, Brettin T, Butler R, Butler RM, et al. The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. Nucleic Acids Research. 2019; gkz943. https://doi.org/10.1093/nar/gkz943 PMID: 31667520

47. Lassmann T, Sonnhammer EL. Kalign–an accurate and fast multiple sequence alignment algorithm. BMC Bioinformatics. 2005; 6: 298. https://doi.org/10.1186/1471-2105-6-298 PMID: 16343337

48. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ. 2015; 3: e1319. https://doi.org/10.7717/peerj.1319 PMID: 26500826

49. Shapiro JW, Putonti C. Gene Co-occurrence Networks Reflect Bacteriophage Ecology and Evolution. MBio. 2018;9. https://doi.org/10.1128/mBio.01870-17 PMID: 29559574

50. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13: 2498–2504. https://doi.org/10.1101/gr.1239303 PMID: 14597658

51. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006; 22: 1658–1659. https://doi.org/10.1093/bioinformatics/btl158 PMID: 16731699

52. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012; 28: 3150–3152. https://doi.org/10.1093/bioinformatics/bts565 PMID: 23060610

53. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics. 2010; 26: 680–682. https://doi.org/10.1093/bioinformatics/btq003 PMID: 20053844

54. Lai S, Jia L, Subramanian B, Pan S, Zhang J, Dong Y, et al. mMGE: a database for human metagenomic extrachromosomal mobile genetic elements. Nucleic Acids Research. 2021; 49: D783–D791. https://doi.org/10.1093/nar/gkaa869 PMID: 33074335

55. Barnes HC, Wolff B, Abdul-Rahim O, Harrington A, Hilt EE, Price TK, et al. A Randomized Clinical Trial of Standard versus Expanded Cultures to Diagnose Urinary Tract Infections in Women. Journal of Urology. 2021 [cited 17 Aug 2021]. https://doi.org/10.1097/JU.0000000000001949 PMID: 34184930

56. Hochstedler BR, Burnett L, Price TK, Jung C, Wolfe AJ, Brubaker L. Urinary microbiota of women with recurrent urinary tract infection: collection and culture methods. Int Urogynecol J. 2021 [cited 17 Aug 2021]. https://doi.org/10.1007/s00192-021-04780-4 PMID: 33852041

57. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics. 2012; 13: 134. https://doi.org/10.1186/1471-2105-13-134 PMID: 22708584

58. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. BMC Genomics. 2008; 9: 75. https://doi.org/10.1186/1471-2164-9-75 PMID: 18261238

59. Dudley EG, Thomson NR, Parkhill J, Morin NP, Nataro JP. Proteomic and microarray characterization of the AggR regulon identifies a pheU pathogenicity island in enteroaggregative Escherichia coli. Mol Microbiol. 2006; 61: 1267–1282. https://doi.org/10.1111/j.1365-2958.2006.05281.x PMID: 16925558

60. Roux S, Hallam SJ, Woyke T, Sullivan MB. Viral dark matter and virus–host interactions resolved from publicly available microbial genomes. eLife. 2015;4. https://doi.org/10.7554/eLife.08490 PMID: 26200428

61. Paez-Espino D, Eloe-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N, et al. Uncovering Earth's virome. Nature. 2016; 536: 425–430. https://doi.org/10.1038/nature19094 PMID: 27533034

62. Paez-Espino D, Chen I-MA, Palaniappan K, Ratner A, Chu K, Szeto E, et al. IMG/VR: a database of cultured and uncultured DNA Viruses and retroviruses. Nucleic Acids Res. 2017; 45: D457–D465. https://doi.org/10.1093/nar/gkw1030 PMID: 27799466

63. Shapiro JW, Putonti C. UPΦ phages, a new group of filamentous phages found in several members of Enterobacteriales. Virus Evolution. 2020.

64. Lima-Mendez G, Toussaint A, Leplae R. A modular view of the bacteriophage genomic space: identification of host and lifestyle marker modules. Research in Microbiology. 2011; 162: 737–746. https://doi.org/10.1016/j.resmic.2011.06.006 PMID: 21767638

65. Iranzo J, Krupovic M, Koonin EV. The Double-Stranded DNA Virosphere as a Modular Hierarchical Network of Gene Sharing. mBio. 2016; 7: e00978–16. https://doi.org/10.1128/mBio.00978-16 PMID: 27486193

66. Koonin EV, Dolja VV, Krupovic M, Varsani A, Wolf YI, Yutin N, et al. Global Organization and Proposed Megataxonomy of the Virus World. Microbiol Mol Biol Rev. 2020; 84: e00061–19. https://doi.org/10.1128/MMBR.00061-19 PMID: 32132243

67. Thomas-White K, Forster SC, Kumar N, Van Kuiken M, Putonti C, Stares MD, et al. Culturing of female bladder bacteria reveals an interconnected urogenital microbiota. Nat Commun. 2018; 9: 1557. https://doi.org/10.1038/s41467-018-03968-5 PMID: 29674608

68. Stapleton AE. The Vaginal Microbiota and Urinary Tract Infection. Microbiol Spectr. 2016;4. https://doi.org/10.1128/microbiolspec.UTI-0025-2016 PMID: 28087949

69. de Jonge PA, Nobrega FL, Brouns SJJ, Dutilh BE. Molecular and Evolutionary Determinants of Bacteriophage Host Range. Trends Microbiol. 2019; 27: 51–63. https://doi.org/10.1016/j.tim.2018.08.006 PMID: 30181062

70. Bolocan AS, Callanan J, Forde A, Ross P, Hill C. Phage therapy targeting Escherichia coli-a story with no end? FEMS Microbiol Lett. 2016;363: fnw256. https://doi.org/10.1093/femsle/fnw256 PMID: 27974392

71. Zalewska-Piątek BM, Piątek RJ. Alternative treatment approaches of urinary tract infections caused by uropathogenic Escherichia coli strains. Acta Biochim Pol. 2019; 66: 129–138. https://doi.org/10.18388/abp.2018_2787 PMID: 31136644

72. Chegini Z, Khoshbayan A, Vesal S, Moradabadi A, Hashemi A, Shariati A. Bacteriophage therapy for inhibition of multi drug-resistant uropathogenic bacteria: a narrative review. Ann Clin Microbiol Antimicrob. 2021; 20: 30. https://doi.org/10.1186/s12941-021-00433-y PMID: 33902597