

LOYOLA UNIVERSITY CHICAGO

CATALOGING AND ENGINEERING TEMPERATE COLIPHAGES OF THE HUMAN
URINARY MICROBIOME

A THESIS SUBMITTED TO
THE FACULTY OF THE GRADUATE SCHOOL
IN CANDIDACY FOR THE DEGREE OF
MASTER OF SCIENCE

PROGRAM IN BIOINFORMATICS

BY
ELIAS D. CRUM
CHICAGO, IL
AUGUST 2022

Copyright by Elias Crum, 2022
All rights reserved.

ACKNOWLEDGEMENTS

I would like to thank Dr. Catherine Putonti for her constant guidance and support throughout my time. I would also like to thank the members of the Putonti lab for their help in all of my projects, specifically Genevieve Johnson and Dr. Jason Shapiro. I want to thank the members of my committee, Dr. Alan Wolfe and Dr. Michael Grillo, for lending their knowledge and assistance to my project. Finally, I would like to thank my family and the many other unnamed mentors that have provided their guidance and support throughout my academic career.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
LIST OF TABLES	v
LIST OF FIGURES	vi
ABSTRACT	viii
CHAPTER ONE: INTRODUCTION	1
The Human Urobiome	1
<i>Escherichia coli</i>	3
Phages - Overview	5
Scope of Thesis	12
CHAPTER 2: BIOINFORMATIC IDENTIFICATION, COMPARISON, AND CHARACTERIZATION OF UROBIOME COLIPHAGES	13
Introduction	13
Methods	19
Results	23
Discussion	29
CHAPTER 3: UROBIOME COLIPHAGE INDUCTION AND HOST CHANGE CONFIRMATION EXPERIMENTS	37
Introduction	37
Methods	42
Results	53
Discussion	63
CHAPTER 4: CONCLUSIONS	69
APPENDIX A: PROPHAGE VIRULENCE FACTOR GENE DESCRIPTIONS	71
APPENDIX B: PROPHAGE ANTIBIOTIC RESISTANCE GENE TARGETS	73
APPENDIX C: PATIENT UMB <i>E. COLI</i> STRAIN ISOLATION IRB DETAILS	75
REFERENCE LIST	78
VITA	92

LIST OF TABLES

Table 1. Results of phage lambda blast screening.	42
Table 2. Pcr primers used to identify induced prophages.	45
Table 3. Superinfection blast hits.	51
Table 4. Umb <i>e. Coli</i> strains used for induction and host range.	51
Table 5. Induced phage stock phaster sequence identities.	54
Table 6. Integrase gene modification details.	57
Table 7. I700 and d700 host range assay outcomes.	60

LIST OF FIGURES

Figure 1. Phage lytic and lysogenic lifecycles.	8
Figure 2. Temperate phage integration and excision diagram.	9
Figure 3. Histogram of prophages per <i>e. Coli</i> genome.	24
Figure 4. Prophage network.	28
Figure 5. Phylogenetic tree of prophage integrase genes.	29
Figure 6. Induced prophage stock identification pcr products.	56
Figure 7. Integrase gene cleavage confirmation pcr results.	59
Figure 8. Example lytic and non-lytic spotting assay results.	61
Figure 9. D700 lysis following transformation.	62
Figure 10. Spot titration assays of i700 and d700.	62

ABSTRACT

In the US, around 60% of females will be diagnosed with a Urinary Tract Infection (UTI) in their lifetime, and *Escherichia coli* is the most implicated etiological agent of UTIs. Despite its frequent association with lower urinary tract symptoms, recent studies have found that the urinary microbiome (UMB), the viral, bacterial, and fungal resident members of the urinary tract of healthy females, can also consist of *E. coli*. While most research has focused on the bacterial constituents of the UMB, bacteriophages, viruses that infect bacteria, are far more abundant. Bacteriophages (phages) of other human microbiomes have been shown to play a significant role in shaping the bacterial community dynamics. While a handful of *E. coli* infecting phages (coliphages) have been characterized, little is known about the diversity of coliphages of the human UMB or their role within this microbial community.

Of the known types of phages in nature, temperate phages have complex, broad impacts on bacterial populations and, therefore, community dynamics, through both lytic and lysogenic means. One such widely studied impact is lysogenic conversion, the alteration of host phenotype by the infection by integration of a temperate phage. Lysogenic conversion has the potential to offer a host bacterium human-relevant phenotypes, including antibiotic resistance and virulence factor production. Since it has been shown that UMB phages can exhibit host ranges broader than just their native host, phages could be facilitating the horizontal gene transfer of potentially harmful genes within the human UMB.

Despite the threat posed by UMB temperate phages, little work has been done characterizing their integrated form, prophages, in UMB *E. coli*. Additionally, the ability of these phages to be induced from their native host, then infect and integrate into a non-native host has not been shown previously by phages of the human UMB.

Here, I present a two-part study into (*i*) the diversity of *E. coli* prophages within the human UMB and (*ii*) the ability of UMB temperate coliphages to be induced from a native host, engineered to be obligately lytic to then infect and lyse a non-native UMB host.

First, 3,177 predicted intact prophages were identified from 961 urinary *E. coli* genome assemblies by the prophage predicting tool PHASTER. Prophages were pervasive; 95% of the strains contained at least one intact prophage (average >3 prophages per genome assembly). Furthermore, ~50% of these predicted prophages did not share significant sequence similarity to characterized phages, thus likely representing novel phages. Some predicted prophage sequences contained antibiotic resistance and virulence genes, but these genes were not common among the prophage sequences. Investigation of the predicted prophages' integrase gene suggests that the UMB coliphages share common attachment sites for integration in the *E. coli* chromosome. Collectively, this study provides the first catalog of urinary *E. coli* temperate phages.

Next, predicted intact prophages were induced from urinary *E. coli* strains using biologically relevant pH conditions. The induced phages were identified using PCR with primers designed based on the PHASTER sequence predictions. One of these induced prophages was then engineered from temperate to obligately lytic through the removal of its integrase gene sequence, and the host range of this engineered phage (d700) and its ancestral strain (i700) was assessed. d700 was

able to lyse one strain of UMB *E. coli* that i700 was not, suggesting that i700 infected and lysogenized the bacterium.

This study shows that there is substantial diversity in UMB *E. coli* prophages and shows that some temperate coliphages can be induced, engineered to be obligately lytic, and lyse non-native hosts.

CHAPTER ONE

INTRODUCTION

The Human Urobiome

In 2010, the healthy human bladder was shown to contain microbial DNA, challenging the dogma that the urine of healthy individuals is sterile (Thomas-White et al., 2016a). The urinary tract contains the kidneys, ureters, bladder, and urethra. The discovery of the Human Urinary Microbiome (UMB), the microbiome found within the human urinary tract, was first made via the identification of ‘uncultivated bacteria’ by DNA sequencing of the highly conserved 16S rRNA genes within voided human urine (Nelson et al., 2010; Siddiqui et al., 2011). Prior to the development and use of 16S sequencing to identify constituent bacteria, there was thought to be no human UMB in healthy individuals because clinical microbiology laboratory culture cultivation of non-symptomatic subject urine historically returned negative results (Dong et al., 2011; Hilt et al., 2014). Instead, a positive clinical laboratory culture was used to diagnose disorders, most commonly urinary tract infections (UTI) (Long and Koyfman, 2018). Therefore, historically in the medical literature, presence of bacteria in the urine accompanied by symptoms was indicative of a disorder (Long and Koyfman, 2018).

While culture negative results were consistently replicated for healthy individuals, the protocols used in clinical practice were optimized during the 1950s for known uropathogen growth conditions in the clinical laboratory setting (Thomas-White et al., 2016a). These laboratory conditions differ from the environment of the human bladder and thus the diverse

range of genera that naturally inhabit the human bladder may not be able to thrive in a clinical laboratory setting (Wolfe et al., 2012; Hilt et al., 2014). By using a larger range of media, incubation temperatures, incubation lengths, and oxygen availability, bacteria was shown to be cultured from the catheterized and voided urine specimens of healthy individuals (Hilt et al., 2014). The variation of culture conditions shown by (Hilt et al., 2014) allowed for the cultivation of 35 different genera of bacteria from the bladder, not just the well-studied uropathogens. Since these initial findings, further developments in urinary microbe cultivation techniques have allowed for the development of Enhanced Quantitative Urine Culture (EQUC) methods that specifically improve the detection of uropathogens (Price et al., 2016a).

Investigations aimed at characterizing constituents of the UMB have been undertaken since the discovery of a UMB in healthy individuals. As in other human microbiomes (Turnbaugh et al., 2007), the diversity of the UMB is highly variable from individual to individual (Nelson et al., 2010; Siddiqui et al., 2011; Wolfe et al., 2012; Thomas-White et al., 2018). It should also be noted that human host genetics have been shown to contribute to the composition and dynamics of the human urinary microbiome (Adebayo et al., 2020). These results suggest that the differences in host genetics may be contributing to the different urinary microbial communities observed between individuals. Furthermore, there are data suggesting significant differences between the UMBs of females of different age groups/menopausal status (Ackerman and Chai, 2019). Similarly, the healthy male UMB was also shown to be both diverse and highly variable (Nelson et al., 2010; Dong et al., 2011). In recent years, it has also been shown that the microbiome of the lower urinary tract of females is also temporally dynamic,

being influenced specifically by menstruation cycles and vaginal intercourse (Price et al., 2020b).

The bacterial species in the UMBs of female individuals presenting with lower urinary tract or bladder symptoms have been shown to be significantly different than healthy individual UMBs (Pearce et al., 2014; Shoskes et al., 2016; Thomas-White et al., 2016a; Wu et al., 2017). Therefore, extensive work has been done assessing the role of the human UMB in causing, contributing to, and being correlated with urinary tract and bladder disorders.

The human female UMB has been shown to contribute to the overall health of the bladder and urinary tract. The female UMB has been shown to differ between asymptomatic patients, overactive bladder (OAB) patients (Pearce et al., 2014; Thomas-White et al., 2016b), and some patients with UTIs (Pearce et al., 2015). A subdivision of OAB is urinary incontinence (UI), the symptoms of which have been shown to be negatively correlated with the frequency of *Lactobacilli*, a common constituent genus within asymptomatic female UMBs (Pearce et al., 2014; Brubaker and Wolfe, 2017). Concurrently, diversity, meaning lack of a dominantly abundant species, is also associated with patients with OAB symptoms in contrast to those without OAB symptoms (Pearce et al., 2014; Ackerman and Chai, 2019). Work pertaining to the discovery of associations between differing levels of UMB diversity and OAB symptomatic presentation is ongoing. In contrast, the association between UTI and a microbial cause – specifically *Escherichia coli* – has been well-established in clinical practice.

Escherichia coli

E. coli is a rod-shaped, gram-negative bacterium from the family Enterobacteriaceae. Strains of *E. coli* are primarily harmless constituents of the human digestive tract, but, in relation

to human health, can be categorized into three groups; commensal strains, intestinal pathogenic strains, and extraintestinal pathogenic strains (Russo and Johnson, 2000). *E. coli* has been found in the UMBs of both females with urinary disorders and healthy controls (Thomas-White et al., 2016a; Garreto et al., 2020; Neugent et al., 2020; Price et al., 2020a). Nevertheless, *E. coli* is among the most implicated etiological agent of UTIs, contributing to both complicated and uncomplicated cases (Subashchandrabose and Mobley, 2015). In the US, around 60% of females will be diagnosed with a UTI in their lifetime (Klein and Hultgren, 2020). There are multiple designations of UTI, complicated and uncomplicated as well as acute and recurrent (for a review of UTI medical terminology see (Geerlings, 2016)). Of the acute UTIs diagnosed, around 80% are attributed to presence of *E. coli*, specifically Uropathogenic *E. coli* (UPEC) strains (Klein and Hultgren, 2020).

UPEC strains are among the most studied bacteria implicated in UTIs (Terlizzi et al., 2017). Despite rigorous investigation, few universal features that enable *E. coli* infection of the urinary tract uropathogenesis have been documented (Schreiber et al., 2017). What is known about UPEC strains is that pathogenesis is related to the bacterial expression of a wide range of genes encoding virulence factors, termed putative urovirulence factors (PUFs) (Schreiber et al., 2017; Klein and Hultgren, 2020). While most PUFs are not characterized enough to suggest a direct role in pathogenicity, they are useful in characterization and prediction of virulence of UPEC strains (Schreiber et al., 2017). However, examination of *E. coli* genomic assemblies of isolates from healthy females and from females with UTI symptoms found no discernable difference in the virulence factors encoded (Garreto et al., 2019). These findings suggest that

there are other factors beyond simply *E. coli* presence in an individual's UMB that contribute to symptom presentation (Garreto et al., 2019; Perez-Carrasco et al., 2021).

Along with virulence, the UPEC strains are also documented to be capable of developing antibiotic resistance, particularly resistance to fluoroquinolones and SMZ-TMP, two of the most frequently prescribed UTI treatments globally (Kumar and Das, 2017; Terlizzi et al., 2017; Waller et al., 2018; Kot, 2019). Given the increasing threat of antibiotic resistance, an alternative treatment strategy has gained renewed interest – bacteriophages, or viruses that infect bacteria. In other human microbiomes, bacteriophages (phages) have been shown effective in combating antibiotic resistant bacterial infections (Kortright et al., 2019; Zalewska-Piątek and Piątek, 2020).

Phages - Overview

In addition to the many genera of bacteria in the UMB, the human UMB comprises an even greater diversity of the bacteria's viral counterparts, phages. Phages are the most numerous biological entities on Earth, numbering near 10^{31} at any given point in time (Hendrix et al., 1999). Despite being only tens to hundreds of nanometers in length, when stretched end to end, the bacteriophages on earth would span farther than the nearest 60 galaxies (Suttle, 2007). Throughout nature, phages seem to be ubiquitous, and observed on all examined human body sites (Navarro and Muniesa, 2017; Batinovic et al., 2019). Phages, when observed in the external environment, are composed of a protein coat that encases a genome (Kakasis and Panitsa, 2019). Phages have genomes that average between 30 and 50 kbp in size but have been shown to range from 2 kbp up to 735 kbp (Hatfull, 2008; Al-Shayeb et al., 2020). There is extensive phage diversity, but most phages isolated and studied in laboratory settings are tailed dsDNA phages (order: *Caudovirales*). Other phages include ssDNA phages that do not have tails (e.g., family:

Microviridae) or filamentous phages (e.g., family: *Inoviridae*), as well as dsRNA and ssRNA phages (Ackermann, 2007).

Phages are obligate intracellular parasites that require a susceptible bacterial host cell to multiply and proliferate. The infection of a host cell is facilitated by recognition of a host cell receptor via phage receptor binding proteins. The bacterial cell surface proteins targeted by phages are extensively diverse. Targeted host receptors include protein receptors (OmpA and OmpC), lipopolysaccharide (LPS) receptors, receptors located in capsular polysaccharides (Vi-antigen), pili, and flagella (Stone et al., 2019). Because of the extensive diversity in cell receptor targeting, phage-bacteria relationships tend to be highly specific, lending phages a relatively narrow viral host range. Host range is defined as the diversity in hosts that can be infected by a phage. Host ranges for phages are usually limited to select strains of a single species but some phages have exhibited the ability to infect hosts of multiple genera (Hyman and Abedon, 2010a). Evolutionarily, phage and bacteria are involved in a constant arms-race; phages are selected to infect bacteria more efficiently, while bacteria are selected to resist phage infection. The result is extensive diversity of both phages and bacteria, contributing to phages earning the title of the most diverse biological entities on the planet (Grose and Casjens, 2014).

Phages follow two main lifecycles: the lytic lifecycle and the lysogenic lifecycle. Phages are separated into multiple categories depending on their lifecycle(s); for a review of phage terminology, see (Hobbs and Abedon, 2016). Here, three main categories of phages will be discussed: obligately lytic phages, prophages, and temperate phages. Obligately lytic phages, as the name suggests, persist exclusively through the lytic lifecycle (phage life cycles are visualized in **Figure 1**). The lytic lifecycle is characterized as phage infection of a host followed by phage

replication via the sequestering of the bacterial host's replication machinery, followed by host death and release of phage progeny into the external environment (Stone et al., 2019). The lysogenic life cycle is different than the lytic life cycle in that upon successful infection, the phage genome integrates into a host's chromosome or become an extrachromosomal plasmid instead of rapid phage progeny replication flowed by lysis. A prophage is a phage genome existing within a bacterial genome or plasmid during the lysogenic cycle that may or may not possess the ability to exist in a form external to the bacterial host. Temperate phages are a hybrid of the two previous designations in that they can participate in both the lytic and lysogenic lifecycles (i.e. both in the external environment as well as within a host bacterium's genome/extrachromosomal plasmid). Thus, temperate phages can both integrate its genomic content into a host bacterium's genome (lysogenic life cycle) and then extricate itself through a process called phage induction (lytic life cycle) (Little, 2014). Therefore, temperate phages can be described as prophages that have the ability to excise from a host's genome and enter the lytic lifecycle (Hobbs and Abedon, 2016).

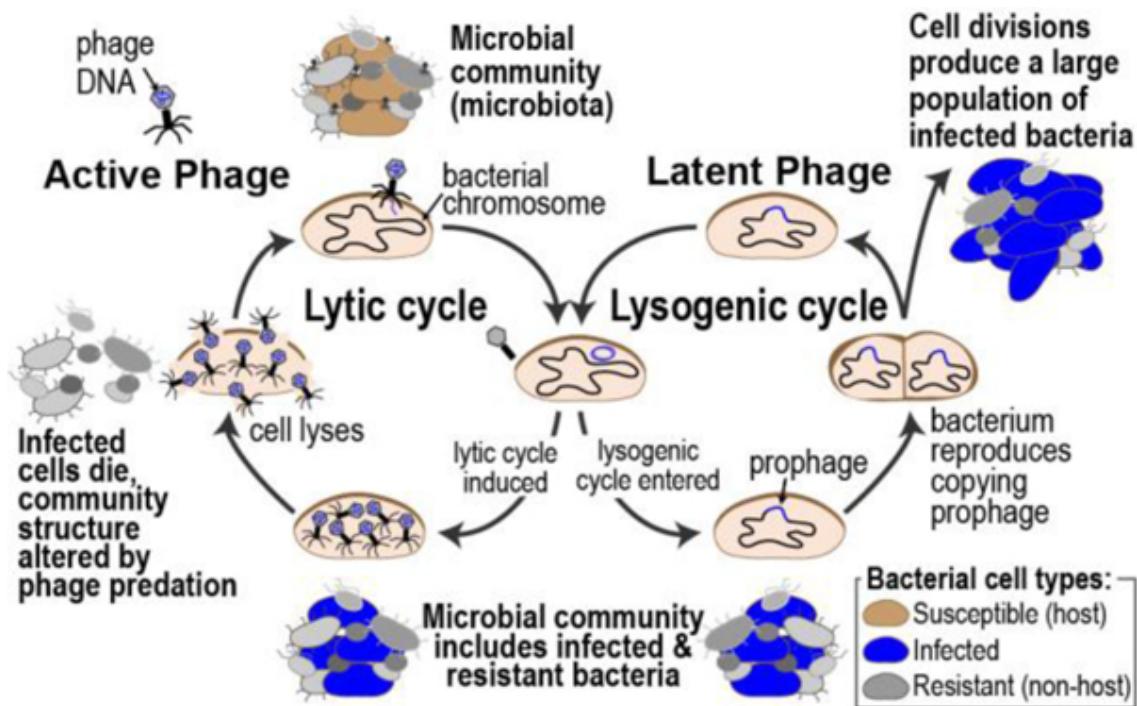


Figure 1. Phage lytic and lysogenic lifecycles. Infection begins by phage attachment to a host cell and insertion of its genome into the cell. The lytic cycle is characterized by immediate phage replication using host cell machinery followed by bursting of the host cell and the release of phage progeny into the surrounding environment, thus killing the host. The lysogenic cycle is characterized by insertion of the phage's DNA genome into the bacterial host's genome. As the host replicates, the phage is replicated as part of the bacteria's own genome. Temperate phages can switch between these two life cycles. Image obtained from (Garreto et al., 2019).

Temperate phage integration is most often accomplished through site-specific recombination mediated by phage-encoded integrase proteins. There are 2 evolutionarily distinct families of site-specific integrase proteins: tyrosine and serine integrases. This study will focus on tyrosine integrases as they are the family of integrases most predominant in *E. coli* phages (coliphages) (Fogg et al., 2014). A hallmark representative of the tyrosine integrase family is the integrase protein of the temperate coliphage λ (Fogg et al., 2014; Nakamura et al., 2021).

λ integrates via a unidirectional, highly site-specific recombination reaction. The recombination reaction is catalyzed by the phage integrase protein, Int, and occurs between homologous stretches of phage (*attP*) and host (*attB*) DNA sequences. The phage attachment site

(*attP*) and the bacterial attachment site (*attB*) are split in half and the phage sequence is inserted into the bacterial genome through site-specific recombination (for review of site-specific recombination mechanisms see (Grindley et al., 2006)). The resulting inserted phage sequence is flanked by new hybrid *attL* and *attR* sites, each composed of half the original *attP* and *attB* regions. The integration process is visualized in **Figure 2** (Fogg et al., 2014).

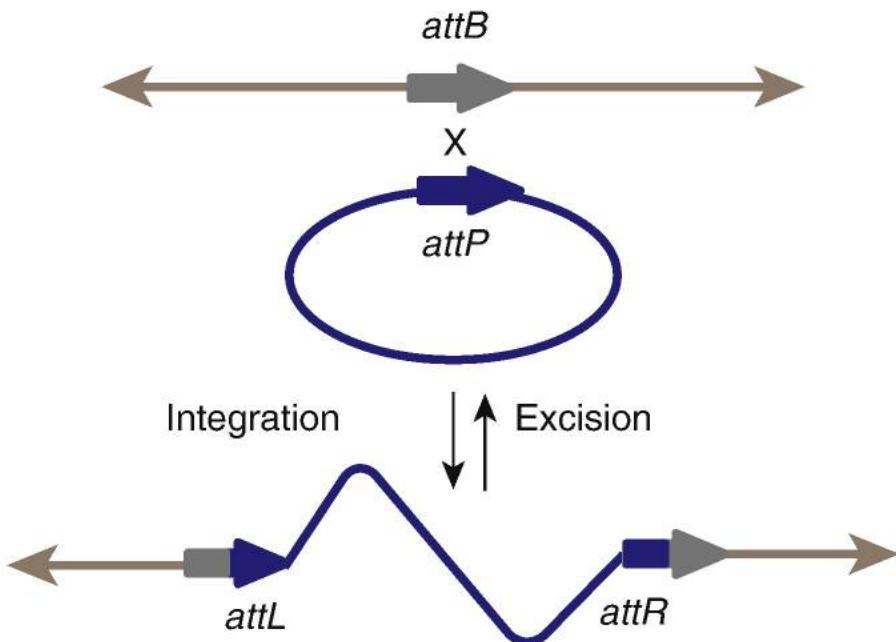


Figure 2. Temperate phage integration and excision diagram. The purple lines represent a temperate phage DNA genome. The gray lines represent a bacterial host's genome. The processes of temperate phage integration and induction occur at the *attP* and *attB* phage and bacterial DNA regions respectively. The integration process is catalyzed by phage integrase proteins. The induction process is essentially the reverse reaction of integration and is catalyzed by phage excisionase or recombination directionality factor proteins. Image obtained from (Fogg et al., 2014).

While integrated, the lysogenic state of λ is maintained by the phage *cI* repressor protein. Its bound state suppresses the transcription of lytic promoting genes integrase (Int) and excisionase (Xis) (Ball and Johnson, 1991). *cI* repressor protein is regulated by other phage and bacterial proteins (such as *cII*, *cIII*, Cro, and Hfl). The regulation of *cI* relies upon cooperative

gene regulation and its understanding was foundational in elucidating the mechanisms of cooperative gene regulation (Dodd et al., 2005). Despite a deep understanding of the mechanistic regulation process of *cI* repressor, the underlying mechanisms that trigger such regulation events in other phages are yet to be exhaustively studied. Current knowledge provides that specific environmental conditions, particularly those that cause transcription of the host DNA-repair protein RecA, result in autoproteolysis of *cI*, transcription of lytic genes (Int and Xis), and entrance into the lytic cycle. Thus, most commonly, severe host cellular distress causes phage induction. The other known primary cause of induction is random genetic mutation in the *cI* gene, but this is thought to be statistically very unlikely (Dodd et al., 2005; Hochschild and Lewis, 2009).

The excision process of the temperate *coli* λ (discussed above as prophage induction) requires two phage and two host transcribed proteins. The phage proteins required are the integrase, Int, and the excisionase protein, Xis. The required host are the nucleoid associated proteins, integration host factor (IHF) and Fis (Ball and Johnson, 1991). The excision reaction follows a mechanism that is the reverse of the integration reaction: *attL* and *attR* recombine into *attB* and *attP*, returning the chromosome to the non-prophage form and forming the circular phage genome that then enters the lytic life cycle (Fogg et al., 2014).

Studies of the human UMB have revealed a diverse range of resident phages (Miller-Ensminger et al., 2018a); prophages significantly outnumber obligately lytic phages (Malki et al., 2016a; Garretto et al., 2018). It also has been shown that *E. coli* strains isolated from patients experiencing the symptoms of urinary disorders often contain prophages (Price et al., 2016b; Miller-Ensminger et al., 2018a). Prophage sequences within *E. coli* of the human UMB have

been predicted using computational methods, the details of which will be talked about in-depth in Chapter 2. The investigation of temperate phages also has been undertaken on several different UMB bacterial species. Temperate phages have been experimentally induced from *E. coli* (Dallas and Kingsbery, 1997; Malki et al., 2016b; Miller-Ensminger et al., 2018a), *Pseudomonas aeruginosa* (Brown-Jaque et al., 2016; Johnson and Putonti, 2019; Johnson et al., 2019), and *Streptococcus anginosus* (Brassil et al., 2020).

Of the studies that examined UMB *E. coli* prophage induction, two studies by our group have experimentally induced and characterized urinary coliphages (Malki et al., 2016b; Miller-Ensminger et al., 2018a). In a study by (Miller-Ensminger et al., 2018a), temperate phages were computationally identified in four UMB *E. coli* strains. One of the predicted prophages was induced and characterized via host range testing and transmission electron microscopy (TEM) imaging. The host range tests revealed that the induced phage was able to lyse laboratory strains of *E. coli* (e.g., *E. coli* K-12, B, and/or C) but unable to lyse other UMB *E. coli* strains. The highly specialized lytic potential of this induced phage suggests that some phages of the UMB are only capable of infecting and lysing their native host. However, not all induced UMB phages have shown lytic host ranges limited to only their native host. In a 2016 study by Malki et al., six coliphages were isolated from the urinary microbiota of females with urinary urge incontinence (UUI) (Malki et al., 2016b). Two of those phages were tested for their ability to lyse non-native host urinary *E. coli* isolates, including UPEC strains (Malki et al., 2016b; Putonti et al., 2017). Notably, both phages exhibited the ability to lyse some UPEC and non-pathogenic UMB *E. coli* strains, but neither phage was able to lyse all tested strains (Putonti et al., 2017). These results

suggest that temperate phages with broad host ranges could affect the *in vivo* population dynamics (lysis) and phenotypes (lysogeny) of multiple strains of UMB *E. coli*.

Scope of Thesis

Despite extensive documentation of laboratory coliphages, such as the temperate phage λ , the diversity of coliphages in the human UMB, specifically temperate coliphages, has not been catalogued. While previous studies have identified and characterized the lytic potential of a few temperate UMB coliphages (Malki et al., 2016b; Miller-Ensminger et al., 2018a), their ability to infect and integrate into the host is unknown. More broadly, the lysogenic potential of UMB phages has not been explored.

Here, I present a study of *E. coli*-infecting temperate phages, integrating bioinformatic and synthetic biological approaches. Chapter 2 presents a survey of urinary *E. coli* prophage diversity through analysis of all publicly available urinary *E. coli* genomes. Chapter 3 shows the induction, characterization, and genetic engineering of a few of the predicted prophages from the urinary *E. coli* sequences analyzed in Chapter 2. Chapter 4 offers conclusions and future directions of this thesis work.

CHAPTER 2

BIOINFORMATIC IDENTIFICATION, COMPARISON, AND CHARACTERIZATION OF UROBIOME COLIPHAGES

Introduction

Prophages can account for up to 20% of the host bacterium's genome and nearly half of all sequenced bacterial genomes contained at least 1 prophage (Touchon et al., 2016). Prophage presence within a bacterial genome can offer the host selective advantage and offer increased individual fitness (Little, 2014). The increase in fitness is due to lysogenic conversion, the expression of genes from the prophage genome that can change the phenotype of the bacterial host (Little, 2014). Examples of lysogenic conversion include extracellular toxins, adhesins to assist in bacterial attachment, effector proteins to assist in bacterial invasion, virulence enzymes, and antibiotic resistance genes (Brüssow et al., 2004; Wendling et al., 2021). Prophages also can confer resistance to their bacterial host from other phage infections, also known as superinfection immunity. Thus, prophages can significantly impact a bacterial host while contributing to greater phenotypic and genotypic bacterial diversity and fitness.

As discussed in Chapter 1, the genes contained in prophages are extremely diverse and specific to the bacterial host they target. Unlike cellular organisms, there is no gene conserved in all or even most phage genomes. For example, the sequences of enterobacterial phages λ , P2, P22, HK97, Mu, and T7 are not recognizably homologous despite these phages exhibiting indistinguishable morphology under an electron microscope and being able to infect hosts within

the same bacterial family (Casjens, 2003). Despite the extensive sequence diversity in bacteriophage sequences, there are some phage proteins that are more highly conserved than others. Genes that encode the large subunit of the terminase, the portal protein, head maturation protease, coat protein, and the tail shaft and fiber proteins (in tailed phages) are more conserved across phage groups than other phage proteins (Casjens, 2003). Specific to temperate phages, site-specific recombinase proteins, known as integrases, that are required for bacterial genome or plasmid integration are also conserved (Groth and Calos, 2004). These are just a few of the genes specific to prophages that can be searched for and used to identify prophages within bacterial genomes.

A subcategory of prophages, known as cryptic prophages, also have been found within bacterial genomes, and notably have been linked to bacterial antibiotic resistant phenotypes (Wang and Wood, 2016). Cryptic prophages are prophages that still possess multiple phage genes but are no longer functional (meaning they cannot be induced) and are in a state of mutational decay (Casjens, 2003). Along with cryptic prophages, satellite phages are parasitic phages that do not contain their own structural protein genes but rely on the structural genes of other specific phages for replication (Casjens, 2003). Cryptic prophages, satellite phages, phage tail protein-like bacteriocins, and bacterial gene transfer agents that resemble phage tail proteins all present a challenge for computational identification of functional prophages because they introduce of non-phage noise and increase in divergent gene presence and structure.

Because prophages can be highly influential to a bacterial host's phenotype, especially if they confer increased virulence or antibiotic resistance, it is imperative to be able to identify and characterize prophages within bacterial genomes. Over the past decade, dozens of computational

tools have been developed for identifying viral elements within bacterial genomes. While the tools vary in their specific design, the general strategy currently used for identifying prophages within bacterial genomes is to use known phage proteins and sequence similarity-based searches (Casjens, 2003; Arndt et al., 2016). While this is an imperfect solution, the most widely-used current computational prophage-finding tools, e.g. PHASTER (Arndt et al., 2016), VirSorter2 (Guo et al., 2021a), PhiSpy (Akhter et al., 2012), Phigaro (Starikova et al., 2020a), DBSCAN-SWA (Gan et al., 2022), Prophage Hunter (Song et al., 2019), primarily rely on this strategy. The sequence similarity-based searches used by prophage finding tools often rely on the NCBI Reference Sequence (RefSeq) database (<https://www.ncbi.nlm.nih.gov/refseq/>). The RefSeq viral collection, as of May 2022, contained nearly 10,000 viral genomes and over 104,000 viral proteins (O’Leary et al., 2016). As more bacterial and phage genomes are sequenced and annotated, the tools become more equipped to find prophages and improve the confidence of their results.

Prophage finding tools differ in how they utilize the results from similarity-based searching. Two of the most common approaches include (*i*) random forest machine learning classifiers followed by *att* site identification and (*ii*) assessment of number of consecutive phage proteins, specifically the semi-conserved structural genes discussed previously. VirSorter2 and PHASTER, the two most widely used tools in current studies, employ the first and second approach, respectively. VirSorter2, the next generation of its predecessor VirSorter (Roux et al., 2015), utilizes a random forest machine learning classifier that varies the parameters of its classifier based on genetic characteristics of different viral families (Guo et al., 2021b). In contrast, PHASTER predicts prophages based upon the number of consecutive phage proteins in

a genomic region and then categorizes that region as ‘not prophage’, ‘incomplete’, ‘questionable’, or ‘intact’, with intact being the highest confidence prophage output (Arndt et al., 2016).

Both tools have been used to predict prophages in microbiota from a wide range of environments and genera. For instance, VirSorter2 was used to identify prophages of the gut microbiome (Bikel et al., 2022), soil microbiome (Sokol et al., 2022), and sediment microbiome (Zhou et al., 2021). Other past work has also used the original release of VirSorter to identify UMB prophages in 51 UMB bacterial genera (Miller-Ensminger et al., 2018b). More recently, PHASTER was used to identify a novel temperate phage Lu-1 in a UMB strain of *Lactobacillus jensenii* (Miller-Ensminger et al., 2020c). In just the past of couple months, PHASTER has been used to predict prophages in studies ranging from vaginal health probiotics analysis (Happel et al., 2022) to surveys of *K. pneumonia* (De Koster et al., 2022) and *E. coli* and *P. aeruginosa* (Giannattasio-Ferraz et al., 2022) from human and bovine samples.

PHASTER is routinely used when benchmarking new prophage finding tools (Song et al., 2019; Starikova et al., 2020b) because it consistently performs among the best of the tools currently available. It is aptly suited to identify prophages within well-characterized species such as *E. coli*. In a 2020 study on Shiga toxin-producing *E. coli* strains (Zhang et al., 2020), PHASTER was used to identify prophages that are known to be involved in the transfer of Shiga toxin-producing virulence genes. The study found 22 prophages, with one or more containing Shiga toxin-producing virulence genes, in each of the 8 Shiga toxin-producing *E. coli* strains analyzed. Additionally, phylogenetic analysis revealed high genetic diversity between the prophages identified, further emphasizing the efficacy of PHASTER’s ability to predict distinct coliphages.

Some of the most well characterized temperate phages have come from *E. coli*. The phage λ (discussed in Chapter 1) is a model that helped elucidate the mechanisms of genetic and

molecular control over gene expression. Consequently, λ has been among the most studied and well-understood bacteriophages. After the characterization of λ , phages subsequently discovered that resemble λ genetically or follow the lysogenic lifecycle have been referred to as ‘lambdoid’ phages (Casjens and Hendrix, 2015). Over the past decade, 80 genetically similar relatives of λ have been isolated and sequenced and hundreds have been identified as prophages integrated into bacterial genomes. Despite the fact that the phages above are referred to as ‘lambdoid’, extensive sequence and genomic structure variation has been observed (Casjens and Hendrix, 2015).

While much is known about the model coliphages mentioned, recent studies have shifted their emphasis towards the diversity and analogy between phages found in *E. coli* of different natural habitats, with particular focus on the implications of prophages acting as vectors of virulence and antibiotic resistance. One area of *E. coli* prophage diversity research is within Shiga toxin-producing *E. coli*. Shiga toxin-related coliphage diversity has been documented (Rodríguez-Rubio et al., 2021). One particular study by Zhang et al., 2020 looked at the diversity of Shiga toxin-producing *E. coli* in 40 strains including the serotypes O26, O45, O103, O111, O121, O145, and O157. The group found 54 Shiga toxin-containing prophages using PHASTER. It was also shown that induction of Shiga toxin-containing prophages could be triggered in most *E. coli* strains via treatment with EDTA or UV light (Zhang et al., 2020).

With the increase in the availability of bioinformatic tools for prophage prediction, investigations into the diversity of prophages within specific bacterial genera and species have been extensively undertaken in recent years. Importantly, despite the increase in phage diversity studies in the human gut (Manrique et al., 2016) and oral microbiomes (Szafranśki et al., 2021), the phage diversity of the human urinary microbiome remains largely uncharacterized. To date,

the general diversity of prophages in the human UMB has been assessed by two studies focusing on the urinary microbiome and lower urinary tract (Miller-Ensminger et al., 2018a; Garreto et al., 2019). VirSorter revealed that 86% of the urinary bacterial genomes analyzed contained at least 1 prophage sequence, suggesting that lysogeny is widespread in the human urinary tract (Miller-Ensminger et al., 2018a). Additionally, it has been shown that different individual's UMBs can harbor related prophages (Brassil et al., 2020; Miller-Ensminger et al., 2020c, 2020c). Furthermore, similar and related prophages have been found in the UMBs of humans and other animals (Giannattasio-Ferraz et al., 2022).

Here, we present a comprehensive catalog of the *E. coli* infecting prophages of the UMB. This chapter investigates the diversity of UMB *E. coli* prophages. Using PHASTER and all publicly available urinary *E. coli* genomic assemblies, we identify the genetic similarities between predicted UMB *E. coli* prophages. We also assess predicted prophage diversity through the identification of antibiotic resistance, virulence, and integrase genes.

Methods

Urinary *E. coli* Genomes.

All publicly available *E. coli* complete and draft genomes in NCBI that were documented as being collected from urine or the urinary tract (in the genome metadata) as of February 2021 were downloaded. 906 genomes meeting this criterion were obtained from NCBI; 2 sequences were removed due to quality concerns, as both genomes contained >300 scaffolds. Thus, the final set includes 904 *E. coli* genome assemblies (**Supplemental Table 1**). Additionally, 57 unpublished, recently assembled *E. coli* genome assemblies, from transurethral catheterized urine samples, were obtained from the Wolfe Lab (Loyola University of Chicago, Maywood, IL). All files were in FASTA format. Thus, the complete set includes 961 genome assemblies.

E. coli Prophage Prediction.

Prophages were predicted using the online phage prediction tool PHASTER (<https://phaster.ca/>) (Arndt et al., 2016). A Python script was written to submit genome assemblies and retrieve results via the PHASTER API. Once the results were retrieved, the Python script separated the predicted prophage sequences based on their PHASTER prediction category: “intact,” “questionable,” and “incomplete” prophage sequences.

All intact prophage sequences were compared via local BLASTn queries to a database of all complete and partial genome sequences in GenBank as of February 2021 (Organism: “Virus” and Division: “PHG”). This database includes 26,381 sequences. BLAST hits with a query coverage greater than 50% and a percent identity greater than 70% were determined to be related phages and the associated taxonomies of the homologous database records were used to predict the taxonomies of the queried prophages.

Assessment of the prevalence of identical prophages within PHASTER-predicted prophages was performed with a Python script.

***E. coli* Prophage Network Construction.**

Using Anvi'o 6.2 (Eren et al., 2015), the intact prophage sequences were annotated, and a coliphage pan-genome 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 mcl-inflation 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 (Shapiro and Putonti, 2018a). 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: $w = (\# \text{ of shared genes between genomes 1 \& 2}) / \sqrt{\text{size genome 1} * \text{size genome 2}}$. By designating a minimum value of w (min_w) that allows for an edge to be drawn between two genomes only if $w > \text{min}_w$, the edges were filtered to construct networks of differing connectivity.

The edge-lists constructed from the edge-drawing Rscript were then visualized using Cytoscape 3.9.1 (Shannon et al., 2003). The prophage gene similarity network was constructed using a min_w of 0.3 ($\geq 30\%$ gene similarity for each edge).

Bridge Prophage Multiple Sequence Alignment

All bridge prophage connections within the phage network (nodes connected to bridge prophages via an edge) were identified for the 6 bridge prophages using a Python script. Identified bridge prophages and connection sequences were appended into a multifasta file using a Python script. The multifasta files were individually inputted into Mauve 2.4.0 (<https://darlinglab.org/mauve/mauve.html>) for alignment (Darling et al., 2004). The resulting alignments were interpreted visually.

Prophage Integrase Gene Annotation, Clustering, and Tree Construction.

The PATRIC online tool (<https://www.patricbrc.org/>) was used to annotate the predicted prophage genomes (Davis et al., 2019). The phage sequences were uploaded onto the web interface, the job was submitted using the following parameters: ‘Domain’: Bacteriophages, ‘Taxonomy Name’: *Escherichia coli*, ‘Genetic Code’: 11 (Archea and most Bacteria), and ‘Annotation Recipe’: Bacteriophage. The resulting .gb files were pulled.

All prophage annotated genes containing ‘integrase’ in the description were added to a multi-FASTA file and clustered using USEARCH 11.0.667 (<https://www.drive5.com/usearch/>)

(Edgar, 2010). Two clustering thresholds were tested, -id 0.9 and 0.7. Kalign 2.04 (<https://github.com/TimoLassmann/kalign>), with default parameters, was used to produce a multiple sequence alignment of all prophage integrase genes (Lassmann and Sonnhammer, 2005). Visual inspection of the multiple sequence alignment was done using Geneious Prime 2021.2.2 (Dotmatics, Auckland, NZ). Partial integrase gene sequences were identified and removed from the data set. The remaining integrase gene sequences were again aligned using the default parameters in Kalign. The resulting FASTA formatted multiple sequence alignment was inputted into FastTree 2.1.11 (<http://www.microbesonline.org/fasttree/>) using the -gtr and -nt settings to produce an approximately-maximum-likelihood phylogenetic tree (Price et al., 2010). The resulting Newick format tree file was visualized using iTOL 6.5.2 (<https://itol.embl.de/about.cgi>) (Letunic and Bork, 2021).

Prophage Antibiotic Resistance + Virulence Factor Identification.

Antibiotic Resistance genes within prophage genomes were predicted via the RGI 5.2.1 tool (<https://github.com/arpcard/rgi>) which utilizes the CARD Antibiotic database (Alcock et al., 2019). RGI was installed locally using conda (<https://docs.anaconda.com/>). The CARD database (Version 3.1.4) was downloaded locally. The RGI program was run using default parameters.

The full dataset of known bacterial virulence factor gene sequences (<http://www.mgc.ac.cn/VFs/download.htm>) was downloaded March 2021. Using the local BLASTn package within BLAST+ 2.9.0, each prophage sequence was queried to a BLAST database, made using the ‘makeblastdb’ command, from the dataset of known bacterial virulence factor gene sequences. The BLASTn was run with the parameter -evalue 0.001. Hits of <90% sequence identity and query length <90% of the virulence factor gene sequence were removed

from consideration. Homologous hits were filtered out using the qstart (query start) and qend (query end) values output by BLAST. If hits had qstart values within 500 bp of each other, only the hit with the highest percent identity value was kept. The virulence factor gene descriptions were determined using VFDB.

Results

Urinary *E. coli* Prophage Prediction.

Prophages were predicted for 961 urinary *E. coli* genome assemblies, including 128 genomes from Loyola's isolate collection. In total, PHASTER predicted 8,987 potential prophage sequences. This included 3,177 intact, 1,656 questionable, and 4,154 incomplete prophage sequences. Only prophage sequences from the highest confidence group, the intact prophages, were used in downstream analysis. Among the PHASTER-predicted intact prophages, there were 1,225 predicted prophages that actually represented 607 unique prophages because they were identical to at least one other prophage identified in another bacterial isolate. While 48 *E. coli* genomes did not contain a predicted intact prophage sequence, most genomes harbored more than one intact prophage (average of 3.30 predicted intact prophages per genome) which is visualized in **Figure 3**. Strains with intact prophage sequences included both genomes from public databases as well as genomes from Loyola's isolate collection. Furthermore, the same propahge sequence was found in different strain; the most prevalent intact prophage sequence was identified in 33 different *E. coli* sequences.

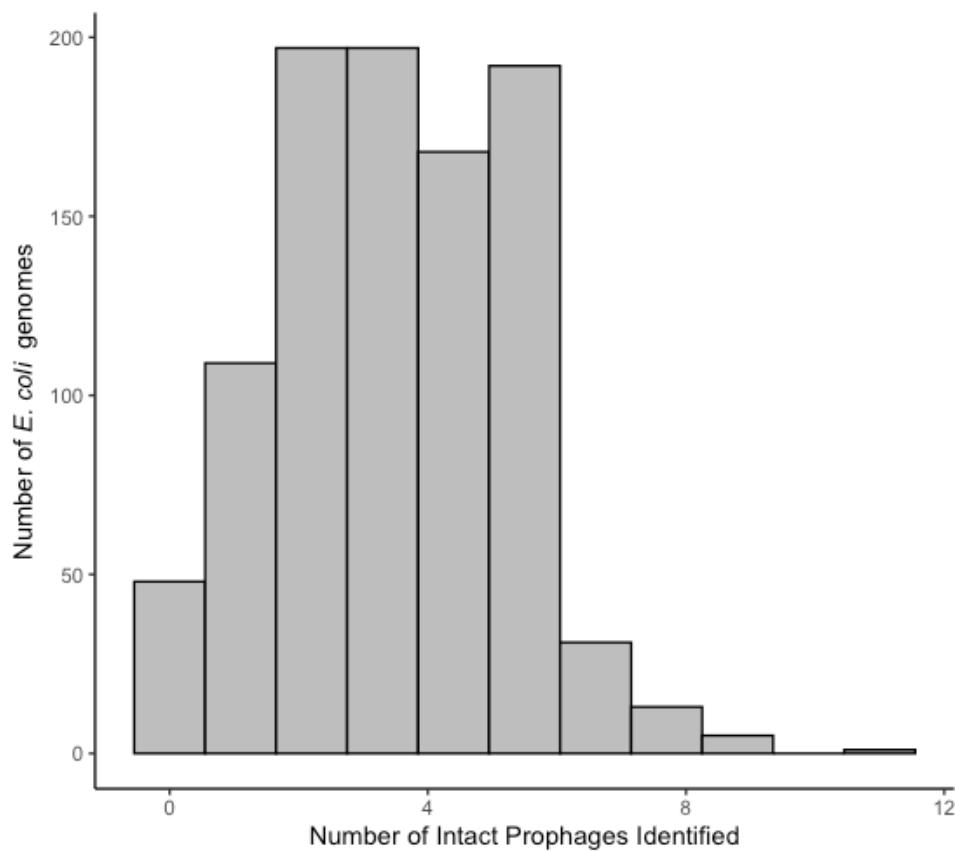


Figure 3. Histogram of prophage per genome distribution of intact prophages predicted by PHASTER among 961 urinary *E. coli* genome assemblies.

Prophage Antibiotic Resistance and Virulence Factor Gene Prevalence in Prophages.

187 of the 3,177 prophage sequences were found to contain antibiotic resistance genes and a total of 314 antibiotic resistance genes were identified. Of the 314 antibiotic resistance genes, 23 drug class resistances were represented, the most common being “aminoglycoside antibiotic; aminocoumarin antibiotic” resistance from 84 of the 314 genes identified. The other drug classes most represented were “macrolide antibiotic”, “fluoroquinolone antibiotic; monobactam; carbapenem; cephalosporin; glycylcycline; cephemycin; penam; tetracycline antibiotic; rifamycin antibiotic; phenicol antibiotic; triclosan; penem”, “aminocoumarin antibiotic”, and “aminoglycoside antibiotic”. The five drug classes listed here represent 74.5% of

the antibiotic resistance genes found in the prophage sequences. Of the individual antibiotic genes, *emrE* was the most frequently identified (n=50). Additionally, *marA*, *baeR*, *baeS*, and *mdtC* were common antibiotic resistance genes; collectively, these five genes accounted for 58.6% of all antibiotic genes found.

In total, 543 of the 3,177 prophage sequences contained known virulence factor genes. There was a total of 1,451 individual virulence factor genes identified, representing 211 unique virulence factor genes. The most common virulence factor gene was *aiiQ* with 137 occurrences. The other most represented virulence factor genes were *sitA* (iron ABC transporter substrate binding component), *YPK_3315* (invasion), *ipaH* (hypothetical prophage protein), *aaiR*, and *sitD* (*sitD* protein) which represent 27.3% of the virulence factor genes predicted.

Exploring the Diversity of Urinary *E. coli* Prophages.

To characterize the genomic diversity of the predicted intact prophage sequences, the content of each pair of prophage sequences was compared (see Methods) and visualized using a network (**Figure 4**). In this network, the nodes represent individual intact prophage sequences, the edges represent prophage genetic similarity, with the darkness of each edge representing the amount of similarity. The network uses a 30% gene similarity metric threshold to draw an edge between two nodes, meaning that for an edge to be drawn, the two prophages need to contain $\geq 30\%$ of the same genes. The 30% gene similarity threshold was selected to maximize the number of nodes (prophages) included in the network while reducing noise caused by low similarity edges between prophages.

The network contains 10 independent connected components. Of the 3,177 prophage genomes predicted by PHASTER, 3,164 are represented in **Figure 4** following application of the

threshold. 451,027 total edges are included in the network connecting the 3,164 prophages. The color of each node in **Figure 4** is representative of the prophage sequence's taxonomic family as determined by a local BLAST analysis against a database of taxonomically known phage sequences. In the network there were 1,402 "Unknown" taxa, 822 "Myoviridae", 763 "Siphoviridae", 120 "Podoviridae", 50 "Unclassified Caudovirales", and 20 "Unclassified Bacterial Virus". The "Unknown" group encompasses sequences that did not return any significant taxonomic hits. Overall, the network exhibits clustering of phage sequences that are from the same taxonomic family reflecting the general phylogenetic organization of the network.

Note: all mentions of lower-case letters in quotations (e.g., "n") are references to independent connected components labelled in **Figure 4**. The Unknown phage sequences also cluster together within independent connected component "a," they are also found in four of the smaller independent connected components ("b", "c", "e", and "f"). The clustering of these intact prophage sequences, especially those in separate independent connected components, suggests the potential presence of novel, unclassified viral taxonomies within urinary *E. coli* isolates.

The largest independent connected component labeled "a" contains 2,864 nodes. The large clusters of nodes generally follow taxonomic classifications and are connected to each other by "bridge prophages" labelled by red roman numerals in **Figure 4**. These bridge prophages represent the intact prophage sequences (I) GCF_002416865_4_NZ_NXIR0, (II) GCF_013372345_18_NZ_CP05, (III) GCF_000503275_1_NZ_AYQK0, (IV) GCF_000460075_6_NZ_KE701, (V) GCF_003886275_12_NZ_RRVO, and (VI) GCF_004116695_2_NZ_CP035. The bridge prophages labelled in the network represent prophage sequences that share gene content with two or more clusters and increase overall

connectivity of the network. Since edges require $\geq 30\%$ of the same genes, these bridge prophages exhibit genetic chimerism suggesting their evolutionary intermediate position between two groups. When bridge prophage sequences were aligned to their adjacent prophage sequences (nodes), it was found that the sequences contained similar modules (contiguous set of genes). These modules were usually $>1\text{kb}$ in size. Each of the bridge prophages contained one or more modules to prophage sequences of one subcluster and one or more modules to prophage sequences of the other subcluster.

Three different viral families as well as intact prophages of “Unknown” taxonomy are connected within the independent connected component “a”. The connectiveness of this large cluster confirms that there are conserved genes among tailed phages (*Caudovirales*; families Myoviridae, Podoviridae, and Siphoviridae). While not a perfect solution, the conserved genes that characterize this large cluster could be used as candidates to search for urinary coliphages. Moreover, “Unknown” prophages within this independent connected component can confidently be predicted to be tailed phages. The independent connected component “b” presents an interesting cluster of intact prophages that share $<30\%$ of genes with all other analyzed intact prophages. Additionally, the classification of all intact prophages within independent connected component “b” as belonging to “Unknown” taxonomies suggests that these may be novel, uncharacterized coliphages.

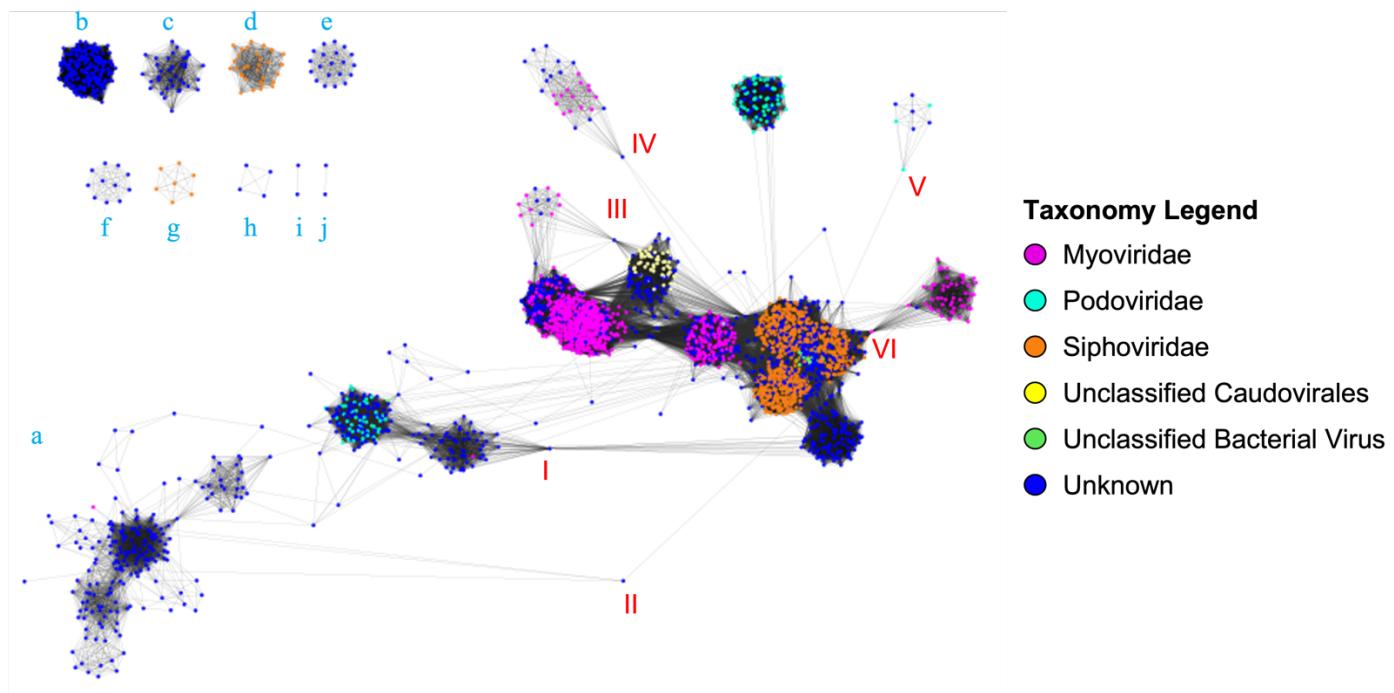


Figure 4. Prophage network produced from the percent similar genes found between intact prophage sequences predicted by PHASTER. Nodes represent intact prophage sequences and are colored based on similarity to known phage sequence taxonomies as determined by BLAST. Edges are representative of $\geq 30\%$ shared genes between two intact prophage sequences. Edge boldness is representative of gene similarity, darker indicates higher proportion of shared genes. Network includes 10 independent components labelled by teal lower-case letters. Within the largest independent connected component, connections between clusters by “bridge prophages” are labelled with red roman numerals I-VI.

Investigating the Diversity of Urinary *E. coli* Prophage Integrase Genes.

Annotation of the predicted prophage sequences identified 1,938 prophages encoding for integrase genes. In total, there were 2,186 integrase genes identified. The average nucleotide length of the integrase genes identified was 1,082 bp while the median nucleotide length was 1,130 bp. Clustering of the prophage integrase gene sequences with a 90% nucleotide sequence similarity threshold resulted in 106 unique integrase gene clusters. The largest cluster contained 130 integrase genes. Most of these clusters were small; 27 contained only one integrase gene and an additional 27 contained ≤ 5 integrase genes. After relaxation of this

threshold, stipulating only 70% nucleotide sequence similarity, all integrase genes formed a single cluster which was then aligned and manually inspected. 162 integrase gene sequences were identified as partial sequences and removed from further analyses. The resulting sequences were realigned, and a phylogenetic tree was derived (**Figure 5**).

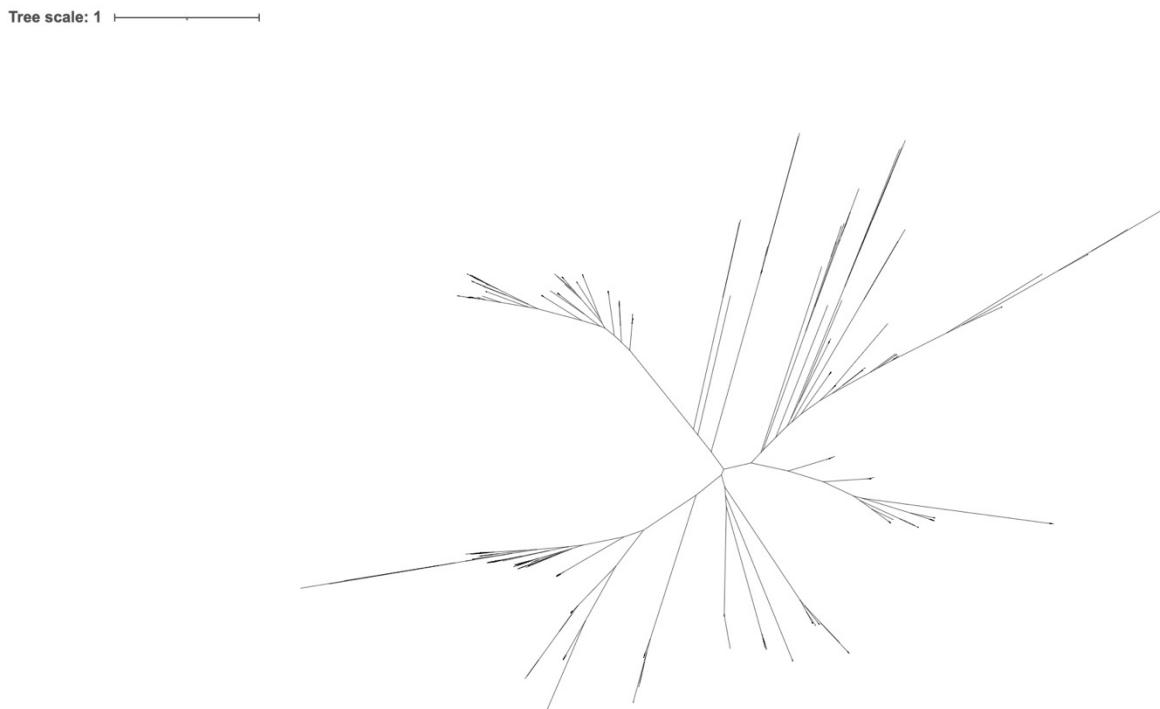


Figure 5. Maximum likelihood phylogenetic tree depicting the sequence similarity of integrase genes annotated in predicted intact urinary *E. coli* prophage sequences. Branch lengths represent evolutionary time.

Discussion

UMB *E. coli* Prophage Prediction.

While UMB prophages have been predicted and catalogued previously (Miller-Ensminger et al., 2018a), only four *E. coli* genomes were considered. This study focuses more specifically on the identification and classification of urinary *E. coli* prophages from

significantly more urinary *E. coli* sequences (n=961). The genetic diversity of *E. coli* strains has been previously documented in the urinary tract (Garreto, 2019; Garreto et al., 2020), and here we have identified prophages that may contribute to increasing this diversity. We found that prophages are ubiquitous in urinary *E. coli* strains; only 5% of UMB *E. coli* sequences contained no prophages. Furthermore, as was previously found (Miller-Ensminger et al., 2018a), urinary *E. coli* strains often contain more than one intact prophage. Here, we found that there was an average of >3 predicted prophages per urinary *E. coli* genome.

Although the first version of VirSorter2 was released at the time this study was initiated, we chose to use PHASTER, given its reliability and its prior success in identifying coliphages (Zhang et al., 2020). It should be noted that PHASTER has limitations as a prophage prediction tool. For instance, our study did not predict any intact prophages that were members of the viral family *Inoviridae* (inoviruses). Inoviruses are different from the more prolific tailed phages (of the viral families *Myoviridae*, *Siphoviridae*, and *Podoviridae*) in that they have a ssDNA genome, can follow a chronic life cycle, and are physically filamentous in form (for more information about inoviruses see (Roux et al., 2019)). Current computational approaches, including PHASTER, are not effective at identifying inoviruses (Roux et al., 2015, 2019). Previously, we have identified inoviruses in UMB *E. coli* (Shapiro and Putonti, 2020). Moreover, novel, previously uncharacterized phage taxonomies are difficult for computational tools to identify, as discussed in Chapter 1. It should also be noted that false positive predictions are likely as well. The prediction of ‘cryptic prophages’ that are no longer functional may be erroneously predicted as intact. Despite these limitations, PHASTER has proven to be a rigorously tested and applied tool for prophage identification and therefore provides an

inaccurate but useful glance at the diversity of currently identifiable prophages within urinary *E. coli*.

Antibiotic Resistance Gene Prevalence.

All predicted prophages were screened for antibiotic resistance and virulence factor genes because of the potential impacts these known, well characterized genes can have on host phenotype and human health. There is little work characterizing the prevalence of the antibiotic resistance genes of bacteriophages in the human UMB. Most of the work done in this field is in relation to wastewater and do not offer quantitative genome analysis of the coliphages studied. Rather, they state the presence of coliphages with antibiotic resistance genes found in wastewater and that they present a public health threat (Calero-Cáceres and Muniesa, 2016). It is not directly known if prophages can influence a host's antibiotic resistance phenotype but studies have reported that antimicrobial resistance genes are present in the phage or prophage region, which suggests prophage involvement in antibiotic resistance phenotypes (Kondo et al., 2021).

There is extensive work on the antibiotic resistance genes found in *E. coli*, especially within urinary pathogenic (UPEC) strains (Kot, 2019). It should be noted that since antibiotic resistance of urinary *E. coli* is a growing health concern, it is relevant to assess the antibiotic resistance gene prevalence within predicted urinary *E. coli* prophages. Interestingly, results suggest that prophages with at least one antibiotic resistance gene usually contained more than one antibiotic resistance gene. In the literature, the most common UPEC antibiotic resistance is to the fluoroquinolone antibiotic class (Kot, 2019). In the antibiotic genes identified, the most prevalent antibiotic drug class resistance targets were aminoglycosides, aminocoumarins, and fluoroquinolones. Combined, the three most targeted antibiotic drug classes accounted for 60%

of observed antibiotic resistance gene targets. In the literature, aminocoumarin and aminoglycoside antibiotics are reported to be commonly used in concert with fluoroquinolone antibiotics to treat complicated, severe urinary and kidney related bacterial infections, including UTIs (Heide, 2014; Goodlet et al., 2019). Thus, it is not surprising to find resistance to the aminocoumarin, aminoglycoside, and fluoroquinolone antibiotic drug classes in urinary coliphages. There is a low prevalence of predicted prophages containing antibiotic resistance genes (5.9%), but the identification of even this small proportion suggests that urinary coliphages could play a role in the phenotypic antibiotic resistance exhibited by some strains of urinary *E. coli*.

Virulence Factor Gene Prevalence.

Unlike antibiotic resistance, significant work has been done on the identification and characterization of virulence factor genes found in bacteriophages (Wagner and Waldor, 2002). Most characterization of virulence factors in coliphage has been done in relation to the known pathogenic *E. coli* O157:H7 strains. O157:H7 strains of *E. coli* possess 'lambdoid' prophages that house genes responsible for enterohemorrhagic Shiga toxin production. It has been found that the induction of virulence factor-possessing prophages, primarily in response to antibiotic treatment, can worsen symptom severity in patients (Łoś et al., 2011). Additionally, there has been work on the presence of virulence factors found in UPEC strains. The most common virulence factor in UPEC strains are 'fimbriae' which aid in cell surface adhesion, tissue adhesion, and biofilm formation (Shah et al., 2019). There has not been previous work assessing the virulence factor prevalence in UMB coliphages. We found that 543 (17%) of the predicted prophages contained at least one virulence factor gene. The most prevalent single gene *aaiQ*,

along with the less predicted but still prevalent gene *aaiR*, are predicted pseudogenes that have been linked to enteroaggregative *E. coli* pathogenesis (Dudley et al., 2006). These virulence factor genes are not fully understood but seem to be involved in the regulation of chromosomal virulence factor transcription (Dudley et al., 2006). As for genes that have known protein products, most belong to groups that function in iron uptake (*sit* and *ius* genes) and as adhesins (*fim* and *pap* genes) (Sarowska et al., 2019). Importantly, this analysis found that urinary coliphages can contain virulence factor genes. Since coliphages with virulence factor genes have been shown to impact the pathogenic phenotypes of their bacterial hosts, particularly with the Shiga toxin in pathogenic *E. coli* (Łoś et al., 2011), the fact that we found virulence genes in urinary coliphages that are common in UPEC suggests that coliphages could be contributing to *E. coli* virulence in the human UMB.

Urinary *E. coli* Prophage Diversity.

To assess the relatedness of phages, we used a gene co-occurrence network, an approach that has been used previously by our and other groups (Lima-Mendez et al., 2008; Shapiro and Putonti, 2018b). The visualization of prophage sequences based on the sharing of genes allows for the reflection of phage similarity despite the lack of a universally conserved set of core phage genes. Importantly, the approach described above allows for the inclusion of distant relationships within the network while capturing taxonomic structure (Lima-Mendez et al., 2008).

Our network shows that the largest independent connected component (component “a”), contains 90% of predicted prophages. This suggests a shared gene pool among urinary temperate phages. This could carry implications for the curation of a set of highly conserved coliphage genes that could be used for taxonomic prophage identification within urinary *E. coli* isolates.

From visual inspection of the network (**Figure 4**), we can see that prophages of the same predicted viral family are often more tightly connected, i.e., have more and/or darker edges connecting the nodes. The prophages that connect the different clusters of like taxonomies (bridge prophages) are of interest for their insight into phage evolution and phage genetic chimerism. Because phages are known to be mosaic in genomic make-up, exhibiting abrupt changes between homologous and analogous genomic segments in relation to other highly-related phages (Campbell, 2003), and experience recombination events frequently (Hendrix, 2002), the bridge prophages found between two taxonomically distinct prophage clusters could represent prophages that defy a traditional Linnean taxonomic classification. Much debate over phage taxonomic classification has been presented in recent decades (Nelson, 2004) and phage taxonomic classification in the current Linnean system is still limited (Turner et al., 2021). Thus, these bridge prophages present an interesting example of the limitations of the current taxonomic system. Further, these phages could even provide insight into the interesting evolutionary questions about the expansion or change in host range. Further analysis of these bridge prophages against prophages of other, closely related bacterial genera, such as *Salmonella*, could provide interesting information about phage evolutionary relationships between host genera and prophage genetic composition. Phages capable of infecting across urinary isolates of Proteobacteria have previously been identified by our group (Shapiro and Putonti, 2020).

Notably, there are also 9 smaller independent connected components (“b”–“j”). The most interesting of these genetically isolated prophage groups are the independent connected components “b”, “c”, “e” and “f”. These four independent connected components show high interconnectedness between prophages within these components and contain prophages that are

taxonomically “Unknown.” Because of their unique gene content and high interrelatedness, it is possible that these phages are novel species and should be explored in future studies.

Additionally, the unknown prophages in the main independent connected component could help expand current knowledge of known phage taxonomic groups because they were identified by PHASTER yet are still-to-be characterized in lytic forms. Given their connections with characterized prophages, i.e., number of connections, we could reasonably predict their taxonomy.

Integrase Gene Diversity.

Only 61% of the prophages examined contained an integrase gene. It should be noted that the integrase gene is typically located at one of the termini of the integrated prophage sequence. In prophage prediction, one of the most challenging computational problems is determining where the prophage genome begins and ends in relation to the host bacterium’s genome. Thus, it is likely that the integrase gene of an incomplete predicted prophage would not be present in its predicted genome. Additionally, integrase genes are known to be diverse between and within phage families (Balding et al., 2005), making sequence-similarity annotation approaches challenging and presents a bias that increases the chance of missing integrase genes in uncharacterized phages. Importantly, there is only one known family of integrases in *E. coli*-infecting temperate phages – tyrosine integrases (Fogg et al., 2014). Temperate phages from other bacterial species like *Staphylococcus aureus* have been shown to contain phages exhibiting integrases from multiple integrase gene families (Goerke et al., 2009). Since all known *E. coli* temperate phage integrases are from the same gene family, and we observed that these genes remain relatively conserved, temperate coliphage genes should be studied further for their

potential to be used in identifying, taxonomically classifying, and discerning evolutionary relationships between temperate UMB coliphages.

Studies done on the integrase diversity of *Salmonella enterica*-predicted prophages, an evolutionarily close relative of *E. coli*, have reported similarities to what we found here. Mainly, that the diversity of prophage integrase genes can be a good predictor of the diversity of entire prophage sequences (Colavecchio et al., 2017). It has also been shown that the integration sites of both coliphages and *S. enterica*-infecting phages are within conserved host intergenic regions (Bobay et al., 2013). These findings suggest there is likely homology between the integrase genes of coliphages and *S. enterica*-infecting phages and suggests evolutionary links between the phages of these two closely related bacterial hosts.

CHAPTER 3

UROBIOME COLIPHAGE INDUCTION AND HOST CHANGE CONFIRMATION

EXPERIMENTS

Introduction

As discussed previously, prophage prediction tools can predict the presence of prophages in bacterial genomes, but they are unable to determine if those prophages can enter the lytic life cycle. Thus, the best current method for determining if prophage sequences are temperate phages is experimental induction testing. The mechanisms that trigger prophage induction are still only partially understood and it is likely that, like phages themselves, there is significant diversity between and even among phage families in what triggers their induction (Bruce et al., 2021).

It has been shown that some prophage induction is dependent on signaling molecules. Work done in the bacteria *Bacillus subtilis* and its phage phi3T has shown that phages use signals to make informed decisions regarding prophage induction. These signals are produced by lysogens and over time those signaling molecules are degraded by hosts in a density-dependent manner. The declining signal molecule concentrations therefore potentially indicate the presence of uninfected hosts and trigger prophage induction (Bruce et al., 2021). In other circumstances, bacterial host cell health is regarded as the trigger of prophage induction as discussed by (Ball and Johnson, 1991).

In the laboratory setting, manipulation of a bacteria's growth environment is both a practical and reliable method for prophage induction. Methods such as using Mitomycin C (Raya

and H'bert, 2009), norfloxacin (McDonald et al., 2010), hydrogen peroxide concentrations (Łoś et al., 2010), pH (Miller-Ensminger et al., 2020a), and many others have been documented. Of the prophage induction methods listed, Mitomycin C is the most popular method used (Martín et al., 2006). Notably, Mitomycin C is a chemotherapeutic molecular agent secreted by soil fungi and thus is not naturally found within the human UMB (Raya and H'bert, 2009). In recent years, phage induction strategies have shifted towards conditions relevant to the human body to mimic conditions that may occur *in vivo* (Oh et al., 2019; Miller-Ensminger et al., 2020a).

In multiple studies presented thus far, it has been shown that many urinary prophages can enter the lytic life cycle and can, therefore, be classified as temperate phages. It has been shown that altering the pH of incubation media by biologically relevant amounts can cause the induction of temperate coliphages (Miller-Ensminger et al., 2020b). The pH method for induction was developed because it is a bladder-relevant stressor. For most healthy individuals, the pH of collected urine samples lies between 5 and 8 (Simerville et al., 2005), but is known to deviate for certain health conditions, such as a more acidic urine in diabetic patients (Maalouf et al., 2007). Fluctuations in the pH of urine likely impact the microbiota of the bladder and urinary tract and are therefore applicable for the cataloguing of UMB temperate coliphages.

Temperate phages, after being induced, enter the lytic cycle. The lytic cycle implies virion replication and host cell lysis. The lysis of the host cell causes the death of that cell and the release of phage progeny into the surrounding environment (Hatfull and Hendrix, 2011). Ideally, for the phage, another bacterium within its host range is accessible, allowing for infection and either lysogeny or lysis depending on the environmental circumstances (Hyman and Abedon, 2010b). This reinfection following lysis is impacted by a number of factors,

particularly those that govern lysogenic phage infection in the first place. This includes phage-host cell receptor compatibility, the absence or ineffectiveness of bacterial phage-defense mechanisms (such as superinfection protection or CRISPR-Cas systems), and a proper *attB* integration location within the bacterial host's genome or plasmid (Casjens and Hendrix, 2015). All this amounts to stating phage re-infection and integration are not simple or certain.

Understandably, phage host range is usually narrow because of the highly host-specific nature of many of the host defense evasion strategies employed by phages (Seed, 2015). Thus, characterization of phages based on their host range is useful as a means for describing phages based on phenotype. For obligately lytic phages, those that do not possess the ability to integrate into a host's genome, the test of host range is straightforward – does the phage lyse the host or not? For temperate phages, the testing of host range is more nuanced because of the possibility of lysogeny. With a lysogenic infection, direct phenotypic changes or outwardly observable lysis can be absent; therefore, providing inconclusive results regarding host range (Howard-Varona et al., 2017).

In an *in vivo* environment, phages often drive community dynamics through direct lysis (Santos et al., 2015), as well as horizontal gene transfer within a bacterial population (Borodovich et al., 2022). Recent studies have shown that phages with a broad host range, capable of infection and/or absorption by multiple species, can contribute to horizontal gene transfer between bacterial populations (Chen et al., 2015). This is particularly relevant in the context of temperate phages that contain genes that code for virulence factors, antibiotic genes, or other genes that could impact community dynamics via lysogenic conversion.

Testing the host range of temperate phages experimentally presents challenges that are mentioned above. To test for lysis by mixing induced phage lysate with a host and visually inspecting for lysis is one method, but it could be inconclusive due to the possibility of integration instead of lysis. If lysis is not observed, lysogeny could be tested for by using PCR primers that target the integration site of the temperate phage after induced phage and host incubation (Chen et al., 2020). However, this approach is challenging for phages and bacteria that are not well characterized or studied. Another approach involves the genetic engineering of an induced temperate phage, rendering that phage obligately lytic. This approach involves the identification and removal of the temperate phage's integrase gene via restriction enzyme (RE) cleavage. It is well documented and described in Chapter 1 that for coliphages, the integrase gene product is responsible for catalyzing the reaction that allows the temperate coliphage to integrate into a bacterial host's genome. Thus, by removing all or part of the integrase gene from a temperate phage's genome, the temperate phage can be forced to follow only the lytic life cycle after infection. This phenotypic conversion of the prophage allows for the host range to be assessed using the well documented methods used for obligately lytic phages without the uncertainty associated with the lysogenic cycle (Monteiro et al., 2019; Johnson et al., 2022).

The concept of phage engineering has been in circulation for decades, particularly since the first genomic modifications and genetic cloning experiments of the coliphage λ were presented in the 1970s (Casjens and Hendrix, 2015). Since then the genetic modification of viruses has exploded, providing advances from DNA libraries (Zabarovsky et al., 1993) to cancer treatments (Gutierrez-Guerrero et al., 2020).

Temperate bacteriophage genetic modification has become a hot topic in recent years because of its potential use in phage therapies (Monteiro et al., 2019). Phage therapy is the treatment of a patient that presents a serious bacterial infection exhibiting antibiotic resistance. Obligately lytic bacteriophages that are known be able to infect and lyse the bacteria presumably causing symptoms in the patient are used to treat the antibiotic resistant infection (Monteiro et al., 2019). This approach has been shown to be effective in multiple cases with different bacteria and antibiotic resistance phenotypes (Kortright et al., 2019). Generally, temperate phages are not ideal candidates for phage therapies at present because of the uncertainty they introduce because of their ability to follow both the lytic and lysogenic life cycles. However, the modification of temperate phages to obligately follow the lytic life cycle can vastly expand the available phages for therapeutic use.

Using human relevant UMB conditions, we set out to experimentally induce and propagate temperate UMB coliphages. Once propagated, the temperate UMB coliphages were then screened for sequence identity matches with any of the prophages predicted by PHASTER using PCR. We then used the PHASTER predicted prophage sequences to successfully engineer one of the induced temperate UMB coliphages into an obligately lytic coliphage by the cleavage of its integrase gene. The host range of this engineered phage, d700, was then assessed in comparison to its integrase-containing precursor phage, i700, among different UMB *E. coli* strains to gain insights into the possibility of visibly unidentifiable infection by i700 because of its participation in the lysogenic lifecycle.

Methods

Prophage Induction.

49 urinary isolates of *E. coli*, all collected via transurethral catheterization, were used for induction experiments. All 49 strains had been sequenced and assembled prior to this work. All strains were included in the PHASTER prophage prediction discussed in Chapter 2.

These strains were obtained from the Wolfe lab and were isolated through prior IRB-approved studies see (see **APPENDIX C** for more information). The urinary isolated *E. coli* were selected for attempted induction based upon their PHASTER prophage prediction results (Chapter 2); all isolates were predicted to contain at least one intact prophage sequence.

PHASTER predictions from sequenced isolates within our lab's collection (obtained from Alan J. Wolfe [Loyola University Chicago, Maywood, IL]) were screened for the well characterized λ prophage sequence by BLAST+ 2.9.0. The λ RefSeq entry, GenBank Accession no. NC_001416.1, was used as the query against a database of the predicted prophage sequences from our collection. This database was produced by using the 'makeblastdb' command. *E. coli* isolates with >95% nucleotide sequence identity were considered positive for containing a λ -like phages. Details about the BLAST hits can be found in (**Table 1**).

Table 1. Results of phage λ BLAST screening. The bacterial strains identified below were avoided for attempted phage induction because of the presence of a λ -like prophage.

Prophage Sequence	% Similarity	Region Length
AWS746_S52_L002_contigs_3_NODE_39	96.798	25780
AWS760_S57_L002_contigs_3_NODE_11	95.413	16187
AWS758_S56_L002_contigs_2_NODE_13	95.393	16175
AWS605_S25_L002_contigs_1_NODE_18	96.034	10233
AWS861_S80_L002_contigs_4_NODE_28	97.178	26090
AWS767_S60_L002_contigs_4_NODE_18	97.963	25955

AWS488_S312_L002_contigs_1_NODE_10	95.787	17101
AWS788_S66_L002_contigs_3_NODE_10	95.413	16187
AWS863_S81_L002_contigs_2_NODE_22	98.297	26232

For pH-based induction, we used a previously described protocol (Miller-Ensminger et al., 2020b). Briefly, urinary *E. coli* isolates were first streaked from a freezer stock onto 1.5% LB agar plates and incubated overnight at 37°C. Overnight cultures were then prepared by inoculating a single colony into 5 mL liquid LB and incubating at 37°C with shaking overnight. The overnight culture was then subcultured by adding 1 mL of overnight bacterial culture into 3 mL of pH-adjusted liquid LB. This step was performed for three LB pH values (pH = 4, 7, and 9). pH of 4 was produced using 1M HCl, nothing was altered about LB of pH 7, and pH of 9 was produced using 2.5M NaOH, pH was tested using Hydrion pH paper (<https://www.microessentiallab.com/>). The altered pH subcultures were grown overnight at 37°C with shaking. These pH-adjusted overnight cultures were then filtered using a 0.22 um CA syringe filter. Filtrate was spotted onto 3 different lab strains of *E. coli* (C, B, and K-12): pipetted 10 µL phage filtrate spots onto 3 mL LB soft agar (SA) [0.7% agar], and 0.5 mL overnight bacterial culture mixed and spread atop a 1.7% LB agar plate. The mixture of SA and bacterial culture on a LB agar plate as described above is called a lawn. Each pH treated spotted plate was incubated overnight at 37°C. If individual plaques or clearance was visible, the spot was harvested by scraping the visible clear plaque off the overnight spotted plate and inoculating it into 1 mL liquid LB and vortexed for 10 minutes. The inoculated plaque was then centrifuged at 12,000 x g for 1 minute and filtered using a 0.22 µm CA syringe filter. The filtrate was then used to produce a pour plate with the lab *E. coli* strain that the plaque was harvested from: 100 µL induced phage filtrate, 3 mL LB soft agar (SA) [0.7% agar], and 0.9 mL overnight bacterial

culture mixed and spread atop a 1.7% LB agar plate. The pour plate was incubated overnight at 37°C. If plaques were observed, single plaques, representative of different plaque morphologies if present, were harvested by piercing the visible plaque with a p1000 pipette tip, inoculating the harvested plaque in 1 mL liquid LB, vortexing for 10 minutes, centrifuging for 1 minute at 12,000 x g, and filtering using a 0.22 µm CA syringe filter. The resulting filtrate was then used propagated via liquid propagation: 20 µL of purified phage filtrate was added to 5 mL liquid LB along with 200 µL of lab strain *E. coli* overnight culture (the *E. coli* lab strain that the plaque was harvested from) and incubated overnight at 37°C with shaking. 1 mL of the resulting culture was centrifuged at 12,000 x g and the supernatant was filtered using a 0.22 µm CA syringe filter. The produced phage filtrate was then used to produce a pour plate as described before and this process was repeated until phage plaque morphologies were uniform.

Liquid propagated phage filtrate titer was determined by performing 1:10 serial dilutions of the propagated, filtered phage lysate (typically from 10^0 to 10^{-7} typically). The dilutions were then spotted by pipetting 10 µL of each phage lysate dilution onto a SA lawn made with 500 µL of the lab strain *E. coli* host used for propagation. After overnight incubation at 37°C, titer was determined by the smallest dilution that phage plaques can still be observed at (see **Figure 10** for example).

Induced phages were propagated to a high titer using the methods above. The high-titer lysate was then run through a viral DNA extraction protocol using the Viral DNA Extraction kit from Zymogen. The resulting induced phage DNA yields were quantified using the Qubit dsDNA protocol.

Induced Coliphage PCR Identification.

High titers of the induced phages produced using the methods above and DNA was extracted using the Viral DNA Extraction kit from Zymo, following the manufacturer's protocol. DNA yields were quantified using the Qubit dsDNA protocol.

Primers were designed using Primer-BLAST (Ye et al., 2012) to amplify the PHASTER predicted intact prophage sequences. If more than one intact prophage was predicted for a given strain, primers were designed for each predicted prophage. Primer sequences are listed in (**Table 2**) and were synthesized by Eurofins Genomics LLC (Louisville, KY USA). Temperate phage identification was performed via a 25uL PCR reaction. The PCR reaction mixture contained 1 μ L of 10 μ M forward and 1 μ L of 10 μ M reverse PCR primer, 1 μ L phage DNA, 9.5 μ L nuclease free water, and 12.5 μ L 2x GoTaq DNA polymerase. PCR thermal cycling used a denaturation temperature of 95°C for 30 seconds, an annealing temperature of 57°C for 30 seconds, and an elongation temperature of 72°C for 30 seconds. The PCR reaction was run for 32 cycles. The original bacterial strain (containing the integrated prophage) was used as a positive control for PCR reactions of its respective induced prophages. A negative control with 1 μ L nuclease free water instead of phage DNA was also included. 5 μ L of the PCR product was mixed with 2 μ L of loading dye and added to a 1.0% agarose gel. Electrophoresis was run for 20 minutes. Gels were visualized under UV light.

Table 2. PCR primers designed to identify induced prophages to PHASTER predicted prophage sequences.

Primer Name	Sequence	Predicted Prophage Target	PCR Product (bp)
667_C1_F	GCATACGCTGGCCTTAAGC	AWS667_S32_L002_contigs_1_NODE_7	230
667_C1_R	CTGCGATGCATGGGTTGATG		
667_C3_F	CCCATATCGTTGCGTTGCTG	AWS667_S32_L002_contigs_3_NODE_13	110

667_C3_R	TCTGGTGATCCATCGGGACT		
667_C7_F	AGAGCATCAGGCAGATGTGC	AWS667_S32_L002_contigs_7_NODE_45	414
667_C7_R	CGCTGCATCAGCATCAGTTC		
667_C13_F	GTATCAGGAGGCAGACAGC	AWS667_S32_L002_contigs_13_NODE_76	302
667_C13_R	ATCGCAACTGTTCCCAGATT		
777_C1_F	TCAAACCTTGCCGTTCGCTG	AWS777_S62_L002_contigs_1_NODE_2	344
777_C1_R	GCGATCAGTGTGCATTCGAC		
777_C2_F	TCACGTCGCTGTTTGCAG	AWS777_S62_L002_contigs_2_NODE_3	380
777_C2_R	CGTTTTGTCGCCCTGGTG		
777_C5_F	GAAACGATGTCGTAACGCCG	AWS777_S62_L002_contigs_5_NODE_13	452
777_C5_R	ACATCTGTGAGCTGGTGACG		
777_C6_F	GGTCAGGTGATGCACGGTAA	AWS777_S62_L002_contigs_6_NODE_16	192
777_C6_R	GTGGTGCCACTGTATGACGA		
775_C2_F	ACGCCGTTCACCTCTCAAT	AWS775_S61_L002_contigs_2_NODE_17	209
775_C2_R	TCTTACCAAGCGCGACGATAC		
775_C3_F	CTCGGACAACGACAGACACA	AWS775_S61_L002_contigs_3_NODE_24	426
775_C3_R	ATTTCGAAACCCAAGGTGC		
775_C4_F	TGCAGATCGCGCGTAAATTG	AWS775_S61_L002_contigs_4_NODE_24	300
775_C4_R	TTTGTAGCAACGCAACCTGC		
723_C2_F	GCTTCTGCGCCAAAAGTTGA	AWS723_S48_L002_contigs_2_NODE_8	128
723_C2_R	CCAAAGCGCGTATAGCCAAC		
723_C4_F	CCAGGCAAGGATTGAGTCGT	AWS723_S48_L002_contigs_4_NODE_14	352
723_C4_R	CGCACTGAATACGCTGAACG		
723_C6_F	AGCGAGCCGCTCAATACTAC	AWS723_S48_L002_contigs_6_NODE_28	225
723_C6_R	CTGGTCGCGCGTTATCTTG		
700_C1_F	GTTATGGTGATTCCCGCGCC	AWS700_S39_L002_contigs_1_NODE_2	405
700_C1_R	CTTCATGCACGCATACCACG		
700_C2_F	CCGAATATCTGCCAGCACT	AWS700_S39_L002_contigs_2_NODE_20	151
700_C2_R	TGCCGAAACCTTTAACAG		
700_C4_F	TTCAGCTTGTGCTGTCGC	AWS700_S39_L002_contigs_4_NODE_27	209
700_C4_R	CGTTCAATGCAGGTATGGCG		
831_C2_F	GCGGCATAAATGGCAGACAG	AWS831_S75_L002_contigs_2_NODE_15	368
831_C2_R	TCGGTGGGTGCCTATCAGTA		
831_C6_F	CATCAGTATGCTGCGTGTGC	AWS831_S75_L002_contigs_6_NODE_27	256
831_C6_R	CATCGACGTGGTGAAATCGC		
831_C9_F	GGGGATATTGCCCGTCTCTG	AWS831_S75_L002_contigs_9_NODE_51	221
831_C9_R	CGGTCCGGTGTAAATGCGATA		

Induced Coliphage Propagation and Extraction.

Induced phages were propagated to obtain high titer lysates that were then used for DNA extraction. Phage lysate was spotted onto a Luria broth (LB) agar plate as described previously.

The spots were then harvested, inoculated, and filtered using chloroform. The resulting 1 mL of phage lysate was added to sub-cultured liquid *E. coli* C. The sub-cultured liquid *E. coli* C was prepared using 15 mL liquid LB and 5 mL overnight *E. coli* C culture and incubated with shaking at 37°C for 2 hours. The phage lysate was then added to the sub-cultured liquid *E. coli* C culture and incubated with shaking at 37°C overnight. The resulting culture was filtered using a 0.22 um bottle-top filter from Corning Incorporated. The filtrate was then filtered and concentrated using a MacroSep Advance Centrifugal Device Omega 100K Membrane from Pall Corporation until around 1 mL of phage lysate that had not passed through the filter could be removed from the centrifuge tube. The resulting concentrated phage solution was treated with OPTIZYME DNase I, RNase-Free DNase from Fischer Bioreagents and then extracted using the Viral DNA Extraction kit from Zymogen, following the manufacturer's protocol. The resulting phage DNA yield was quantified using Qubit using their provided dsDNA protocol.

Bacterial DNA Extraction.

An overnight culture of *E. coli* strain AWS723 was made by inoculating 3 mL liquid LB with a loop from the AWS723 freezer stock into and incubating at 37°C overnight with shaking. DNA was extracted using the UltraClean Microbial DNA Extraction kit from Qiagen following the manufacturer's protocol. The DNA concentration was determined using the Qubit dsDNA protocol.

Induced Coliphage Restriction Enzyme Digest.

Induced temperate phage DNA was digested using restriction enzymes (REs). Restriction enzymes that would selectively cleave the integrase gene of the urinary *E. coli* phages were selected using the NEBCutter V2.0 website (<http://nc2.neb.com/NEBcutter2/>). FASTA

sequences of *E. coli* phages (obtained from the PHASTER prediction) were uploaded individually. Using the annotated location of each respective prophage integrase gene, all potential ‘single cutters’ were assessed. The RE was considered viable if it cleaved the phage genome inside or directly flanking the annotated location of the integrase gene.

Six prophages were identified for engineering. The FastDigest *Kpn*I, *Eco*47III, and *Bst*I from ThermoFischer were used for RE digests. RE digests were performed using 2 µL phage DNA, 2 µL FastDigest Buffer, 1 µL FastDigest RE, and 15 µL nucleotide free water for a total reaction volume of 20 µL. For each phage, multiple reactions were run with varied concentrations of phage DNA (dilutions of 10^{-1} and 10^{-2}) and FastDigest RE (dilutions of 10^{-1} and 10^{-2}). Digests were incubated for 10 minutes, 1 hour, and overnight. Digest reactions were stopped following the RE protocol in the manufacturer’s instructions.

PCR primers to detect integrase gene cleavage were developed using Primer-BLAST (Ye et al., 2012). Primers were designed so that one primer was upstream of the RE cleavage site and the other primer was downstream of the RE cleavage site. Primers were synthesized by Eurofins Genomics LLC (Louisville, KY USA).

Confirmation of RE cleavage was performed via a 25 µL PCR reaction. The PCR reaction mixture contained the same volumes and concentrations, as described in the induced prophage identification PCR protocol, but the DNA template used was digested phage DNA. A positive control of undigested induced phage DNA was used alongside a negative control of nuclease free water. PCR thermal cycling used a denaturation temperature of 95°C for 1 minute, an annealing temperature of 65°C for 30 seconds, and an elongation temperature of 72°C for 1 minute for 30 cycles. 5 µL of the PCR product was mixed with 2 µL of loading dye and added to

a 1.0% agarose gel. Electrophoresis was run for 20 minutes. Gels were visualized under UV light. PCR and agarose gel electrophoresis (AGE) methods above were performed on each digested PCR product to determine RE cleavage success.

Digested Coliphage Transformation.

Digested phage DNA was transformed into competent *E. coli* C. Competent *E. coli* C competent cells were prepared by inoculating 1 mL of overnight liquid cultured *E. coli* into 100 mL of LB and incubating at 37°C with shaking for 2 hours. The culture was then chilled on ice for 15 minutes. The culture was split into two 50 mL centrifuge tubes, which were then centrifuged at 3300 x g for 10 minutes at 4°C. The supernatant was discarded, and the cell pellet was resuspended in 40 mL of 0.1M CaCl₂ and incubated on ice for 30 minutes. The cell suspension was then centrifuged as before, the supernatant was removed, and the pellet was resuspended in a mixture of 6 mL of 0.1M CaCl₂ and 0.9 mL of 15% glycerol. The resulting mixture was aliquoted into microcentrifuge tubes, 60 µL each, and stored at -80°C.

Next, 60 µL competent *E. coli* C cells were inoculated with 10 µL of digested phage DNA and incubated with shaking at 37°C overnight. The culture was centrifuged and 10 µL was spotted, without filtering, onto an LB agar plate with 3mL soft LB agar and 500 µL *E. coli* C overnight culture. Spotted plates were incubated overnight at 37°C. Spots were harvested, inoculated into 1 mL liquid LB, vortexed for 10 minutes, and filtered using a 0.22um CA syringe filter. 100 µL of the filtered phage lysate was added to a 1 mL overnight *E. coli* C culture and incubated overnight at 37°C with shaking. This process was repeated twice to ensure the digested prophage was sufficiently recovered from transformation to be used for spotting. The phage was then propagated to a high titer, as determined via serial dilution spotting.

DNA was extracted, as described previously, and treated with DNase, also as previously described. The phage was then screened for the presence of the integrase gene via the same PCR protocol described in the initial integrase gene screening section. The transformed, propagated, and extracted digested phage DNA was used alongside the undigested phage DNA as a positive control. The results of the PCR reaction were assessed using AGE as previously described.

Host Range Determination.

30 UMB isolates of *E. coli* (obtained from the Wolfe lab) and 3 laboratory *E. coli* strains (B, C, and K12) were first streaked from a freezer stock onto 1.5% LB agar plates and incubated overnight at 37°C. A single colony from each streaked *E. coli* strain was selected to inoculate 3 mL liquid LB and was incubated overnight at 37°C with shaking. Each culture was then lawned using SA, as described previously. On each *E. coli* lawn, 10 µL of digested phage lysate (at titer 10⁹ phage per mL) and 10 µL of undigested phage lysate (at titer 10⁶ phage per mL) was spotted on each lawn in triplicate. The phage lysates used were the digested d700 phage and the initially induced, undigested, i700 phage.

Representative UMB *E. coli* Selection for Host Range Characterization.

Six unique induced and identified urinary *E. coli* intact prophage sequences predicted by PHASTER (identities in **Table 5**), identified in previous steps, were compared to 49 UMB *E. coli* isolate assembled genomes using BLAST+ 2.9.0. The 6 induced urinary coliphages were used as queries against a BLAST database of the 49 UMB *E. coli* isolate sequences, produced by using the ‘makeblastdb’ command. Default BLASTn parameters were used. Hits with >95% sequence identity and >90% query length were determined possible strains with superinfection immunity and were not considered for host range characterization.

Table 3. Superinfection BLAST hits for each PHASTER predicted sequence of the induced temperate phages.

Induced Phage	Bacteria
i667	AWS667
i700	GCF_003886275 (UMB 1348)
	GCF_003892605 (UMB 1160)
	AWS700
i723	AWS723
i775 / i831	AWS688
	GCF_003885915 (UMB 5978)
	GCF_003892445 (UMB 928)
	AWS775
	AWS695
i777	AWS734
	AWS831
	AWS777
	GCF_003886005 (UMB 5924)
	GCF_003892445 (UMB 928)
i831-1	AWS831
	GCF_003885215 (UMB 1093)
	AWS667

Serotyping of the 49 UMB *E. coli* isolate sequences was performed using the web version of SerotypeFinder 2.0.1 (<https://cge.food.dtu.dk/services/SerotypeFinder/>) (Joensen et al., 2015). Bacterial genomes were input as FASTA files. From the 49 UMB *E. coli* isolates available in the Loyola collection, 30 isolates were chosen for temperate and engineered phage host range characterization. These isolates have distinct serotypes and are not likely to have superinfection immunity. The UMB *E. coli* strains chosen are detailed in (Table 4).

Table 4. List of UMB *E. coli* strains used for phage induction (†) and host range characterization (*). Strains avoided for phage induction due to λ -like predicted prophages (^).

Bacteria Strain	Serotype	# Predicted Prophages
AWS788†	O2/50:H7	5
AWS623†	O25:H4	5
AWS760†	O2/50:H7	5

AWS833*†	O71:H12	4
AWS757†	O25:H4	4
AWS777†	O15:H6	4
AWS667†	O1:H7	4
AWS752†	O46/134:H31	4
AWS605†	O21:H25	4
AWS699*†	O132:H2	3
AWS823*†	O15:H1	3
AWS714*†	O18:H7	3
UMB6653*	O11:H18	3
AWS831†	O75:H5	3
AWS708†	O25:H4	3
AWS711†	O6:H31	3
AWS724†	O1:H7	3
AWS734†	O75:H5	3
AWS764†	O25:H4	3
AWS775†	O75:H5	3
AWS785†	O8:H28	3
AWS700†	O75:H5	3
AWS723*†	O46/134:H31	2
AWS767*^	O150:H8	2
AWS786*†	O105:H34	2
AWS790*†	O35:H4	2
AWS822*	O101:H9	2
AWS688*	O75:H5	2
AWS746*^	O25:H18	2
UMB1180*	O21:H21	2
UMB1225*	H34; No O	2
UMB1346*	O166:H15	2
AWS709†	O25:H4	2
AWS863^	O7:H15	2
AWS799†	O8:H19	2
AWS600*†	O45:H8	1
AWS466*†	O1:H7	1
AWS488*†	O16:H5	1
AWS492*	O7:H45	1
AWS758*	O4:H5	1
UMB1202*	O8:H10	1
UM2328*	O17/77:H31	1
AWS659†	O45:H70	1
AWS681†	O6:H31	1
AWS698*	O2/O50:H1	0
AWS498*	O6:H1	0
AWS845*	O19:H4	0
UMB149*	O86:H30	5
UMB2019*	O19:H4	3
UMB2055*	O59:H23	2
UMB4716*	O22:H1	6
UMB1358*	O147:H21	1

Results

Experimental Temperate Phage Induction.

Candidates for prophage induction included 49 of the UMB *E. coli* strains that were included in the prophage sequence analysis of Chapter 2. 3 of these isolates were predicted to contain 0 prophages and another 9 were removed from consideration because the predicted prophages were ‘λ-like’ (**Table 1**). Of the remaining 43 *E. coli* isolates, 30 were grown in three different liquid media conditions, with pH of 4, 7, and 9, and the resulting lysate was spotted on laboratory *E. coli* strains B, C, and K-12. In total, there were 42 observed instances where induced temperate phage(s) were able to infect and lyse lab strains of *E. coli*. Of these, 12 individual phages were able to be purified via plaque purification and consistently lyse lab strains of *E. coli* during the process of plaque purification. The identities and details of these 12 purified phages are detailed in (**Table 5**).

Because the UMB *E. coli* hosts that the 12 purified temperate coliphage stocks were induced from were all predicted to contain >1 prophage by PHASTER, primers were designed for each PHASTER-predicted prophage such that PCR could identify if the induced phage stocks were predicted by PHASTER, as well as confirm purification of the induced phage stocks (**Figure 6**). The induced phages from AWS667 were all the PHASTER-predicted prophage AWS667_S32_L002_contigs_1_NODE_7. Of the two phages purified from AWS777, both were shown to be the PHASTER-predicted prophage AWS777_S62_L002_contigs_2_NODE_3. The purified phage stocks induced from AWS700, AWS723, and AWS775 all showed positive PCR products for only one PHASTER-predicted intact prophage sequence, respectively. The phage stocks produced from induced AWS831 could be propagated using two different lab *E. coli*

strains (C and K-12). Through purification and PCR identification, it was determined that there were two different phages induced from AWS831 and these two phages were from different viral families.

The PCR methods described previously confirmed that there is only one PHASTER-predicted prophage in each of these purified induced phage stocks (suggesting that the stocks were indeed purified). Additionally, two phages were induced from AWS831 and one of these PHASTER predicted phages (AWS831_S75_L002_contigs_2_NODE_15) was able to be propagated by two different lab *E. coli* strains (K-12 and C). Via sequence comparison, it was discovered that the induced, PHASTER-predicted prophages i775 (AWS775_S61_L002_contigs_2_NODE_17) and i831-2 (AWS831_S75_L002_contigs_9_NODE_51) shared 99% sequence similarity and only i775 was used for further analysis. The 12 induced purified phages were found to represent 6 unique induced purified phage stocks. Only the six unique purified phages were used for downstream analysis, these phages are demarcated by an ‘*’ in (**Table 5**).

Table 5. Induced purified phage stock PHASTER sequence identities, PHASTER predicted prophage sequence length, experimental titer obtained, and integrase gene presence. ‘*’ denotes phage stocks that were used for downstream experiments.

Purified Phage Stock	Propagation Host	Induction pH	PHASTER Identity (Predicted Viral Taxonomic Family)	PHASTER Sequence Length	Titer (pfu/mL)	Integrase Gene
i667-1*	<i>E. coli</i> C	4	AWS667_S32_L002_contigs_1_NODE_7 (<i>Myoviridae</i>)	33.6 kb	10 ⁷	Yes
i667-2	<i>E. coli</i> C	4				
i667-3	<i>E. coli</i> C	7				
i667-4	<i>E. coli</i> C	7				
i700*	<i>E. coli</i> C	7	AWS700_S39_L002_contigs_4_NODE_27 (<i>Myoviridae</i>)	34.7 kb	10 ⁵	Yes
i723*	<i>E. coli</i> C	4	AWS723_S48_L002_contigs_4_NODE_14 (<i>Myoviridae</i>)	32.4 kb	10 ⁷	Yes
i775*	<i>E. coli</i> C	4	AWS775_S61_L002_contigs_2_NODE_17 (<i>Myoviridae</i>)	35.0 kb	10 ⁷	Yes

i777-1*	<i>E. coli</i> C	7	AWS777_S62_L002_contigs_2_NODE_3 (<i>Myoviridae</i>)	36.9 kb	10^7	Yes
i777-2	<i>E. coli</i> C	7				
i831-1*	<i>E. coli</i> K-12	7	AWS831_S75_L002_contigs_9_NODE_51 (<i>Siphoviridae</i>)	15.7 kb	10^7	No
i831-2	<i>E. coli</i> C	7	AWS831_S75_L002_contigs_2_NODE_15 (<i>Myoviridae</i>)	35.0 kb	10^7	Yes
i831-3	<i>E. coli</i> K-12	7			10^6	

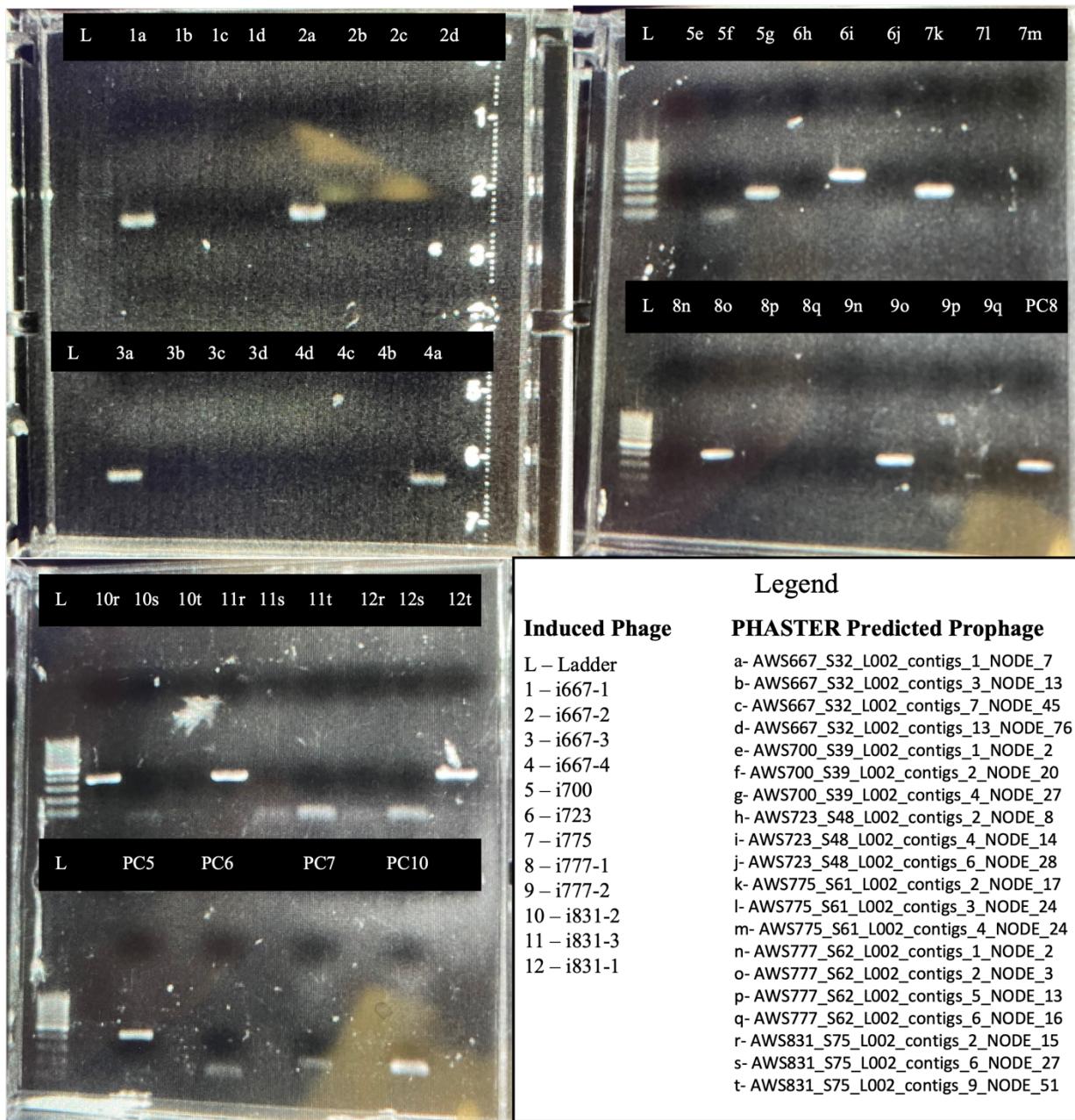


Figure 6. Induced prophage PCR products run on 1.0% agarose gels. Lanes are labeled to reflect the combination of induced prophage DNA and PCR primer used in PCR reactions. The Legend indicates the different combinations of experimentally induced phage stock DNA and intact prophage sequence-designed PCR primer. Additional positive and negative control PCR reactions were run (results not shown).

From temperate to obligately lytic.

The predicted sequences of the six identified induced phages were annotated. 5 of the 6 induced phages contained integrase genes, all of which were located at a terminal location within the linear prophage sequence (details in **Table 6**). Of these 5 phages with integrase genes, 4 (i667-1, i700, i723, and i777-1) contained restriction enzyme cleavage sites within or just upstream of the integrase gene and 3 of the 4 (i667-1, i700, and i723) were able to be produced at a high titer for DNA extraction (i667-1 = 16.8 ug/mL, i700 = 14.9 ug/mL, and i723 = 21.4 ug/mL).

Table 6. Integrase gene modification details.

Prophage Sequence	Prophage Sequence Length (bp)	Integrase Gene Location (bp..bp)	RE Cleavage Site (bp)	RE Enzyme (Fischer)	PCR Primer	Primer Sequence
AWS667_S32_L002_contigs_1_NODE_7	33622	31078..32088	31824	Eco47III [AfeI]	Eco47III_667_F	CCGATTCTGAGGAACGTGGA
					Eco47III_667_R	TCCCAGAACCTGGCAGTG
AWS700_S39_L002_contigs_4_NODE_27	34704	32189..33169	31698	KpnI	KpnI_700_F	AGCGCGATTAAACTCAGGAA
					KpnI_700_R	GTGGTTAGAAAGGGCTGGA
AWS723_S48_L002_contigs_4_NODE_14	32432	29888..30898	30634	Eco47III [AfeI]	Eco47III_723_F	CACGCTAAAACCAGCACAGA
					Eco47III_723_R	TCTGGCTTCACTCTGCGTAA
AWS775_S61_L002_contigs_2_NODE_17	35066	32522..33532	None	None	N/A	N/A
AWS777_S62_L002_contigs_2_NODE_3	36882	c(3897..4910)	c(5617)	Bst1107I [BstZ17I]	Bst1107I_777_F	AGCTGTAAGGGCATGATTG
					Bst1107I_777_R	TGGTCACCGAGACCTAACCC
AWS831_S75_L002_contigs_9_NODE_51	15659	None	N/A	N/A	N/A	N/A

Restriction digests were performed for each of these DNA samples: i667-1 cleavage by Eco47III, i700 cleavage by KpnI, i723 cleavage by Eco47III, and i777-1 cleavage by Bst1107I. Cleavage of the integrase gene for each of these phages was ascertained via PCR using primers to target the integrase (see Methods). i700 was the only phage in which the integrase gene was fully cleaved and showed no PCR signal. The other two phages (i667-1 and i723) both showed

PCR bands following the restriction digests signifying that complete digestion did not occur. The results of these PCR reactions are shown in (**Figure 6**).

The digested phage i700 lacking the integrase gene is henceforth referred to as d700. Following PCR confirmation of integrase cleavage, d700 was successfully transformed into and propagated within competent *E. coli* C (**Figure 9**). The phage d700 was able to be propagated and to reach a titer of 10^9 . After transformation and propagation, d700 was again tested for presence of its integrase gene and showed no amplification (**Figure 7**). This confirms that the lytic phage, d700, was successfully engineered. Notably the engineered phage d700 was able to reach a titer 10-fold higher than the integrase-containing phage i700 (**Figure 10**).

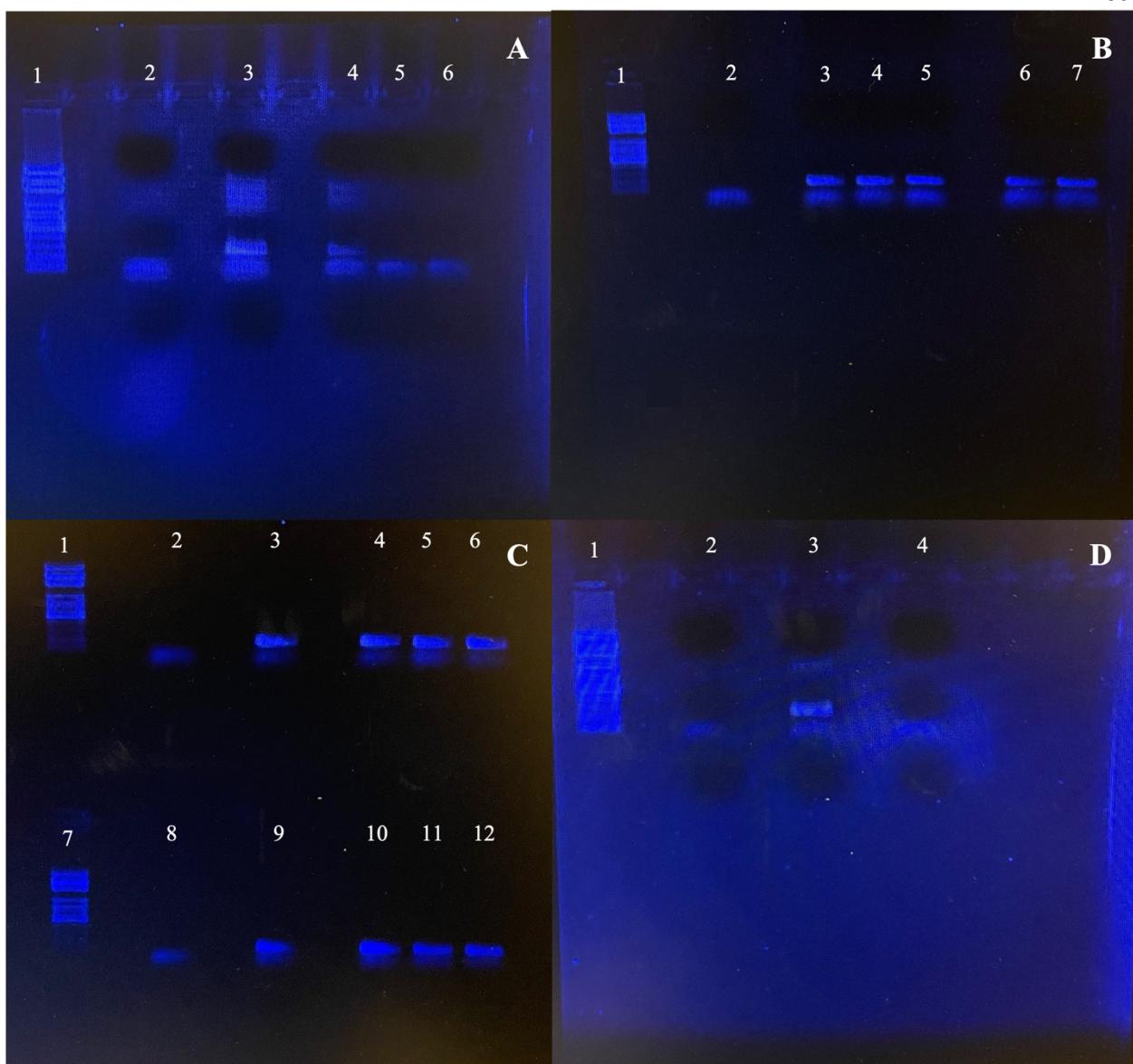


Figure 7. Gel A shows the results of PCR following the RE digest of i700. Lane 1 is a 100bp ladder, lane 2 is a NC, lane 3 is a PC with 10^{-1} DNA, lane 4 is a 10^0 DNA/ 10^{-1} RE digest PCR, lane 5 is a 10^{-1} DNA/ 10^{-1} RE digest PCR, lane 6 is a 10^{-1} DNA/ 10^{-2} RE digest PCR. Gel B shows the results of PCR following the RE digest of i723. Lane 1 is a 100bp ladder, lane 2 is a NC, lane 3 is a 10^{-1} DNA/ 10^{-1} RE digest PCR, lane 4 is a 10^{-1} DNA/ 10^{-2} RE digest PCR, lane 5 is a 10^{-2} DNA/ 10^{-2} RE digest PCR, lane 6 is a PC with 10^{-1} DNA, lane 7 is a PC with 10^{-2} DNA. Gel C shows the results of PCR following the RE digest of i667-1 (top) and i723 (bottom) with the digest conditions identical to those used in Gel A for both induced phage DNA samples. Gel D shows PCR verification of d700 phage DNA isolated and extracted after transformation and propagation into *E. coli* C. Lane 1 is a 100bp ladder, lane 2 is a NC, lane 3 is a PC with 10^0 i700 DNA, lane 4 is a 10^0 d700 phage DNA.

Induced Temperate phage Host-Range.

The genetically modified temperate phage d700 and the initially induced phage i700 were both spotted on 30 UMB *E. coli* strains, representative of 30 different serotypes. The d700 phage lysate had a titer of 10^9 . The i700 phage lysate had a titer of 10^8 . The results of phage spotting are shown in (Table 7). Although the phage i700 was able to lyse the laboratory strain *E. coli* C, it was unable to lyse any of the UMB *E. coli* strains tested. However, d700 was able to lyse UMB isolate AWS758. The lysis observed by d700 on UMB *E. coli* strain AWS758 as well as an example of no lysis by both phages are shown in (Figure 8).

Table 7. Host Range assay outcomes for the 30 UMB *E. coli* isolates and 3 lab *E. coli* strains tested. Left most column represents UMB *E. coli* strains, the second column from the right represents the serotype of each UMB *E. coli* strain respectively. The third and fourth columns from the left represent the two phages used for spotting, i700 containing integrase (temperate phage) and d700 without an integrase (obligately lytic). Red reflects no lysis observed; green reflects lysis.

	Serotype	i700	d700
AWS698	O2/O50:H1		
AWS498	O6:H1		
AWS845	O19:H4		
AWS600	O45:H8		
AWS699	O132:H2		
AWS723	O46/134:H31		
AWS767	O150:H8		
AWS786	O105:H34		
AWS790	O35:H4		
AWS822	O101:H9		
AWS823	O15:H1		
AWS833	O71:H12		
AWS466	O1:H7		
AWS488	O16:H5		
AWS492	O7:H45		
AWS714	O18:H7		
AWS688	O75:H5		
AWS746	O25:H18		
AWS758	O4:H5		Green
UMB149	O86:H30		
UMB2019	O19:H4		
UMB2055	O59:H23		

UMB4716	O22:H1		
UMB1358	O147:H21		
UMB1180	O21:H21		
UMB1202	O8:H10		
UMB1225	H34; No O		
UMB1346	O166:H15		
UMB2328	O17/77:H31		
UMB6653	O11:H18		
Ec C	N/A		
Ec B	N/A		
Ec K-12	N/A		

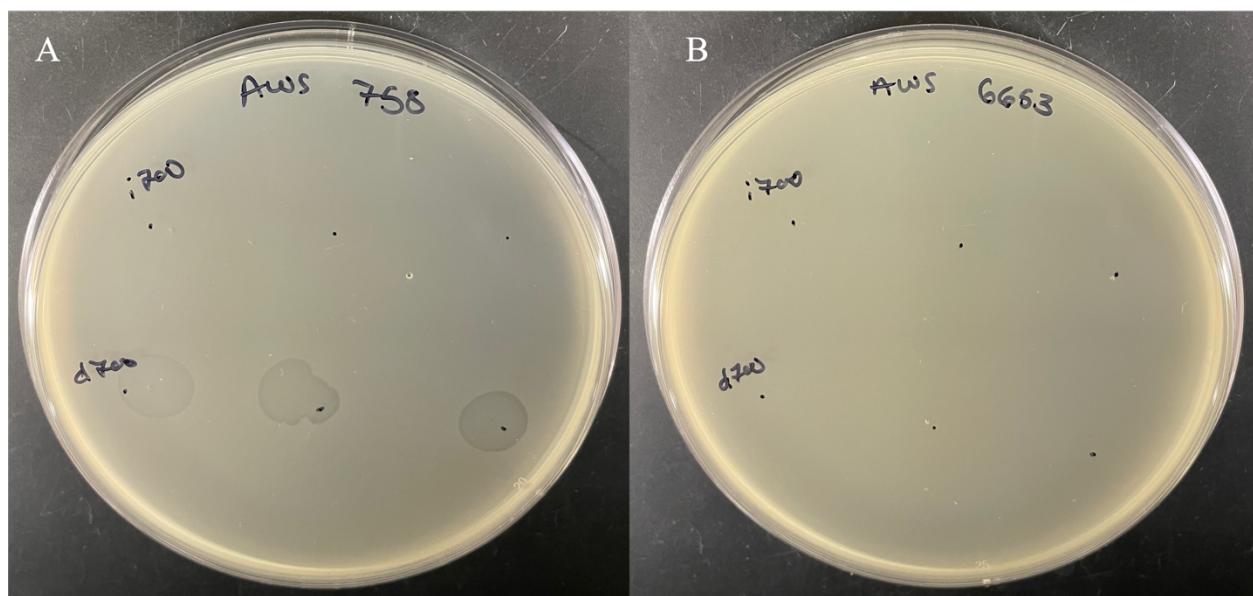


Figure 8. Image A shows the lysis caused by d700 alongside the absence of lysis by i700. Image B shows the absence of lysis by both phages. Three spots of each phage were applied in a straight line on every host tested to provide technical replicates.

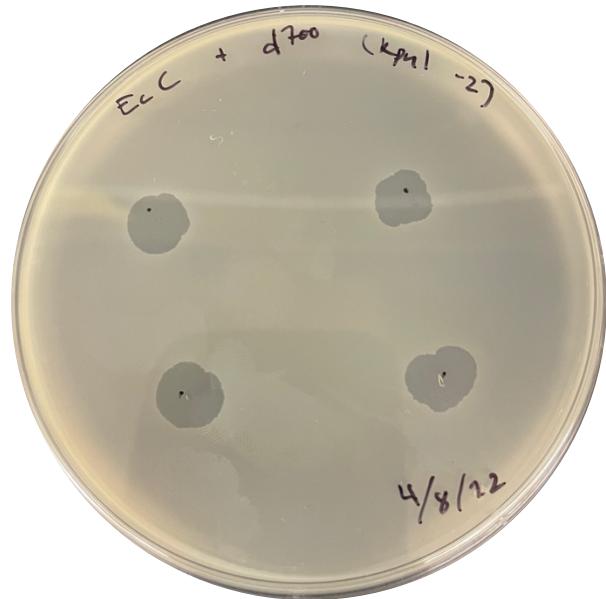


Figure 9. An incubated lawn that was spotted with transformed d700 phage to prove successful transformation and propagation following RE cleavage of its integrase gene. Bacterial clearance at each spot shows successful phage lysis.

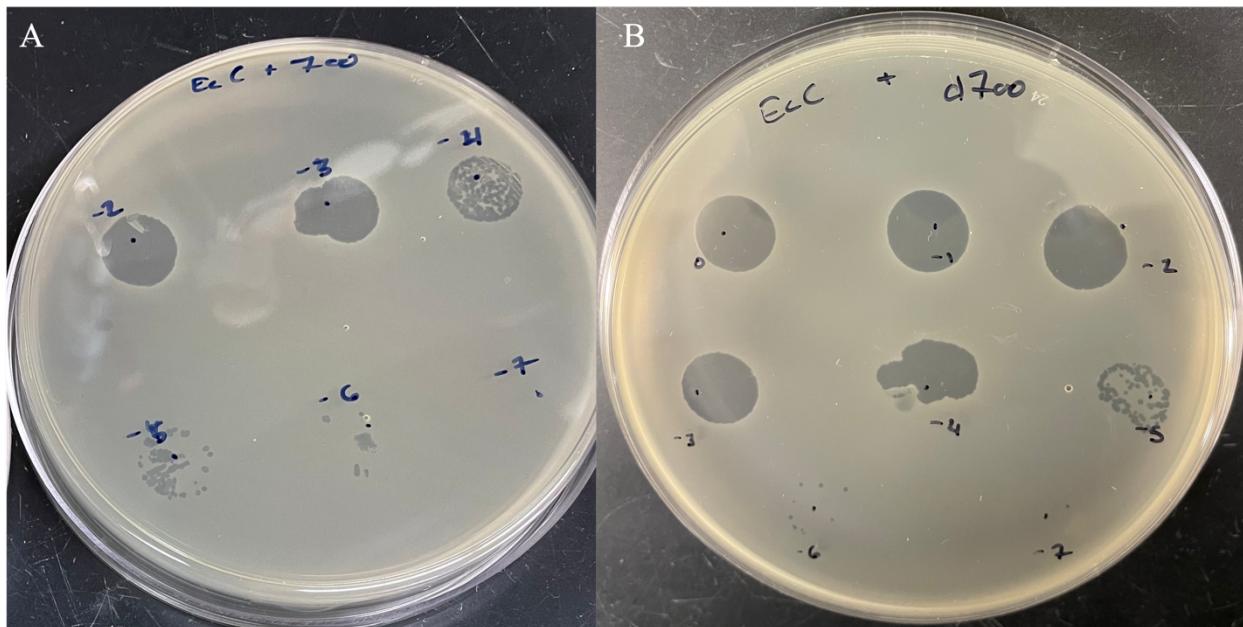


Figure 10. *Image A* shows an incubated lawn spotted with serially diluted i700 phage for titer determination. Negative numbers labelled on the plate correspond to that 10^n dilution. *Image B* shows the same titer determination methods with the phage d700. Since each spot is 10 uL, the last dilution with visible plaques and the amount of phage lysate spotted are multiplied to

determine the number of plaque forming units (pfu). *Image A* shows a titer of $\sim 10^8$ pfu; *Image B* shows a titer of $\sim 10^9$ pfu.

Discussion

UMB *E. coli* Prophage Induction.

Using the pH method for urinary temperate coliphage induction (Miller-Ensminger et al., 2020b), 42 temperate coliphages were induced from 30 different UMB *E. coli* isolates. This identification of temperate phage induction was reliant on induced phages' abilities to lyse an *E. coli* lab strain. As a result, it is very likely that there were induced phages that were not detected by this method. The pH protocol was designed to stress the *E. coli* cells using lower/higher pH, thus triggering phage induction. 5 of the 6 unique induced and purified temperate coliphages were detected without changing the pH typically used to culture *E. coli* (pH 7) (**Figure 6**). This suggests that stress did not cause prophage induction, rather the observed induction was likely the result of spontaneous induction. Spontaneous prophage induction is the entrance of a prophage into the lytic cycle (induction) in the absence of any apparent external trigger (Nanda et al., 2015). The spontaneous induction of temperate phages in this study suggests that these phages could directly contribute to shaping the UMB.

Only 12 phage stocks could be purified and replicated to a titer $> 10^5$. The other 30 induced phages, after being plated on naïve *E. coli* C and plaquing during the first iteration, failed to display lysis in subsequent iterations when attempting to purify. This could be the result of reentrance into the lysogenic cycle and integration into the *E. coli* strains being used for propagation, or the general inability of these phages to continually infect and/or lyse lab *E. coli* strains used. To encourage the identification and high titer propagation of multiple urinary coliphages, 30 *E. coli* isolate cultures were treated with the pH induction protocol as mentioned

above. As previously mentioned, 9 UMB isolates were determined to contain phages highly similar to phage λ . As this phage is well characterized, we instead focused our effort on other phages to increase our understanding of the coliphage populations within the urinary tract.

We found that all 12 phage stocks used to determine phage identity via PCR were accurately matched to an intact prophage predicted by PHASTER. In other words, we did not find any induced phages that were not predicted by PHASTER. Notably, all of the induced high-titer phage stocks were found to be from the viral order *Caudovirales*, which are the most well-characterized of coliphages (Roux et al., 2015), lending to their accurate predictions by PHASTER. In two cases (AWS667 and AWS777), our PCR found the same phage induced from the same UMB strain under different pH conditions. This suggests that a range of pH levels can cause induction and/or spontaneous induction and that induction is frequent within these *E. coli* populations. Furthermore, two of the PCR-identified phages (i775 and i831-1) that were induced from different UMB *E. coli* isolate strains were found to share 99.9% sequence similarity. This discovery further substantiates the point made by (Miller-Ensminger et al., 2018a) that the same phage species can be found in UMB isolates from multiple individuals. This phage, along with λ -like phages which were found in 9 UMB strains, may also be a common constituent of the female UMB.

Going lytic – Transforming temperate coliphages.

Removal of the integrase gene sequence has been an effective strategy for engineering a temperate phage to an obligately lytic phage. This approach has been applied to phages infectious of different bacterial taxa (Zhang et al., 2013; Kilcher et al., 2018; Johnson et al., 2022) and even applied in clinical settings (Dedrick et al., 2019). 5 of our 6 unique induced and

purified UMB coliphage sequences contained an identifiable integrase gene sequence and thus were candidates for genetic engineering. Of those 5, only 4 contained a single RE cleavage site close to or within the identified integrase gene. This method of genetic engineering via cleaving the temperate phage integrase has been successfully employed by others (Kilcher et al., 2018; Johnson et al., 2022). Details about integrase gene location and modifications can be found in (**Table 6**). The one PHASTER predicted phage sequence that did not contain an annotated integrase gene was AWS831_S75_L002_contigs_2_NODE_15. However, because there was no annotated integrase gene does not mean that the phage did not have an integrase gene; rather, it may be that the PHASTER predicted genome was not complete or that the integrase gene was too divergent from known phage integrase genes to be accurately identified during annotation. Since this phage was predicted to have a genome of nearly half the size of the other PHASTER predicted genomes, it is likely that the PHASTER prediction simply did not capture the whole phage sequence.

While we attempted to remove the integrase genes for all 4 candidate induced phages, we were only successful in digesting one of the phages – i700. The digest was confirmed via PCR (**Figure 7**). As discussed in the Methods, PCR primers were designed to flank the digest site. Thus, if full cleavage is successful, PCR will not produce DNA copies and there will be no amplicon. It should be noted that several different DNA concentrations, enzyme concentrations, and incubation times were tested for the candidate phages to find the optimal conditions for engineering. The manufacturer’s suggested protocol using a RE digest incubation time of 10 minutes and an undiluted concentration of phage DNA was unsuccessful; gels for all phage samples displayed amplicons. The only phage for which digestion was confirmed via PCR was

d700, the digested, integrase-free phage i700 seen in (**Figure 7**). The other digested phages d667 and d723, which notably both used the RE *Eco47III* in their digest reactions, were unable to produce PCR results confirming successful cleavage. Future work is required to optimize this protocol.

After confirmation of integrase cleavage in d700, the phage DNA was transformed into competent *E. coli* C cells. The successful transformation of d700 and subsequent plaquing on an *E. coli* C lawn confirms that this phage is still viable without the integrase gene (**Figure 9** and **Figure 10**). Removal was confirmed again via PCR after transformation and propagation (**Figure 7**). When the host range of d700 was compared to that of its ancestral temperate phage i700, one change in lytic efficiency was observed. On the UMB strain AWS758, i700 showed no visible lysis while d700 showed obvious lytic clearance (**Figure 8**). The host range spotting results can be seen in (**Table 7**). This suggests that the temperate phage could infect the AWS758 strain but did not reproduce via the lytic cycle. We hypothesize that it switched from the lytic to the lysogenic life cycle and potentially integrated into the bacterial genome, but more work is needed to confirm this.

Engineering temperate phages for phage therapy.

The engineering protocol used here not only provides a method for assessing the ability of a temperate phage to infect a host cell but also is a potential method for increasing the phages in our arsenal for use as phage therapy. While d700 is unlikely a viable candidate, given its narrow host range, this protocol could be applied to other urinary coliphages with an identifiable and excisable integrase gene. As our analysis of urinary *E. coli* prophages (Chapter Two) showed, there is a wealth of different phages lysogenizing urinary *E. coli*. Their presence in the

genomes examined shows their ability to infect urinary *E. coli* strains. While here we targeted engineering of understudied phages, selecting temperate phages that are commonly found in urinary *E. coli* genomes have the potential of displaying broader host ranges.

In recent decades, various novel phage therapy strategies have been developed (for a review of phage therapy methodology, see ref Zalewska-Piątek and Piątek, 2020). Historically, phage therapy relied strictly on obligately lytic phages. Temperate phages, however, significantly outnumber obligately lytic phages in nature (Monteiro et al., 2019). Recently, advances in synthetic biology and sequencing technology have made it possible to genetically alter temperate phages, such as the methods employed for d700. This makes them an attractive and even preferable option for development of phage therapies (Fiddian-Green and Silen, 1975; Monteiro et al., 2019). In the past, temperate phages were avoided because they were not as effective at killing bacterial hosts and they posed the threat of mediating horizontal gene transfer that could result in the transmission of unwanted trait transmission, such as antibiotic resistance (Monteiro et al., 2019). In the past decade, numerous studies have documented the potential efficacy of temperate phages being used for phage therapy (see Monteiro et al., 2019).

With the advent of engineered temperate phage-based phage therapies, studies of the phage abundance and diversity of the UMB is warranted. One of the primary challenges posed by phage therapy in the case of antibiotic resistant bacterial infections of the human bladder and lower urinary tract is the lack of candidate lytic phages capable of infecting and clearing the antibiotic-resistant bacteria implicated (Caflisch et al., 2019). One potential solution that has been recently made possible is the consideration of using engineered temperate phages for such phage therapy treatments. As has been established, *E. coli* is disproportionately implicated in

complex, antibiotic resistant bladder disorders, and is, therefore, an obvious candidate for novel temperate phage therapy applications.

CHAPTER 4

CONCLUSIONS

This study provides an improved catalogue of urinary *E. coli* prophage diversity through the analysis of all publicly available urinary *E. coli* genomic assemblies. We were able to identify taxonomically diverse *E. coli* prophages that represented the diversity of dsDNA tailed phage families *Siphoviridae*, *Myoviridae*, and *Podoviridae*. The production of a network through the connection of prophages with shared genes shows that homologous genes are shared between UMB prophages of different viral families providing evidence for genetic mosaicism among urinary coliphages. Nearly half of the intact prophages predicted were taxonomically classified as “Unknown” due to divergence from characterized phages. Future characterization of these novel prophages would expand our knowledge of coliphage taxonomy and genetic diversity.

Our analyses found that antibiotic resistance and virulence factor genes are rare within urinary prophages. Thus, we conclude that lysogenic conversion is not a significant contribution to uropathogenicity. This finding also has implications for use of native urinary coliphages for therapeutic use. Lysogenic conversion is a frequent concern when considering the use of temperate phages for phage therapy. Given the low incidence of antibiotic resistance and virulence genes carried by these temperate phages, the risk is mitigated. Thus, the catalog of prophages identified here presents numerous candidates, granted they can replicate efficiently through the lytic life cycle, for phage therapy use.

Using biologically relevant pH conditions, we showed that urinary *E. coli* prophages can be induced and propagated to high titer through the lytic life cycle. Transforming temperate lytic phages to obligately lytic would provide for more effective phage therapy, as demonstrated by our engineering of the temperate phage i700. We showed the successful selective cleavage of the integrase gene from the temperate phage i700 without inhibition of the phage's ability to infect and lyse *E. coli* C. The engineered coliphage, d700, was able to reach a higher titer than the integrase-containing unmodified i700 coliphage. The host ranges of the engineered phage d700 and integrase-containing i700 were compared among various UMB *E. coli* strains with different serotypes. We showed that d700 was able to lyse one strain of UMB *E. coli* that i700 was not. These results suggest that i700 can infect the AWS758 strain of UMB *E. coli*, but instead of proceeding through the lytic cycle, i700 proceeds though the lysogenic cycle and integrates into that host cell's genome.

Employing an engineering strategy, such as that demonstrated here, could be used in future studies as an assay to determine whether the temperate phages are integrating into bacterial host genomes or if they are unable to infect cells when lytic activity is not observed. Furthermore, it provides an effective strategy for developing new obligately lytic phages for therapeutic use. Our analysis of the predicted intact prophages of the urinary *E. coli* strains found 1,938 prophages with recognizable integrase genes. All of these prophages are candidates for engineering as was performed for i700.

APPENDIX A

TOP 30 VIRULENCE FACTOR GENE DESCRIPTIONS IDENTIFIED IN PROPHAGE GENOME

Virulence Factor Description	Occurrences
(aaIQ) -	137
(sitA) Iron ABC transporter substrate binding component	63
(YPK_3315) invasin	54
(ipaH) hypothetical prophage protein	49
(aaIR) -	48
(sitD) SitD protein	46
(sitB) manganese/iron transport system ATP-binding protein	44
(sitC) SitC protein	43
(flk) flagella biosynthesis regulator	42
(aaIW) transposase	38
(gtrA) bactoprenol-linked glucose translocase	32
(ipaH) hypothetical protein	32
(iucD) L-lysine 6-monooxygenase	31
(sat) putative secreted autotransporter toxin sat	29
(iucB) aerobactin siderophore biosynthesis protein lucB	27
(sitD) iron transport protein, inner membrane component	27
(iucA) lucA protein	25
(sitC) Iron transporter inner membrane component	23
(c3610) bifunctional enterobactin receptor/adhesin protein	21
(papB) pap operon regulatory protein PapB	18
(sitB) iron ABC transporter ATP-binding protein	18
(sitC) iron ABC transporter permease	18
(iutA) ferric aerobactin receptor precursor IutA	17
(papA) major pilin subunit PapA	17
(cdtA) cytolethal distending toxin type IV subunit A	15
(cdtB) cytolethal distending toxin type IV subunit B	15
(cdtC) cytolethal distending toxin type IV subunit C	15
(iucC) aerobactin siderophore biosynthesis protein lucC	15
(papl) PapI protein	15

APPENDIX B

ALL ANTIBIOTIC DRUG CLASSES TARGETED BY INTACT PROPHAGE ANTIBIOTIC
RESISTANCE GENE

Antibiotic Drug Class Target	Occurrences
aminoglycoside antibiotic; aminocoumarin antibiotic	84
macrolide antibiotic	55
fluoroquinolone antibiotic; monobactam; carbapenem; cephalosporin; glycyclcline; cephamicin; penam; tetracycline antibiotic; rifamycin antibiotic; phenicol antibiotic; triclosan; penem	39
aminocoumarin antibiotic	34
aminoglycoside antibiotic	22
tetracycline antibiotic; benzalkonium chloride; rhodamine	18
sulfonamide antibiotic	10
tetracycline antibiotic	8
acridine dye; disinfecting agents and intercalating dyes	5
diaminopyrimidine antibiotic	5
fluoroquinolone antibiotic; cephalosporin; glycyclcline; penam; tetracycline antibiotic; rifamycin antibiotic; phenicol antibiotic; triclosan	5
carbapenem; cephalosporin; penam	4
cephalosporin; penam	4
macrolide antibiotic; aminoglycoside antibiotic; cephalosporin; tetracycline antibiotic; peptide antibiotic; rifamycin antibiotic	4
macrolide antibiotic; fluoroquinolone antibiotic; penam; tetracycline antibiotic	3
monobactam; cephalosporin; penam; penem	3
phenicol antibiotic	3
fluoroquinolone antibiotic; glycyclcline; tetracycline antibiotic; diaminoypyrimidine antibiotic; nitrofuran antibiotic	2
cephalosporin	1
fluoroquinolone antibiotic	1
fluoroquinolone antibiotic; aminoglycoside antibiotic	1
monobactam; carbapenem; cephalosporin; penam	1
rifamycin antibiotic	1

APPENDIX C

PATIENT UMB *E. COLI* STRAIN ISOLATION IRB DETAILS

Bacteria Strain	UMB #	Study Patient ID	IRB #	University for IRB	Citations
AWS466	7965	RUTISD-4	170077A W	UCSD	https://pubmed.ncbi.nlm.nih.gov/34036621/ ; https://pubmed.ncbi.nlm.nih.gov/33852041/
AWS488	8034	SvE029	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS492	8038	SvE031	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS498	8045	SvE033	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS600	8611	SVE090	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS605	8625	SvE093	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS623	8727	RUTISD33	170077A W	UCSD	https://pubmed.ncbi.nlm.nih.gov/34036621/ ; https://pubmed.ncbi.nlm.nih.gov/33852041/
AWS659	8988	SvE102	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS667	9006	RUTISD038	170077A W	UCSD	https://pubmed.ncbi.nlm.nih.gov/34036621/ ; https://pubmed.ncbi.nlm.nih.gov/33852041/
AWS681	9077	SvE111	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS688	9084	RUTISD026	170077A W	UCSD	https://pubmed.ncbi.nlm.nih.gov/34036621/ ; https://pubmed.ncbi.nlm.nih.gov/33852041/
AWS698	9094	SVE113	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS699	9101	SVE114	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS700	9105	SVE116	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS708	9135	SvE121	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS709	9136	SvE122	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS711	9138	SvE124	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS714	9141	SvE125	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS723	9208	SVE128	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS724	9209	SvE130	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS734	9227	SvE131	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS746	9246	RUTISD43	170077A W	UCSD	https://pubmed.ncbi.nlm.nih.gov/34036621/ ; https://pubmed.ncbi.nlm.nih.gov/33852041/
AWS752	9257	SVE138	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS757	9268	SVE140	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS758	9269	SVE141	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS760	9288	SVE142	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS764	9318	SVE144	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS767	9330	SvE148	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS775	9344	SVE150	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS777	9346	SVE153	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS785	9590	sve157	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS786	9596	SvE158	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS788	9598	SvE160	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS790	9616	SVE161	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS799	9853	RUTISD47	170077A W	UCSD	https://pubmed.ncbi.nlm.nih.gov/34036621/ ; https://pubmed.ncbi.nlm.nih.gov/33852041/

AWS822	9892	SVE171	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS823	9901	SvE172	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS831	9930	SvE176	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS833	9932	SvE178	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS845	9974	SVE181	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
AWS863	10045	SvE188	209545	Loyola	https://pubmed.ncbi.nlm.nih.gov/34184930/
UM2328	328	OAB031	203986	Loyola	PMID: 26423260, 25006228, 26210757, 24371246
UMB1180	1180	EQUC129	206469	Loyola	https://pubmed.ncbi.nlm.nih.gov/26962083/
UMB1202	1202	EQUC132	206469	Loyola	https://pubmed.ncbi.nlm.nih.gov/26962083/
UMB1225	1225	EQUC141	206469	Loyola	https://pubmed.ncbi.nlm.nih.gov/26962083/
UMB1346	1346	EQUC0152	206469	Loyola	https://pubmed.ncbi.nlm.nih.gov/26962083/
UMB1358	1358	EQUC0155	206469	Loyola	https://pubmed.ncbi.nlm.nih.gov/26962083/
UMB149	149	OAB033	203986	Loyola	PMID: 26423260, 25006228, 26210757, 24371246
UMB2019	2019	EST07	207152	Loyola	PMID: 29674608
UMB2055	2055	EST07	207152	Loyola	PMID: 29674608
UMB4716	4716	EST22B	207152	Loyola	PMID: 29674608
UMB6653	6653	RUTI20	204195	Loyola	https://pubmed.ncbi.nlm.nih.gov/34036621/ ; https://pubmed.ncbi.nlm.nih.gov/33852041/

REFERENCE LIST

1. Ackerman, A. L., and Chai, T. C. (2019). The Bladder is Not Sterile: an Update on the Urinary Microbiome. *Curr. Bladder Dysfunct. Rep.* 14, 331–341. doi: 10.1007/s11884-019-00543-6.
2. Ackermann, H.-W. (2007). 5500 Phages examined in the electron microscope. *Arch. Virol.* 152, 227–243. doi: 10.1007/s00705-006-0849-1.
3. Adebayo, A. S., Ackermann, G., Bowyer, R. C. E., Wells, P. M., Humphreys, G., Knight, R., et al. (2020). The Urinary Tract Microbiome in Older Women Exhibits Host Genetic and Environmental Influences. *Cell Host Microbe* 28, 298–305.e3. doi: 10.1016/j.chom.2020.06.022.
4. Akhter, S., Aziz, R. K., and Edwards, R. A. (2012). PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. *Nucleic Acids Res.* 40, e126–e126. doi: 10.1093/nar/gks406.
5. Alcock, B. P., Raphenya, A. R., Lau, T. T. Y., Tsang, K. K., Bouchard, M., Edalatmand, A., et al. (2019). CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.*, gkz935. doi: 10.1093/nar/gkz935.
6. Al-Shayeb, B., Sachdeva, R., Chen, L.-X., Ward, F., Munk, P., Devoto, A., et al. (2020). Clades of huge phages from across Earth’s ecosystems. *Nature* 578, 425–431. doi: 10.1038/s41586-020-2007-4.
7. Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., et al. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44, W16–W21. doi: 10.1093/nar/gkw387.
8. Balding, C., Bromley, S. A., Pickup, R. W., and Saunders, J. R. (2005). Diversity of phage integrases in Enterobacteriaceae: development of markers for environmental analysis of temperate phages. *Environ. Microbiol.* 7, 1558–1567. doi: 10.1111/j.1462-2920.2005.00845.x.
9. Ball, C. A., and Johnson, R. C. (1991). Multiple effects of Fis on integration and the control of lysogeny in phage lambda. *J. Bacteriol.* 173, 4032–4038. doi: 10.1128/jb.173.13.4032-4038.1991.
10. Batinovic, Wasif, Knowler, Rice, Stanton, Rose, et al. (2019). Bacteriophages in Natural and Artificial Environments. *Pathogens* 8, 100. doi: 10.3390/pathogens8030100.
11. Bikel, S., Gallardo-Becerra, L., Cornejo-Granados, F., and Ochoa-Leyva, A. (2022). Protocol for the isolation, sequencing, and analysis of the gut phageome from human fecal samples. *STAR Protoc.* 3, 101170. doi: 10.1016/j.xpro.2022.101170.

12. Bobay, L.-M., Rocha, E. P. C., and Touchon, M. (2013). The Adaptation of Temperate Bacteriophages to Their Host Genomes. *Mol. Biol. Evol.* 30, 737–751. doi: 10.1093/molbev/mss279.
13. Borodovich, T., Shkoporov, A. N., Ross, R. P., and Hill, C. (2022). Phage-mediated horizontal gene transfer and its implications for the human gut microbiome. *Gastroenterol. Rep.* 10, goac012. doi: 10.1093/gastro/goac012.
14. Brassil, B., Mores, C. R., Wolfe, A. J., and Putonti, C. (2020). Characterization and spontaneous induction of urinary tract *Streptococcus anginosus* prophages. *J. Gen. Virol.* 101, 685–691. doi: 10.1099/jgv.0.001407.
15. Brown-Jaque, M., Muniesa, M., and Navarro, F. (2016). Bacteriophages in clinical samples can interfere with microbiological diagnostic tools. *Sci. Rep.* 6, 33000. doi: 10.1038/srep33000.
16. Brubaker, L., and Wolfe, A. J. (2017). Associating infection and incontinence with the female urinary microbiota. *Nat. Rev. Urol.* 14, 72–74. doi: 10.1038/nrurol.2016.262.
17. Bruce, J. B., Lion, S., Buckling, A., Westra, E. R., and Gandon, S. (2021). Regulation of prophage induction and lysogenization by phage communication systems. *Curr. Biol.* 31, 5046-5051.e7. doi: 10.1016/j.cub.2021.08.073.
18. Brüssow, H., Canchaya, C., and Hardt, W.-D. (2004). Phages and the Evolution of Bacterial Pathogens: from Genomic Rearrangements to Lysogenic Conversion. *Microbiol. Mol. Biol. Rev.* 68, 560–602. doi: 10.1128/MMBR.68.3.560-602.2004.
19. Caflisch, K. M., Suh, G. A., and Patel, R. (2019). Biological challenges of phage therapy and proposed solutions: a literature review. *Expert Rev. Anti Infect. Ther.* 17, 1011–1041. doi: 10.1080/14787210.2019.1694905.
20. Calero-Cáceres, W., and Muniesa, M. (2016). Persistence of naturally occurring antibiotic resistance genes in the bacteria and bacteriophage fractions of wastewater. *Water Res.* 95, 11–18. doi: 10.1016/j.watres.2016.03.006.
21. Campbell, A. (2003). The future of bacteriophage biology. *Nat. Rev. Genet.* 4, 471–477. doi: 10.1038/nrg1089.
22. Casjens, S. (2003). Prophages and bacterial genomics: what have we learned so far?: Prophage genomics. *Mol. Microbiol.* 49, 277–300. doi: 10.1046/j.1365-2958.2003.03580.x.
23. Casjens, S. R., and Hendrix, R. W. (2015). Bacteriophage lambda: Early pioneer and still relevant. *Virology* 479–480, 310–330. doi: 10.1016/j.virol.2015.02.010.

24. Chen, J., Carpena, N., Quiles-Puchalt, N., Ram, G., Novick, R. P., and Penadés, J. R. (2015). Intra- and inter-generic transfer of pathogenicity island-encoded virulence genes by cos phages. *ISME J.* 9, 1260–1263. doi: 10.1038/ismej.2014.187.
25. Chen, Y., Yang, L., Yang, D., Song, J., Wang, C., Sun, E., et al. (2020). Specific Integration of Temperate Phage Decreases the Pathogenicity of Host Bacteria. *Front. Cell. Infect. Microbiol.* 10, 14. doi: 10.3389/fcimb.2020.00014.
26. Colavecchio, A., D’Souza, Y., Tompkins, E., Jeukens, J., Freschi, L., Emond-Rheault, J.-G., et al. (2017). Prophage Integrase Typing Is a Useful Indicator of Genomic Diversity in *Salmonella enterica*. *Front. Microbiol.* 8, 1283. doi: 10.3389/fmicb.2017.01283.
27. Dallas, S. D., and Kingsberry, L. (1997). Bacteriophage Plaques on Primary Isolation Media of a Urine Culture Growing *Escherichia coli*. *Clin. Microbiol. Newslet.* 19, 53–56. doi: 10.1016/S0196-4399(97)87667-2.
28. Darling, A. C. E., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. *Genome Res.* 14, 1394–1403. doi: 10.1101/gr.2289704.
29. Davis, J. J., Wattam, A. R., Aziz, R. K., Brettin, T., Butler, R., Butler, R. M., et al. (2019). The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. *Nucleic Acids Res.*, gkz943. doi: 10.1093/nar/gkz943.
30. De Koster, S., Rodriguez Ruiz, J. P., Rajakani, S. G., Lammens, C., Glupczynski, Y., Goossens, H., et al. (2022). Diversity in the Characteristics of *Klebsiella pneumoniae* ST101 of Human, Environmental, and Animal Origin. *Front. Microbiol.* 13, 838207. doi: 10.3389/fmicb.2022.838207.
31. Dedrick, R. M., Guerrero-Bustamante, C. A., Garlena, R. A., Russell, D. A., Ford, K., Harris, K., et al. (2019). Engineered bacteriophages for treatment of a patient with a disseminated drug-resistant *Mycobacterium abscessus*. *Nat. Med.* 25, 730–733. doi: 10.1038/s41591-019-0437-z.
32. Dodd, I. B., Shearwin, K. E., and Egan, J. B. (2005). Revisited gene regulation in bacteriophage lambda. *Curr. Opin. Genet. Dev.* 15, 145–152. doi: 10.1016/j.gde.2005.02.001.
33. Dong, Q., Nelson, D. E., Toh, E., Diao, L., Gao, X., Fortenberry, J. D., et al. (2011). The Microbial Communities in Male First Catch Urine Are Highly Similar to Those in Paired Urethral Swab Specimens. *PLoS ONE* 6, e19709. doi: 10.1371/journal.pone.0019709.
34. Dudley, E. G., Thomson, N. R., Parkhill, J., Morin, N. P., and Nataro, J. P. (2006). Proteomic and microarray characterization of the AggR regulon identifies a pheU pathogenicity island in enteroaggregative *Escherichia coli*. *Mol. Microbiol.* 61, 1267–1282. doi: 10.1111/j.1365-2958.2006.05281.x.

35. Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461.
36. Eren, A. M., Esen, Ö. C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., et al. (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* 3, e1319. doi: 10.7717/peerj.1319.
37. Fiddian-Green, R. G., and Silen, W. (1975). Mechanisms of disposal of acid and alkali in rabbit duodenum. *Am. J. Physiol.* 229, 1641–1648. doi: 10.1152/ajplegacy.1975.229.6.1641.
38. Fogg, P. C. M., Colloms, S., Rosser, S., Stark, M., and Smith, M. C. M. (2014). New Applications for Phage Integrases. *J. Mol. Biol.* 426, 2703–2716. doi: 10.1016/j.jmb.2014.05.014.
39. Gan, R., Zhou, F., Si, Y., Yang, H., Chen, C., Ren, C., et al. (2022). DBSCAN-SWA: An Integrated Tool for Rapid Prophage Detection and Annotation. *Front. Genet.* 13, 885048. doi: 10.3389/fgene.2022.885048.
40. Garretto, A. (2019). Exploring associations between lysogeny and host abundance.
41. Garretto, A., Miller-Ensminger, T., Ene, A., Merchant, Z., Shah, A., Gerodias, A., et al. (2020). Genomic Survey of *E. coli* From the Bladders of Women With and Without Lower Urinary Tract Symptoms. *Front. Microbiol.* 11, 2094. doi: 10.3389/fmicb.2020.02094.
42. Garretto, A., Miller-Ensminger, T., Wolfe, A. J., and Putonti, C. (2019). Bacteriophages of the lower urinary tract. *Nat. Rev. Urol.* 16, 422–432. doi: 10.1038/s41585-019-0192-4.
43. Garretto, A., Thomas-White, K., Wolfe, A. J., and Putonti, C. (2018). Detecting viral genomes in the female urinary microbiome. *J. Gen. Virol.* 99, 1141–1146. doi: 10.1099/jgv.0.001097.
44. Geerlings, S. E. (2016). Clinical Presentations and Epidemiology of Urinary Tract Infections. *Microbiol. Spectr.* 4, 4.5.03. doi: 10.1128/microbiolspec.UTI-0002-2012.
45. Giannattasio-Ferraz, S., Ene, A., Gomes, V. J., Queiroz, C. O., Maskeri, L., Oliveira, A. P., et al. (2022). Escherichia coli and Pseudomonas aeruginosa Isolated From Urine of Healthy Bovine Have Potential as Emerging Human and Bovine Pathogens. *Front. Microbiol.* 13, 764760. doi: 10.3389/fmicb.2022.764760.
46. Goerke, C., Pantucek, R., Holtfreter, S., Schulte, B., Zink, M., Grumann, D., et al. (2009). Diversity of Prophages in Dominant *Staphylococcus aureus* Clonal Lineages. *J. Bacteriol.* 191, 3462–3468. doi: 10.1128/JB.01804-08.

47. Goodlet, K. J., Benhalima, F. Z., and Nailor, M. D. (2019). A Systematic Review of Single-Dose Aminoglycoside Therapy for Urinary Tract Infection: Is It Time To Resurrect an Old Strategy? *Antimicrob. Agents Chemother.* 63, e02165-18. doi: 10.1128/AAC.02165-18.
48. Grindley, N. D. F., Whiteson, K. L., and Rice, P. A. (2006). Mechanisms of site-specific recombination. *Annu. Rev. Biochem.* 75, 567–605. doi: 10.1146/annurev.biochem.73.011303.073908.
49. Grose, J. H., and Casjens, S. R. (2014). Understanding the enormous diversity of bacteriophages: The tailed phages that infect the bacterial family Enterobacteriaceae. *Virology* 468–470, 421–443. doi: 10.1016/j.virol.2014.08.024.
50. Groth, A. C., and Calos, M. P. (2004). Phage Integrases: Biology and Applications. *J. Mol. Biol.* 335, 667–678. doi: 10.1016/j.jmb.2003.09.082.
51. Guo, J., Bolduc, B., Zayed, A. A., Varsani, A., Dominguez-Huerta, G., Delmont, T. O., et al. (2021a). VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* 9, 37. doi: 10.1186/s40168-020-00990-y.
52. Guo, J., Bolduc, B., Zayed, A. A., Varsani, A., Dominguez-Huerta, G., Delmont, T. O., et al. (2021b). VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* 9, 37. doi: 10.1186/s40168-020-00990-y.
53. Gutierrez-Guerrero, A., Cosset, F.-L., and Verhoeven, E. (2020). Lentiviral Vector Pseudotypes: Precious Tools to Improve Gene Modification of Hematopoietic Cells for Research and Gene Therapy. *Viruses* 12, 1016. doi: 10.3390/v12091016.
54. Happel, A.-U., Kullin, B. R., Gamieldien, H., Jaspan, H. B., Varsani, A., Martin, D., et al. (2022). In Silico Characterisation of Putative Prophages in Lactobacillaceae Used in Probiotics for Vaginal Health. *Microorganisms* 10, 214. doi: 10.3390/microorganisms10020214.
55. Hatfull, G. F. (2008). Bacteriophage genomics. *Curr. Opin. Microbiol.* 11, 447–453. doi: 10.1016/j.mib.2008.09.004.
56. Hatfull, G. F., and Hendrix, R. W. (2011). Bacteriophages and their genomes. *Curr. Opin. Virol.* 1, 298–303. doi: 10.1016/j.coviro.2011.06.009.
57. Heide, L. (2014). New aminocoumarin antibiotics as gyrase inhibitors. *Int. J. Med. Microbiol.* 304, 31–36. doi: 10.1016/j.ijmm.2013.08.013.
58. Hendrix, R. W. (2002). Bacteriophages: Evolution of the Majority. *Theor. Popul. Biol.* 61, 471–480. doi: 10.1006/tpbi.2002.1590.

59. Hendrix, R. W., Smith, M. C. M., Burns, R. N., Ford, M. E., and Hatfull, G. F. (1999). Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. *Proc. Natl. Acad. Sci.* 96, 2192–2197. doi: 10.1073/pnas.96.5.2192.
60. Hilt, E. E., McKinley, K., Pearce, M. M., Rosenfeld, A. B., Zilliox, M. J., Mueller, E. R., et al. (2014). Urine Is Not Sterile: Use of Enhanced Urine Culture Techniques To Detect Resident Bacterial Flora in the Adult Female Bladder. *J. Clin. Microbiol.* 52, 871–876. doi: 10.1128/JCM.02876-13.
61. Hobbs, Z., and Abedon, S. T. (2016). Diversity of phage infection types and associated terminology: the problem with ‘Lytic or lysogenic.’ *FEMS Microbiol. Lett.* 363, fnw047. doi: 10.1093/femsle/fnw047.
62. Hochschild, A., and Lewis, M. (2009). The bacteriophage λ CI protein finds an asymmetric solution. *Curr. Opin. Struct. Biol.* 19, 79–86. doi: 10.1016/j.sbi.2008.12.008.
63. Howard-Varona, C., Hargreaves, K. R., Abedon, S. T., and Sullivan, M. B. (2017). Lysogeny in nature: mechanisms, impact and ecology of temperate phages. *ISME J.* 11, 1511–1520. doi: 10.1038/ismej.2017.16.
64. Hyman, P., and Abedon, S. T. (2010a). “Bacteriophage Host Range and Bacterial Resistance,” in *Advances in Applied Microbiology* (Elsevier), 217–248. doi: 10.1016/S0065-2164(10)70007-1.
65. Hyman, P., and Abedon, S. T. (2010b). “Bacteriophage Host Range and Bacterial Resistance,” in *Advances in Applied Microbiology* (Elsevier), 217–248. doi: 10.1016/S0065-2164(10)70007-1.
66. Joensen, K. G., Tetzschner, A. M. M., Iguchi, A., Aarestrup, F. M., and Scheutz, F. (2015). Rapid and Easy *In Silico* Serotyping of Escherichia coli Isolates by Use of Whole-Genome Sequencing Data. *J. Clin. Microbiol.* 53, 2410–2426. doi: 10.1128/JCM.00008-15.
67. Johnson, G., Banerjee, S., and Putonti, C. (2022). Diversity of *Pseudomonas aeruginosa* Temperate Phages. *mSphere* 7, e01015-21. doi: 10.1128/msphere.01015-21.
68. Johnson, G., and Putonti, C. (2019). Genome Sequence of *Pseudomonas* Phage UMP151, Isolated from the Female Bladder Microbiota. *Microbiol. Resour. Announc.* 8, e00853-19. doi: 10.1128/MRA.00853-19.
69. Johnson, G., Wolfe, A. J., and Putonti, C. (2019). Characterization of the φ CTX-like *Pseudomonas aeruginosa* phage Dobby isolated from the kidney stone microbiota. *Access Microbiol.* 1. doi: 10.1099/acmi.0.000002.

70. Kakasis, A., and Panitsa, G. (2019). Bacteriophage therapy as an alternative treatment for human infections. A comprehensive review. *Int. J. Antimicrob. Agents* 53, 16–21. doi: 10.1016/j.ijantimicag.2018.09.004.
71. Kilcher, S., Studer, P., Muessner, C., Klumpp, J., and Loessner, M. J. (2018). Cross-genus rebooting of custom-made, synthetic bacteriophage genomes in L-form bacteria. *Proc. Natl. Acad. Sci.* 115, 567–572. doi: 10.1073/pnas.1714658115.
72. Klein, R. D., and Hultgren, S. J. (2020). Urinary tract infections: microbial pathogenesis, host-pathogen interactions and new treatment strategies. *Nat. Rev. Microbiol.* 18, 211–226. doi: 10.1038/s41579-020-0324-0.
73. Kondo, K., Kawano, M., and Sugai, M. (2021). Distribution of Antimicrobial Resistance and Virulence Genes within the Prophage-Associated Regions in Nosocomial Pathogens. *mSphere* 6, e00452-21. doi: 10.1128/mSphere.00452-21.
74. Kortright, K. E., Chan, B. K., Koff, J. L., and Turner, P. E. (2019). Phage Therapy: A Renewed Approach to Combat Antibiotic-Resistant Bacteria. *Cell Host Microbe* 25, 219–232. doi: 10.1016/j.chom.2019.01.014.
75. Kot, B. (2019). Antibiotic Resistance Among Uropathogenic *Escherichia coli*. *Pol. J. Microbiol.* 68, 403–415. doi: 10.33073/pjm-2019-048.
76. Kumar, M. S., and Das, A. P. (2017). Emerging nanotechnology based strategies for diagnosis and therapeutics of urinary tract infections: A review. *Adv. Colloid Interface Sci.* 249, 53–65. doi: 10.1016/j.cis.2017.06.010.
77. Lassmann, T., and Sonnhammer, E. L. (2005). Kalign – an accurate and fast multiple sequence alignment algorithm. *BMC Bioinformatics* 6, 298. doi: 10.1186/1471-2105-6-298.
78. Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301.
79. Lima-Mendez, G., Van Helden, J., Toussaint, A., and Leplae, R. (2008). Reticulate Representation of Evolutionary and Functional Relationships between Phage Genomes. *Mol. Biol. Evol.* 25, 762–777. doi: 10.1093/molbev/msn023.
80. Little, J. W. (2014). “Lysogeny, Prophage Induction, and Lysogenic Conversion,” in *Phages*, eds. M. K. Waldor, D. I. Friedman, and S. L. Adhya (Washington, DC, USA: ASM Press), 37–54. doi: 10.1128/9781555816506.ch3.
81. Long, B., and Koyfman, A. (2018). The Emergency Department Diagnosis and Management of Urinary Tract Infection. *Emerg. Med. Clin. North Am.* 36, 685–710. doi: 10.1016/j.emc.2018.06.003.

82. Łoś, J. M., Łoś, M., Węgrzyn, A., and Węgrzyn, G. (2010). Hydrogen peroxide-mediated induction of the Shiga toxinconverting lambdoid prophage ST2-8624 in *Escherichia coli* O157:H7. *FEMS Immunol. Med. Microbiol.* 58, 322–329. doi: 10.1111/j.1574-695X.2009.00644.x.
83. Łoś, J. M., Łoś, M., and Węgrzyn, G. (2011). Bacteriophages carrying Shiga toxin genes: genomic variations, detection and potential treatment of pathogenic bacteria. *Future Microbiol.* 6, 909–924. doi: 10.2217/fmb.11.70.
84. Maalouf, N. M., Cameron, M. A., Moe, O. W., Adams-Huet, B., and Sakhalee, K. (2007). Low Urine pH: A Novel Feature of the Metabolic Syndrome. *Clin. J. Am. Soc. Nephrol.* 2, 883–888. doi: 10.2215/CJN.00670207.
85. Malki, K., Shapiro, J. W., Price, T. K., Hilt, E. E., Thomas-White, K., Sircar, T., et al. (2016a). Genomes of Gardnerella Strains Reveal an Abundance of Prophages within the Bladder Microbiome. *PLOS ONE* 11, e0166757. doi: 10.1371/journal.pone.0166757.
86. Malki, K., Sible, E., Cooper, A., Garretto, A., Bruder, K., Watkins, S. C., et al. (2016b). Seven Bacteriophages Isolated from the Female Urinary Microbiota. *Genome Announc.* 4, e01003-16. doi: 10.1128/genomeA.01003-16.
87. Manrique, P., Bolduc, B., Walk, S. T., van der Oost, J., de Vos, W. M., and Young, M. J. (2016). Healthy human gut phageome. *Proc. Natl. Acad. Sci.* 113, 10400–10405. doi: 10.1073/pnas.1601060113.
88. Martín, M. C., Ladero, V., and Alvarez, M. A. (2006). PCR Identification of Lysogenic *Lactococcus lactis* Strains. *J. Für Verbraucherschutz Leb.* 1, 121–124. doi: 10.1007/s00003-006-0020-7.
89. McDonald, J. E., Smith, D. L., Fogg, P. C. M., McCarthy, A. J., and Allison, H. E. (2010). High-Throughput Method for Rapid Induction of Prophages from Lysogens and Its Application in the Study of Shiga Toxin-Encoding *Escherichia coli* Strains. *Appl. Environ. Microbiol.* 76, 2360–2365. doi: 10.1128/AEM.02923-09.
90. Miller-Ensminger, T., Garretto, A., Brenner, J., Thomas-White, K., Zambom, A., Wolfe, A. J., et al. (2018a). Bacteriophages of the Urinary Microbiome. *J. Bacteriol.* 200. doi: 10.1128/JB.00738-17.
91. Miller-Ensminger, T., Garretto, A., Brenner, J., Thomas-White, K., Zambom, A., Wolfe, A. J., et al. (2018b). Bacteriophages of the Urinary Microbiome. *J. Bacteriol.* 200. doi: 10.1128/JB.00738-17.
92. Miller-Ensminger, T., Garretto, A., Stark, N., and Putonti, C. (2020a). Mimicking prophage induction in the body: induction in the lab with pH gradients. *PeerJ* 8, e9718. doi: 10.7717/peerj.9718.

93. Miller-Ensminger, T., Garreto, A., Stark, N., and Putonti, C. (2020b). Mimicking prophage induction in the body: induction in the lab with pH gradients. *PeerJ* 8, e9718. doi: 10.7717/peerj.9718.
94. Miller-Ensminger, T., Mormando, R., Maskeri, L., Shapiro, J. W., Wolfe, A. J., and Putonti, C. (2020c). Introducing Lu-1, a Novel *Lactobacillus jensenii* Phage Abundant in the Urogenital Tract. *PLOS ONE* 15, e0234159. doi: 10.1371/journal.pone.0234159.
95. Monteiro, R., Pires, D. P., Costa, A. R., and Azereido, J. (2019). Phage Therapy: Going Temperate? *Trends Microbiol.* 27, 368–378. doi: 10.1016/j.tim.2018.10.008.
96. Nakamura, K., Ogura, Y., Gotoh, Y., and Hayashi, T. (2021). Prophages integrating into prophages: A mechanism to accumulate type III secretion effector genes and duplicate Shiga toxin-encoding prophages in *Escherichia coli*. *PLoS Pathog.* 17, e1009073. doi: 10.1371/journal.ppat.1009073.
97. Nanda, A. M., Thormann, K., and Frunzke, J. (2015). Impact of Spontaneous Prophage Induction on the Fitness of Bacterial Populations and Host-Microbe Interactions. *J. Bacteriol.* 197, 410–419. doi: 10.1128/JB.02230-14.
98. Navarro, F., and Muniesa, M. (2017). Phages in the Human Body. *Front. Microbiol.* 8. doi: 10.3389/fmicb.2017.00566.
99. Nelson, D. (2004). Phage Taxonomy: We Agree To Disagree. *J. Bacteriol.* 186, 7029–7031. doi: 10.1128/JB.186.21.7029-7031.2004.
100. Nelson, D. E., Van Der Pol, B., Dong, Q., Revanna, K. V., Fan, B., Easwaran, S., et al. (2010). Characteristic Male Urine Microbiomes Associate with Asymptomatic Sexually Transmitted Infection. *PLoS ONE* 5, e14116. doi: 10.1371/journal.pone.0014116.
101. Neugent, M. L., Hulyalkar, N. V., Nguyen, V. H., Zimmern, P. E., and De Nisco, N. J. (2020). Advances in Understanding the Human Urinary Microbiome and Its Potential Role in Urinary Tract Infection. *mBio* 11, e00218-20. doi: 10.1128/mBio.00218-20.
102. Oh, J.-H., Alexander, L. M., Pan, M., Schueler, K. L., Keller, M. P., Attie, A. D., et al. (2019). Dietary Fructose and Microbiota-Derived Short-Chain Fatty Acids Promote Bacteriophage Production in the Gut Symbiont *Lactobacillus reuteri*. *Cell Host Microbe* 25, 273–284.e6. doi: 10.1016/j.chom.2018.11.016.
103. O’Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, D733–D745. doi: 10.1093/nar/gkv1189.

104. Pearce, M. M., Hilt, E. E., Rosenfeld, A. B., Zilliox, M. J., Thomas-White, K., Fok, C., et al. (2014). The Female Urinary Microbiome: a Comparison of Women with and without Urgency Urinary Incontinence. *mBio* 5, e01283-14. doi: 10.1128/mBio.01283-14.
105. Pearce, M. M., Zilliox, M. J., Rosenfeld, A. B., Thomas-White, K. J., Richter, H. E., Nager, C. W., et al. (2015). The female urinary microbiome in urgency urinary incontinence. *Am. J. Obstet. Gynecol.* 213, 347.e1-347.e11. doi: 10.1016/j.ajog.2015.07.009.
106. Perez-Carrasco, V., Soriano-Lerma, A., Soriano, M., Gutiérrez-Fernández, J., and Garcia-Salcedo, J. A. (2021). Urinary Microbiome: Yin and Yang of the Urinary Tract. *Front. Cell. Infect. Microbiol.* 11, 617002. doi: 10.3389/fcimb.2021.617002.
107. Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE* 5, e9490. doi: 10.1371/journal.pone.0009490.
108. Price, T., Hilt, E., Thomas-White, K., Mueller, E., Wolfe, A., and Brubaker, L. (2020a). The urobiome of continent adult women: a cross-sectional study. *BJOG Int. J. Obstet. Gynaecol.* 127, 193–201. doi: 10.1111/1471-0528.15920.
109. Price, T. K., Dune, T., Hilt, E. E., Thomas-White, K. J., Kliethermes, S., Brincat, C., et al. (2016a). The Clinical Urine Culture: Enhanced Techniques Improve Detection of Clinically Relevant Microorganisms. *J. Clin. Microbiol.* 54, 1216–1222. doi: 10.1128/JCM.00044-16.
110. Price, T. K., Mehrtash, A., Kalesinskas, L., Malki, K., Hilt, E. E., Putonti, C., et al. (2016b). Genome sequences and annotation of two urinary isolates of *E. coli*. *Stand. Genomic Sci.* 11, 79. doi: 10.1186/s40793-016-0202-6.
111. Price, T. K., Wolff, B., Halverson, T., Limeira, R., Brubaker, L., Dong, Q., et al. (2020b). Temporal Dynamics of the Adult Female Lower Urinary Tract Microbiota. *mBio* 11, e00475-20. doi: 10.1128/mBio.00475-20.
112. Putonti, C., Malki, K., Garretto, A., Shapiro, J. W., Wolfe, A. J., and Brubaker, L. (2017). BACTERIAL VIRUSES IN THE FEMALE URINARY MICROBIOME. *Neurourol. Urodyn.* Vol. 36, S194–S195.
113. Raya, R. R., and H'bert, E. M. (2009). “Isolation of Phage via Induction of Lysogens,” in *Bacteriophages Methods in Molecular Biology.*, eds. M. R. J. Clokie and A. M. Kropinski (Totowa, NJ: Humana Press), 23–32. doi: 10.1007/978-1-60327-164-6_3.

114. Rodríguez-Rubio, L., Haarmann, N., Schwidder, M., Muniesa, M., and Schmidt, H. (2021). Bacteriophages of Shiga Toxin-Producing *Escherichia coli* and Their Contribution to Pathogenicity. *Pathogens* 10, 404. doi: 10.3390/pathogens10040404.
115. Roux, S., Enault, F., Hurwitz, B. L., and Sullivan, M. B. (2015). VirSorter: mining viral signal from microbial genomic data. *PeerJ* 3, e985. doi: 10.7717/peerj.985.
116. Roux, S., Krupovic, M., Daly, R. A., Borges, A. L., Nayfach, S., Schulz, F., et al. (2019). Cryptic inoviruses revealed as pervasive in bacteria and archaea across Earth's biomes. *Nat. Microbiol.* 4, 1895–1906. doi: 10.1038/s41564-019-0510-x.
117. Russo, T. A., and Johnson, J. R. (2000). Proposal for a New Inclusive Designation for Extraintestinal Pathogenic Isolates of *Escherichia coli*: ExPEC. *J. Infect. Dis.* 181, 1753–1754. doi: 10.1086/315418.
118. Santos, S. B., Carvalho, C., Azeredo, J., and Ferreira, E. C. (2015). Correction: Population Dynamics of a *Salmonella* Lytic Phage and Its Host: Implications of the Host Bacterial Growth Rate in Modelling. *PLOS ONE* 10, e0136007. doi: 10.1371/journal.pone.0136007.
119. Sarowska, J., Futoma-Koloch, B., Jama-Kmiecik, A., Frej-Madrzak, M., Ksiazczyk, M., Bugla-Ploskonska, G., et al. (2019). Virulence factors, prevalence and potential transmission of extraintestinal pathogenic *Escherichia coli* isolated from different sources: recent reports. *Gut Pathog.* 11, 10. doi: 10.1186/s13099-019-0290-0.
120. Schreiber, H. L., Conover, M. S., Chou, W.-C., Hibbing, M. E., Manson, A. L., Dodson, K. W., et al. (2017). Bacterial virulence phenotypes of *Escherichia coli* and host susceptibility determine risk for urinary tract infections. *Sci. Transl. Med.* 9, eaaf1283. doi: 10.1126/scitranslmed.aaf1283.
121. Seed, K. D. (2015). Battling Phages: How Bacteria Defend against Viral Attack. *PLOS Pathog.* 11, e1004847. doi: 10.1371/journal.ppat.1004847.
122. Shah, C., Baral, R., Bartaula, B., and Shrestha, L. B. (2019). Virulence factors of uropathogenic *Escherichia coli* (UPEC) and correlation with antimicrobial resistance. *BMC Microbiol.* 19, 204. doi: 10.1186/s12866-019-1587-3.
123. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303.
124. Shapiro, J. W., and Putonti, C. (2018a). Gene Co-occurrence Networks Reflect Bacteriophage Ecology and Evolution. *mBio* 9. doi: 10.1128/mBio.01870-17.
125. Shapiro, J. W., and Putonti, C. (2018b). Gene Co-occurrence Networks Reflect Bacteriophage Ecology and Evolution. *mBio* 9. doi: 10.1128/mBio.01870-17.

126. Shapiro, J. W., and Putonti, C. (2020). UPΦ phages, a new group of filamentous phages found in several members of Enterobacteriales. *Virus Evol.* 6, veaa030. doi: 10.1093/ve/veaa030.
127. Shoskes, D. A., Altemus, J., Polackwich, A. S., Tucky, B., Wang, H., and Eng, C. (2016). The Urinary Microbiome Differs Significantly Between Patients With Chronic Prostatitis/Chronic Pelvic Pain Syndrome and Controls as Well as Between Patients With Different Clinical Phenotypes. *Urology* 92, 26–32. doi: 10.1016/j.urology.2016.02.043.
128. Siddiqui, H., Nederbragt, A. J., Lagesen, K., Jeansson, S. L., and Jakobsen, K. S. (2011). Assessing diversity of the female urine microbiota by high throughput sequencing of 16S rDNA amplicons. *BMC Microbiol.* 11, 244. doi: 10.1186/1471-2180-11-244.
129. Simerville, J. A., Maxted, W. C., and Pahira, J. J. (2005). Urinalysis: a comprehensive review. *Am. Fam. Physician* 71, 1153–1162.
130. Sokol, N. W., Slessarev, E., Marschmann, G. L., Nicolas, A., Blazewicz, S. J., Brodie, E. L., et al. (2022). Life and death in the soil microbiome: how ecological processes influence biogeochemistry. *Nat. Rev. Microbiol.* doi: 10.1038/s41579-022-00695-z.
131. Song, W., Sun, H.-X., Zhang, C., Cheng, L., Peng, Y., Deng, Z., et al. (2019). Prophage Hunter: an integrative hunting tool for active prophages. *Nucleic Acids Res.* 47, W74–W80. doi: 10.1093/nar/gkz380.
132. Starikova, E. V., Tikhonova, P. O., Prianichnikov, N. A., Rands, C. M., Zdobnov, E. M., Ilina, E. N., et al. (2020a). Phigaro: high-throughput prophage sequence annotation. *Bioinformatics* 36, 3882–3884. doi: 10.1093/bioinformatics/btaa250.
133. Starikova, E. V., Tikhonova, P. O., Prianichnikov, N. A., Rands, C. M., Zdobnov, E. M., Ilina, E. N., et al. (2020b). Phigaro: high-throughput prophage sequence annotation. *Bioinformatics* 36, 3882–3884. doi: 10.1093/bioinformatics/btaa250.
134. Stone, E., Campbell, K., Grant, I., and McAuliffe, O. (2019). Understanding and Exploiting Phage–Host Interactions. *Viruses* 11, 567. doi: 10.3390/v11060567.
135. Subashchandrabose, S., and Mobley, H. L. T. (2015). Virulence and Fitness Determinants of Uropathogenic *Escherichia coli*. *Microbiol. Spectr.* 3, 3.4.20. doi: 10.1128/microbiolspec.UTI-0015-2012.
136. Suttle, C. A. (2007). Marine viruses — major players in the global ecosystem. *Nat. Rev. Microbiol.* 5, 801–812. doi: 10.1038/nrmicro1750.
137. Szafrański, S. P., Slots, J., and Stiesch, M. (2021). The human oral phageome. *Periodontol. 2000* 86, 79–96. doi: 10.1111/prd.12363.

138. Terlizzi, M. E., Gribaudo, G., and Maffei, M. E. (2017). UroPathogenic Escherichia coli (UPEC) Infections: Virulence Factors, Bladder Responses, Antibiotic, and Non-antibiotic Antimicrobial Strategies. *Front. Microbiol.* 8, 1566. doi: 10.3389/fmicb.2017.01566.
139. Thomas-White, K., Brady, M., Wolfe, A. J., and Mueller, E. R. (2016a). The Bladder Is Not Sterile: History and Current Discoveries on the Urinary Microbiome. *Curr. Bladder Dysfunct. Rep.* 11, 18–24. doi: 10.1007/s11884-016-0345-8.
140. Thomas-White, K., Forster, S. C., Kumar, N., Van Kuiken, M., Putonti, C., Stares, M. D., et al. (2018). Culturing of female bladder bacteria reveals an interconnected urogenital microbiota. *Nat. Commun.* 9, 1557. doi: 10.1038/s41467-018-03968-5.
141. Thomas-White, K. J., Hilt, E. E., Fok, C., Pearce, M. M., Mueller, E. R., Kliethermes, S., et al. (2016b). Incontinence medication response relates to the female urinary microbiota. *Int. Urogynecology J.* 27, 723–733. doi: 10.1007/s00192-015-2847-x.
142. Touchon, M., Bernheim, A., and Rocha, E. P. (2016). Genetic and life-history traits associated with the distribution of prophages in bacteria. *ISME J.* 10, 2744–2754. doi: 10.1038/ismej.2016.47.
143. Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., and Gordon, J. I. (2007). The Human Microbiome Project. *Nature* 449, 804–810. doi: 10.1038/nature06244.
144. Turner, D., Kropinski, A. M., and Adriaenssens, E. M. (2021). A Roadmap for Genome-Based Phage Taxonomy. *Viruses* 13, 506. doi: 10.3390/v13030506.
145. Wagner, P. L., and Waldor, M. K. (2002). Bacteriophage Control of Bacterial Virulence. *Infect. Immun.* 70, 3985–3993. doi: 10.1128/IAI.70.8.3985-3993.2002.
146. Waller, T. A., Pantin, S. A. L., Yenior, A. L., and Pujalte, G. G. A. (2018). Urinary Tract Infection Antibiotic Resistance in the United States. *Prim. Care Clin. Off. Pract.* 45, 455–466. doi: 10.1016/j.pop.2018.05.005.
147. Wang, X., and Wood, T. K. (2016). Cryptic prophages as targets for drug development. *Drug Resist. Updat.* 27, 30–38. doi: 10.1016/j.drup.2016.06.001.
148. Wendling, C. C., Refardt, D., and Hall, A. R. (2021). Fitness benefits to bacteria of carrying prophages and prophage-encoded antibiotic-resistance genes peak in different environments. *Evolution* 75, 515–528. doi: 10.1111/evo.14153.
149. Wolfe, A. J., Toh, E., Shibata, N., Rong, R., Kenton, K., FitzGerald, M., et al. (2012). Evidence of Uncultivated Bacteria in the Adult Female Bladder. *J. Clin. Microbiol.* 50, 1376–1383. doi: 10.1128/JCM.05852-11.

150. Wu, P., Chen, Y., Zhao, J., Zhang, G., Chen, J., Wang, J., et al. (2017). Urinary Microbiome and Psychological Factors in Women with Overactive Bladder. *Front. Cell. Infect. Microbiol.* 7, 488. doi: 10.3389/fcimb.2017.00488.
151. Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 13, 134. doi: 10.1186/1471-2105-13-134.
152. Zabavsky, E. R., Klein, G., and Winberg, G. (1993). λ SK diphasmids: phage lambda vectors for genomic, jumping, linking and cDNA libraries. *Gene* 127, 1–14. doi: 10.1016/0378-1119(93)90610-F.
153. Zalewska-Piątek, B., and Piątek, R. (2020). Phage Therapy as a Novel Strategy in the Treatment of Urinary Tract Infections Caused by *E. Coli*. *Antibiotics* 9, 304. doi: 10.3390/antibiotics9060304.
154. Zhang, H., Fouts, D. E., DePew, J., and Stevens, R. H. (2013). Genetic modifications to temperate *Enterococcus faecalis* phage ϕ Ef11 that abolish the establishment of lysogeny and sensitivity to repressor, and increase host range and productivity of lytic infection. *Microbiology* 159, 1023–1035. doi: 10.1099/mic.0.067116-0.
155. Zhang, Y., Liao, Y.-T., Salvador, A., Sun, X., and Wu, V. C. H. (2020). Prediction, Diversity, and Genomic Analysis of Temperate Phages Induced From Shiga Toxin-Producing *Escherichia coli* Strains. *Front. Microbiol.* 10, 3093. doi: 10.3389/fmicb.2019.03093.
156. Zhou, H., Chen, P., Zhang, M., Chen, J., Fang, J., and Li, X. (2021). Revealing the Viral Community in the Hadal Sediment of the New Britain Trench. *Genes* 12, 990. doi: 10.3390/genes12070990.

VITA

Elias D. Crum was born in Anchorage, Alaska and raised in Scandia, Minnesota. He attended Loyola University Chicago for his undergraduate degree, a Bachelor's of Science; he double majored in Bioinformatics and Biology.