2021

F Plasmids in Escherichia coli Decrease Permissivity to Coliphage

Cesar Eber Montelongo Hernandez

LOYOLA UNIVERSITY CHICAGO

F PLASMIDS IN ESCHERICHIA COLI

DECREASE PERMISSIVITY TO COLIPHAGE

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE GRADUATE SCHOOL

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

PROGRAM IN

MICROBIOLOGY AND IMMUNOLOGY

BY

CESAR E. MONTELONGO HERNANDEZ

CHICAGO, IL

AUGUST, 2021

ACKNOWLEDGEMENTS

I would like to thank the Loyola University of Chicago MD-PhD program for providing this opportunity to me. Admission and funding of these doctorate degrees would have been incredibly difficult as an undocumented immigrant, but you gave me the chance. Foremost I would like to thank the program directors, Charles Hemenway and Andrew Dingwall, for their continued support even when I steer off the beaten path.

I thank my lab members and all the people that helped me during graduate school. I thank my mentor Alan Wolfe for his patience and commitment to giving my ideas a chance. I thank my committee members Francis Alonzo, Thomas Gallagher, Bryan Mounce, and Catherine Putonti for having an open mind towards my project, especially in the more nebulous early days. I would like to thank my fellow grad student lab members (Krystal, David, Evann, Travis, Sarah, Baylie, Omar, Tom, Brian, Nancy, and Michael), especially for all the Dungeons and Dragons sessions. Thank you to the Putonti lab who adopted me for an entire summer. Thank you to all the medical residents that worked in our lab, especially Michelle, Petar, Ryan, and Chirag. Thank you to the techs and lab supporters across the years, especially Tommy, Mark, Carine, Bozena, Vibha, Bob, Dr. Hecht, and Roberto. Thank you to all the administrative people that make our work possible, especially Lishune Mahone, Donna Buczek, Jennifer Bojorquez, and Margarita Quesada.

Finally, thank you to my family who were patient and understanding both when I needed it and when I was not there. I never imagined all the changes that would occur in these four years, especially all the political and pandemic turmoil of the last one, and I would not be here without all of you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

16S RNA	The ribonucleic acid of the 16S subunit of a bacterial ribosome
Abi	Abortive infection systems
Am	Aminoglycoside
Ap	Ampicillin
ASB	Asymptomatic bacteriuria
Cm	Chloramphenicol
Col	Colicin plasmid
Coliphage	*E. coli* phage
CFU	Colony forming units
DMSO	Dimethyl sulfoxide
EIEC	Enteroinvasive *E. coli*
EQUC	Expanded Quantitative Urine Culture
Eval	E-value
F	Fertility plasmid
Fl	Fluoroquinolone
FUM	Female urobiome
HMP	Human Microbiome Project
Inc	Incompatibility
Kn	Kanamycin
LB	Lysogeny broth

LUTS	Lower urinary tract symptoms
Ma	Macrolide
MALDI-TOF	Matrix-assisted laser desorption ionization-time of flight
MGE	Mobile genetic element
MOI	Multiplicity of infection
Pe	Penicillin
PFU	Plaque forming unit
Phage	Bacteriophage
OAB	Overactive bladder
OD	Optical density
ORF	Open reading frame
Pident	Percent identity
Qcover	Query coverage
R	Resistance plasmid
Rep	Replication
SPA	Suprapubic aspiration
Spc	Spectinomycin
STEC	Shiga toxin-producing *E. coli*
Su	Sulfonamide
SUC	Standard urine culture
TA	Toxin-antitoxin
Tc	Tetracycline
TUC	Transurethral catheterization
Tr	Trimethoprim

UPEC	Uropathogenic *E. coli*
UMB	Urinary microbiota
Uropathogen	Urinary pathogen
Urobiome	Urinary microbiome
Urobiota	Urinary microbiota
Urotype	Urinary phylotype
UTI	Urinary tract infection
UUI	Urinary incontinence
WGS	Whole genome sequence

ABSTRACT

The urinary tract contains a community of bacteria called the urinary microbiota (urobiota) that may be relevant to health; the genomic component of the urobiota is the urinary microbiome (urobiome). Urinary bacteria have been associated with both asymptomatic states and disease conditions, such as urinary tract infection (UTI), overactive bladder (OAB), and urge urinary incontinence (UUI). Some bacteria, such as *E. coli*, are considered urinary pathogens (uropathogens) but also can be commensals. Bacteriophage (phage) are ubiquitous in nature and likely shape bacterial populations in every niche; thus, phage may be one factor that modulates the urobiota. Phages have a specific host range dictated not just by host receptor compatibility, but also by traits of the bacterial host. To understand the genetic determinants of phage infection in urinary bacteria, we have used a model system consisting of urinary *E. coli* and the lytic *E. coli* phages (coliphages). Urinary *E. coli* that are less permissive to coliphage infection often carry plasmid-related genes. To determine whether these genes relate to permissivity, plasmids present in urinary microbiota (UMB) were conjugated into a naïve *E. coli* K-12 background; *E. coli* K-12 acquisition of F plasmids from urinary isolates UMB0928 and UMB1284 decreased permissivity to infection by the lytic coliphages P1vir, Greed, and Lust. Analysis of the plasmidome of urinary *E. coli* indicated that more than half of these isolates are predicted to contain a plasmid; most of these urinary plasmids are of the F plasmid group. Antibiotic resistance and virulence genes were common in F plasmids. The F plasmids pU0928 and pU1284 reduced permissiveness to phage in *E. coli* K-12. These two plasmids were stable and conferred multiple antibiotic resistances

Given the selective pressure imposed by the rapid propagation and evolution of phages, plasmids could be a vehicle to deliver and maintain anti-phage genes in a bacteria population. Phage selective pressure also could result in the acquisition and maintenance of plasmid-linked content, such

as genes for antibiotic resistance and virulence factors. Urinary bacteria, phage, and plasmid dynamics could be important for clinically relevant traits of urinary bacteria and overall urobiota dynamics.

CHAPTER ONE:

BACKGROUND

The Urine is Sterile Dogma

The urinary tract contains a community of bacteria called the urinary microbiota (urobiota) that may be relevant to health; the urobiome is the genetic content of the urobiota, which may be relevant to urinary health[1]. For centuries, urine was assumed to be sterile in part due to the seminal Germ theory experiments by Dr. Louis Pasteur and Dr. William Roberts[2]. They showed that, unlike urine exposed to air, urine in a sealed container did not become cloudy with microbes. This led to the conclusion that, "…fresh and healthy urine is perfectly free from bacteria or other minute organisms." While Germ theory was fundamental to microbiology, Pasteur and Roberts did not set out to test the sterility of urine as a primary hypothesis, but rather these observations were byproducts of testing Germ theory[2].

The "urine is sterile" dogma has persisted perhaps due to a self-perpetuated belief in the absence of rigorous testing of what was assumed to be self-evident[1,3]. The dogma was unchallenged also in part due to the adoption of evidence and techniques not initially intended to support the sterility of urine. A prominent example of this was the development by Dr. Edward Kass of a culture method to distinguish UTI in patients with pyelonephritis. Under Dr. Kass' guidelines, uropathogens present greater than or equal to 10^5 colony forming units (CFU)/ml would indicate pyelonephritis over cystitis; this information could be applied to decrease post-operative sepsis in kidney surgery. Unfortunately, this technique was later adopted as a UTI diagnostic beyond this focused intent[4,5]; the practice continued despite later evidence that the threshold specified by Dr. Kass was not sufficient to specifically detect UTI[6–8]. Furthermore, Dr. Kass' culture test was specific for fast-growing, facultative aerobes, such as *E.*

coli, meaning that slow-growing anaerobes would likely not meet the positive culture threshold[9]. In this scenario, not only would the patient be deemed to not have a UTI, but their urine often would be qualified as sterile[9,10]. These diagnostic shortcomings continued until the 21st century and even now there is no standard threshold or test to report a true UTI[11].

The dogma that urine is sterile persisted despite evidence to the contrary, dating back to the initial Germ theory experiments of the 19th century. For example, Dr. Roberts showed that some urines could decompose (i.e. become cloudy and more alkaline) and bacteria be detectable; these samples were correlated to urinary tract symptoms and likely urinary infection[2]. In contrast, urine could be found in a non-decomposing state, remaining clear and acidic but with bacteria still detectable. This urine was from asymptomatic individuals and Dr. Roberts hypothesized that the bacteria originated not from infection but from the mucus membrane of the bladder[2]. Granted, despite the evidence, Roberts did not directly challenge the sterility of urine. A more direct challenge would not occur until the 1980s when Dr. Rosalind Maskell provided evidence of bacteria in suprapubic aspirate (SPA) urine samples from women with various urinary conditions outside of infection[12]. Dr. Maskell proposed that non-UTI urinary disorders could be due to dysbiosis of commensal flora and/or that uropathogens could induce disease that was not necessarily a UTI[12]. Unfortunately, Dr. Maskell's findings and hypotheses were rejected by the clinical community. However, Dr. Maskell would be vindicated in the 2010s when evidence increasingly supported the presence of a urinary bacterial community that could be associated with urinary disease[3,13–16].

The Discovery of the Urobiota/Urobiome

Recently, researchers have identified and characterized the urobiota/urobiome via enhanced culture methods and computational biology[1,17,18]. Viable bacteria have been isolated both from catheterized and voided urine, and grown in a reproducible manner under laboratory conditions[19,20]. Bacteria also have been characterized via computational biology, predominantly by metagenomic

sequencing (e.g., 16S RNA gene sequencing and shotgun metagenomic sequencing) and whole genome sequencing of individual bacteria isolates[21]. The biological niche of the urinary tract, specifically the bladder, has led to challenges, such as sterile sampling technique, low-biomass sampling, and amplification of genetic content for sequencing. The entrenched "urine is sterile" dogma also has placed the burden of proof on urobiome researchers, which is counter-intuitive given that urinary health and conditions, such as UTI, are understudied fields.

One of the highlights of early 21st century biomedical research was the recognition that the human microbiota are a key component of human physiology and health. Commensal bacteria in humans are estimated to rival the number of human cells in the body, with physiological functions that may be just as essential[1,22,23]. A cornerstone of microbiome research was the Human Microbiome Project (HMP), an NIH-funded effort to collect samples from various body sites (i.e., the gut, vagina, skin, mouth, and nasal cavity) and profile their bacterial content via metagenomic amplicon sequencing[24]. At present, the microbiota in these sites is accepted to influence health, with dysbiosis potentially being fatal (e.g., in the gut). During the research stages of the HMP, urine was sampled by the midstream voided collection method, and its bacterial content profiled, yet because urine was assumed to be sterile, these findings were not reported.

The urobiome was first reported by Nelson and co-workers, who collected midstream voided urines of men with and without STIs[25]. Urine genetic content was tested via 16S rRNA gene sequencing and DNA was detected, even in the absence of UTI. Soon afterwards, Wolfe and co-workers used 16S rRNA gene sequencing to identify bacterial DNA in the culture-negative urines of adult women undergoing urogynecological surgery[3]. A vital concern in this analysis was the possibility of vulvo-vaginal contamination. Thus, the authors assessed sampling technique efficacy and bacterial background by comparing multiple urine collection methods and adjacent anatomical sites. They sampled bladder urine using 1) suprapubic aspiration (SPA), which allows direct needle sampling of the bladder in an aseptic

manner and 2) transurethral catheterization (TUC), which involves passing a catheter through the urethra into the bladder. To control for non-bladder microbes during SPA sampling, they also 3) swabbed the skin where the needle was inserted and 4) obtained a sham needle stick which was inserted into the abdomen but did not enter the bladder. Finally, they obtained 5) a vaginal swab and 6) a midstream voided urine, the so-called "clean catch" method. The microbial profiles of the vaginal and midstream voided urine samples often resembled each other, evidence that the mid-stream voided urine was often contaminated with post-urethral (vulvo-vaginal) microbes. The microbial profiles of the SPA and TUC samples resembled each other, but these profiles differed from the skin, sham needle stick, vaginal, and midstream voided urine samples. Since the SPA sampling technique bypassed the vulva and vagina, the authors concluded that the SPA profile reflected the urobiome of the bladder. Since the TUC and SPA samples resembled each other, the authors also concluded that the TUC method of urine collection would suffice to obtain bladder urine for future studies[3].

Until this point, profiling was based on bacterial DNA detection in urine; therefore, an outstanding question was if this genetic content came from viable organisms[3]. Because the standard urine culture method developed by Dr. Kass and used by clinical laboratories was designed to grow fast-growing, non-fastidious, facultative anaerobes, it was imperative to develop a new culture method to unbiasedly grow the bacteria that could be present in urine. Thus, Hilt and co-workers established an enhanced urine culture (metaculturomic) method called Expanded Quantitative Urine Culture (EQUC), which could grow a wider range of urinary bacteria[17]. EQUC provided clear evidence of live bacteria in standard urine culture (SUC)-negative urine samples and of bacteria in urine even in the absence of a clinical UTI. Taken together, EQUC and 16S rRNA gene sequencing provided evidence of associations between the urobiota/urobiome and urinary disorders that had been assumed to be unrelated to bacteria.

Clinically Relevant Bacteria Species in the Urobiota

Hundreds of species have been identified in the urinary tract, though in terms of clinical relevance one can argue that some of the highest yields are *Aerococcus urinae, Corynebacterium amycolatum, Enterococcus faecalis, E. coli, Gardnerella vaginalis, K. pneumoniae, Lactobacillus species, Staphylococcus epidermidis, Streptococcus anginosus,* and *Streptococcus mitis*[18,26–30]. While most of these species can be present during asymptomatic periods, they can likewise be associated with urinary conditions such as OAB, UUI, and UTI[31–33]. *E. coli* is the premier uropathogen, but now other *A. urinae, C. amycolatum, E. faecalis, K. pneumoniae, S. epidermidis, S. anginosus,* and *S. mitis* have been associated with UTI, although all these can be present in the absence of symptoms, especially in the context of asymptomatic bacteriuria (ASB)[33–35]. In addition, *A. urinae* has been associated with OAB and *G. vaginalis* with UUI[17,26]. Often considered commensal bacteria, *S. epidermis* and *Streptococcus* species can be associated with urinary conditions as well[36,37]. Lactobacilli are generally thought of as protective bacteria, although they can be found in women with and without lower urinary tract symptoms (LUTS)[26]. *Lactobacillus jensenii* has even been shown to be protective against UTI[38]. In contrast, *Lactobacillus gasseri* is more frequently detected in women with UUI[26]. Pertinent questions are why and when these bacteria species are associated with urinary conditions, and what are the urobiota mechanisms that influence asymptomatic and symptomatic states.

Dynamics of the Urobiota/Urobiome

A recent study by Price and co-workers showed that the urobiota are not static, but rather seem to fluctuate between a small number of community state types (phylotypes or urotypes)[39,40]. This fluctuation can be influenced by the patient's physiology and health. For example, menstruation and vaginal intercourse in young asymptomatic adult women have been shown to influence the composition of the urobiota. The composition of the urobiota may affect urinary health given the individual species that make up that composition. *Lactobacillus crispatus* and *Lactobacillus iners* may be protective and

desirable in the urinary space, whereas *L. gasseri* has been associated with lower urinary tract symptoms[41]. Urobiota composition and species proportion fluctuations may play a role in potentiating urinary symptoms. It has been demonstrated that, in an asymptomatic individual, daily fluctuations can occur in the proportion of the bacteria isolated from urine[39,40] Potentially, the fluctuation of "good" and "bad" bacteria could influence symptoms or disease states[39,42].

The female urobiome (also known as the female urinary microbiome or FUM) has been studied in the context of urinary disorders with primary examples being OAB and UUI[33,35]. OAB consists of frequency of urination and nocturia; participants with OAB have been noted to have differences in their urobiota compared to non-OAB[17,43]. OAB participants had greater urobiome diversity and a larger proportion of *Gardnerella* bacteria[17,43]. UUI consists of involuntary urine leakage and a sense of urgency to urinate; participants with UUI and a detectable microbiome via 16S rRNA sequencing had more urgency episodes per day compared to those that were sequence-negative[31]. *L. crispatus* was associated with non-UUI whereas ten other species were associated with UUI[26]. In terms of other urinary conditions, it is worth noting that kidney stones, renal masses often composed of calcium-associated minerals, also have been associated with bacteria, specifically species known to be uropathogens[44].

A urinary disease of primary interest is UTI[45]. Under the premise that urine is sterile, UTI was primarily explained as caused by uropathogen invasion of the urinary space (e.g., moving from the gut)[46]. Existence of the urobiota adds complexity to UTI mechanism(s). While it may be true that some UTIs occur by uropathogen invasion, there is evidence that the urobiota itself could be a source of uropathogens[28]. For example, *E. coli* with virulence factors and antibiotic resistance has been identified in the bladder of women even in the absence of infection symptoms; these *E. coli* strains are genomically indistinguishable to *E. coli* in the bladder of women with UTI. Potentially, the inciting event for UTI could be fluctuations in the composition and proportions of the urobiota[39].

It is important to understand urobiota dynamics and determine the mechanisms that influence

fluctuations and population composition. Urobiota dynamics could be influenced by a multitude of factors; to generate hypotheses for these mechanisms, we can extrapolate from the major influencers of bacteria in niches worldwide. Bacteriophages (phages) are the top predator of bacteria and likely modulate these populations in every niche on the planet. Thus, phages likely influence the dynamics of urinary bacteria[20].

Phage Biology and Bacteria-Phage Dynamics

Phages are the most ubiquitous biological entity on Earth, shaping biomass daily on a planetary scale[47]. Phage infection begins with attachment to the bacterial host surface and injection of the phage genetic material into the cytosol[48]. Inside the host, the phage can undergo two general lifestyles, lytic or lysogenic. The lytic pathway begins immediately after phage genetic injection, with phage genome replication and gene expression inside the host, propagation and assembly of phage particles, and then lysis of the host so that particles can exit to the environment and renew the life cycle[48–50]. In contrast, the lysogenic lifestyle is characterized by integration of the phage genomic material into the host genome[51]. Depending on the type of phage, integration of the phage genome either involves site-specific recombination into the host genome (e.g., chromosome, plasmid) or circularization such that the phage genome replicates as a plasmid. The phage genome is now called a prophage and the host is called a lysogen[52]. Under the proper conditions, for some lysogens, often linked to stressors, the lytic lifestyle is induced for some phages (i.e., the prophage begins to replicate)[53]. Phages do not infect indiscriminately, but rather have a limited host range, largely dictated by the type of host receptor that allows the phage to adsorb to the host surface and inject its genome into the cell cytosol[50]. Also, phage infection is not a binary state (i.e., either infection or no infection), but rather a spectrum of permissiveness modulated by genetic determinants within both the host and the phage.

The phage life cycle results in a multitude of relevant bacteria-phage interactions[52,54–56]. The main implication of host lysis is that phages can potentially consume a population of bacteria[57]. This

cannot only decimate a species, but it can alter the overall microbial community by creating or destroying niches. In contrast to lysis, integration of phage into the host genome may not only delay killing of the host, but the host gains access to phage genetic material, which may provide biological advantages[51,52]. This effect can be biologically relevant, such as the transfer of the phage-coded Shiga toxin that converts a harmless *E. coli* into the deadly Shiga toxin-producing *E. coli* (STEC)[58]. For some phages, the assembly of phage particles may allow encapsulation of host genetic material, which can lead to horizontal transfer of traits between bacteria in a population (i.e., transduction)[59]. Expression of phage genes also may protect against infection by other phages (i.e., superinfection immunity) and lysogens may secrete phages to harm competitors[20,60].

Phages were identified in the five microbiome niches studied in the Human Microbiome Project[23]. Phages have been linked to pathology, such inflammatory bowel disease in the gut and periodontal disease in the mouth[61,62]. The most extensively studied bacteriophage community is that of the gastrointestinal tract, where phages are estimated to number 2 trillion[63]. The core phage community shared among individuals is called the phageome[64]. Gut phages have been linked to GI symptoms and disease, including Chron's disease and type 2 diabetes[61,65]. Despite their relevance to the microbiome, there is much we do not know about phages. This lack of information is due in part to the complexity of the phage life cycle and the variation of the structural and biological properties of phages. Sequencing technology has allowed the identification of phage sequences in microbiomes based on broad homology, but many phage gene sequences have not been annotated in databases[20,66]. Currently, only a small portion of phage genetic sequence material can be identified computationally, which limits our studies of phage in the microbiome[66].

Multiple laboratory methods have revealed phages in the urinary tract[19,29,67,68]. Phage have been identified from shotgun metagenomic sequence data of urine samples and by identification of prophages in individual bacterial genomes[19,29,67]. At the bench, phage have been found via phage

particle isolation from the urine and isolated as free-living phage by induction of prophage from bladder bacteria[68,69]. If bacteria exist in a niche, phage are likely preying on them.

Phage in the Urinary Tract

Phages have been isolated from the urine, dating back to 1917, when Felix d' Hérelle's initial phage experiments identified particles in the urine that could lyse the Shiga bacillus[70]. At present, urinary phages have been isolated for *Pseudomonas aeruginosa* and *E. coli*, both prominent uropathogens[68,71]. From urine obtained from adult women by transurethral catheterization, Putonti and co-workers isolated seven coliphages, some of which could infect strains of *E. coli* isolated from urine collected by catheter from adult female bladders[68]. The first metagenome analysis of bacteriophage was performed by Santiago-Rodriguez and co-workers, who aimed to identify free-living (extracellular) phage in the bladder, in addition to their presence relative to eukaryotic viruses and bacterial cells[19]. The total viral fraction, composed of eukaryotic and extracellular phage, was analyzed in catheterized and voided urine of 10 individuals diagnosed with UTI and 10 individuals asymptomatic for UTI. Extracellular phages were identified in the urine via metagenomic techniques; phages were considerably more abundant than either eukaryotic viruses or bacterial cells. Despite the abundance of phage in urine, only 27% of phage sequences identified were homologous to those present in public databases. A subsequent metagenomic study by Moustafa and co-workers studied the urinary microbiome of 49 individuals, including those with UTI[72]. Phage sequences were identified as homologous to viruses that infect the genera *Escherichia*, *Enterococcus*, *Lactobacillus* and *Pseudomonas*. A study by Putonti and co-workers corroborated this pattern by reconstructing viral genomes that were homologous to phage that infect the genera *Gardnerella*, *Lactobacillus*, and *Streptococcus*[73].

Phage integrated into the host chromosome (i.e., prophage) can be identified in genomes via sequencing of bacterial metagenome samples or from sequencing individual isolates[74]. Prophages were identified in the genomes of *Gardnerella* strains isolated from catheterized urine collected from adult

women with UUI[29]. A follow-up study analyzed these strains and publicly available *Gardnerella* strains isolated from the bladder and other sites; this study reported the presence of prophages and provided evidence of horizontal exchange of their genomes[75]. Lysogenic phage were present in 181 bacteria bacterial bladder isolates[75]. The authors identified more than 400 phage sequences in these 181 genomes; most genomes (86%) contained at least one prophage. Of note is the observation that at over half (57%) of phages identified had no homology to phage sequences in publicly database, highlighting how understudied this field of research is[20,75].

A challenge in studying urinary phage is that the urinary microbiome is a low biomass environment[9,17]. A metagenomic study attempted to solve the low DNA concentration issue by amplifying genetic material prior to sequencing; unfortunately, it was reported this method could potentially bias results[19,76,77]. Sample collection is a point of contention in urinary phage research. The studies of Santiago-Rodriguez et al., Rani et al., and Moustafa et al. recovered urine samples using multiple different techniques; however, for certain collection methods, such as that of voided urine, there are concerns of contamination with phages from other body sites[19,20,67,72]. Because catheterization retrieves urine from the bladder and does not contain post-urethral contamination, it is the best urine collection method for the study of bladder phages; however, its invasiveness limits the patient populations that can be studied[78].

Current expectations for how phage may affect human health are based on ongoing urobiota/urobiome research, findings from other microbiome sites, and knowledge of basic phage biology[75]. For example, the abundance of prophage in people with overactive bladder differs from those that are asymptomatic[75]. When assessing the abundance of extracellular phage in people with UTI, there was no significant variation with those that were asymptomatic[19]. Previous studies have noted that phage in the epithelium may protect against bacteria, in addition to enhancing the offensive potential of human cells against bacteria[79,80]. It has been reported that phage can enter eukaryotic cells and thus

access intracellular bacterial pathogens[81]. Phage may also be able to modulate human immune activity and even antagonize cancer growth[82]. The ability of phage to horizontally transfer genetic material between hosts (transduction) raises concerns about the transfer of virulent and fitness traits to uropathogens, but this genetic transduction associated with commensal and probiotic bacteria is still understudied[83,84]. Genetic factors in urinary bacteria that protect against phage infection could be incorporated into novel probiotics, especially if these traits are already endemic to the urobiome. Likewise, lytic phages are a promising tool that may be employed in treating UTIs. Cataloguing of phage populations and profiling of phage genetic content is ongoing, but unaddressed questions remain regarding the genetic determinants of phage-bacteria interactions[20].

Phage-bacteria interactions may influence the broad dynamics of the urobiota[62,64,75]. If the genetic content of phage is essential to understanding phage-bacteria dynamics, then it follows that the genetic content in bacteria that interacts with phage is also essential[48]. Currently, we understand very little about the traits in urinary bacteria that influence phage interactions. Depending on the genetic traits present in a bacteria strain, phage could interact with that bacteria population in wildly different ways: phage could prey on the bacteria, confer beneficial traits, transfer genetic material between cells, or protect from predation by another phage. Understanding the traits linked to phage-bacteria interactions may allow us to better understand the urobiota and the urobiota dynamics that impact health[85].

Urinary *E. coli*

Uropathogenic *E. coli* (UPEC) is the most common cause of UTI[86,87]. It is so common that it could be considered near pathognomonic for the condition, even despite accumulating evidence that several emerging pathogens are also associated with UTIs[12,33]. Approximately 150 million people per year suffer from a UTI, and UPEC has been estimated to account for over 70% of community-acquired UTIs[88,89]. UPEC's dominance belies our lack of understanding of its pathology in the urinary tract[28]. For decades

researchers have attempted to identify the UPEC "signature" that would unlock rapid diagnosis and the pathogenicity mechanism, but even now this goal is unachieved[90,91].

A logical solution to identify UPEC is to study its virulence factors and pathogenicity islands, yet the evidence indicates that the potential of an *E. coli* strain to be UPEC cannot be predicted by any genetic elements analyzed thus far[28,92,93]. Outside of genetic signatures, differential expression of genes involved in transport of potassium, nickel, and copper have been associated with UPEC, but there is no conclusive link to UPEC's etiology[94]. There is evidence that the gut can be a source of *E. coli* with the potential to cause UTI, which gave support to the hypothesis that gut UPEC invades the urinary tract[95]. The mechanisms of UPEC may be more complex, however, as *E. coli* is now recognized as a resident of the urobiota in some asymptomatic people[28]. Some asymptomatic people have *E. coli* in their urinary tract that code for virulence factors associated with UPEC potential[28]. This counters the notion that UPEC can only be an invader but most of all it goes against the dogma that the presence of *E. coli* in the urinary tract invariably results in infection and symptoms.

No genetic marker has been identified that differentiates urinary *E. coli* associated with UTI from *E. coli* present in asymptomatic people[28]. Likewise, the genomes of urinary *E. coli* from women diagnosed with UTI do not differ from women with OAB or UUI[28]. Given the conclusion that *E. coli*'s ability to cause disease cannot be solely linked to its genetic content, the impetus is in finding additional factors that could explain pathogenicity. Currently, it is hypothesized that the urinary tract itself could be a source of UPEC given specific conditions, such as fluctuations (dysbiosis) in the composition and counts of bacteria species in the urobiota[18,28].

Phage could be a factor that influences the composition and count of bacteria species in the urinary tract[69,75]. While the dynamics of coliphage and urinary *E. coli in vivo* have not been studied, there is enough evidence in other niches to infer their dynamic in the urinary tract[50]. The complexity of the phage life cycle allows for various interactions with bacteria, both beneficial and detrimental to the

host[55,96]. For example, lytic coliphage could modulate the overall *E. coli* population and/or disproportionally prey on non-UPEC strains. Coliphage could also lead to the acquisition of genetic content via transduction or lysogeny, thus increasing fitness and virulence in UPEC strains[51,74]. Finally, phage could be a factor in the dynamics of urobiome population flux by disrupting niches, creating a power vacuum for UPEC to exploit[50,97]. Fundamentally, we need to understand the interactions of *E. coli* and coliphage before we can assess their impact on the microbiome and associated pathology.

E. coli and Coliphage as a Model to Study Urinary Bacteria-Phage Interactions

Studying phage infection in *E. coli* necessitates the use of coliphages. Compared to lysogenic phages, lytic phages are advantaged in bench research since they can be more easily propagated, they can achieve higher titers, and they have a visually distinct clear phenotype during assays[96,98,99]. When exposed to a permissible host, lytic phage will decrease the turbidity of bacteria in liquid culture; when spotted on bacteria spread on media plates, lytic phage will result in clear spots/plaques. A prototypical lytic phage that can serve as a phenotypic example is the coliphage P1vir, commonly used in the laboratory setting[98,100]. P1 is a temperate coliphage of the *Myoviridae* family and *Punavirus* genus; P1 has been bioengineered for genetic techniques, used as a cloning vector and for generalized transduction[98,100]. P1vir is a P1 variant that lacks the integrase genes that would enable lysogeny and thus remains strictly lytic.

Outside of a lytic phage laboratory standard, a urinary *E. coli*-coliphage model necessitates the use of urinary lytic phages. The urinary lytic phages Greed and Lust were identified in urine from the bladder of adult females and their draft genomes published[68]. Greed and Lust are tailed phages (*Caudovirales*) hypothesized to be *Siphoviridae*. Despite morphological similarities and overlap in some genetic content, Greed and Lust grouped separately in terms of genomic homology when assessed with other coliphages isolated from bladder urine[68]. Greed and Lust are noted for robust propagation in permissive *E. coli*, resulting in lysate titers comparable to P1vir.

E. coli is very convenient in bench research as it has a relatively short doubling time, can grow well on commonly used laboratory media, and has a high number of well-optimized protocols and tools available[101]. *E. coli* is advantaged by decades of academic research, having its genome annotated, and having many of its genes validated biologically. *E. coli* is a preferred model system in large part due to its tractability with genetic techniques (e.g., transduction, transformation, conjugation) that allow consistent and rapid testing of hypotheses[101,102]. This ease and widespread use have led to powerful resources in the *E. coli* community, such as the ASKA collection (a large set of *E. coli* W3110 ORFs cloned into the vector pCA24n) and the KEIO collection (single-gene deletions of all nonessential genes in *E. coli* K-12)[103,104]. These two genetic collections allow for rapid, versatile, and pin-pointed testing of genetic-based hypotheses in *E. coli*.

Through an ongoing effort with Dr. Catherine Putonti, we have isolated 67 isolates of bladder *E. coli* from catheterized urine samples and sequenced their genomes[28,69]. These isolates were present in the urine of women with a UTI, other lower urinary tract symptoms, or without symptoms. The Putonti lab analyzed these *E. coli* genomes both in the context of bacterial comparative genomics and in terms of phage genetic content and abundance[28,69]. Whereas the Putonti lab has studied the genetic traits of phage that are important for infection, it is equally important to study the genetic content of *E. coli* that interact with phage[85]. Traditionally, it was thought that a phage's ability to infect was determined by receptor compatibility (i.e., adsorption), but the host range of a phage also can be dictated by genes in the host that antagonize the various steps of the phage life cycle[85].

E. coli Anti-phage Genes

Phage predation acts as a selective pressure on *E. coli*; therefore, there is an incentive for *E. coli* to obtain and retain genetic content that will either result in resistance or decrease permissivity to phage infection[85,105]. Antagonism of phage infection can occur at all the different steps of the phage life cycle, such as prior to adsorption, after phage genetic injection, or during hijacking of host machinery for

phage propagation[79,105–108]. A prototypical method to acquire phage infection resistance is mutation of the receptor for phage adsorption[109,110]. Phage recognition and binding of the adsorption receptor requires specific residues on surface proteins; therefore, changes in these can block adsorption and all subsequent phage life cycle steps. Access to the phage receptor can be altered in other ways as well, such as loss of a phage receptor gene or production of an extracellular matrix[109]. *E. coli* can also express traits that block receptors, such as the surface protein TraT; TraT is a plasmid-borne trait that mainly blocks binding of plasmid transfer factors to prevent plasmid invasion, but it also has been reported to block phage that utilize plasmid transfer surface receptors[107].

Even if the phage can adsorb and inject its genetic content, the bacterial host can still target the phage genetic content itself. Restriction enzymes can recognize sites predominant or specific to phage genetic sequences and result in phage genome cleavage[111]. Bacteria also code for CRISPR-Cas systems, which are adaptive immune defense systems that target invading genetic material[48]. Bacteria will maintain genetic spacers that can recognize the genetic sequence of past invaders; when the sequence from a spacer is complementary to a current invader, it will then complex with a Cas protein system with the help of a guide RNA[112]. The activated Cas complex will then be able to target and cleave the invading genetic content.

If the phage is successful in injecting its genetic content, subsequent life cycle steps may be inhibited by abortive infection systems (Abi)[105,108]. For example, a widespread and robust Abi are the toxin-antitoxin (TA) modules, such as *hok/sok* or *pemIK* to name two examples[113,114]. The primary role of TA modules is to stably maintain plasmids in the bacterial host. In unstressed conditions, the antitoxin will bind and neutralize the toxin; loss of the plasmid or stressors (e.g., starvation, phage infection) will result in degradation of the antitoxin by proteases. If the antitoxin no longer neutralizes the toxin, the latter will begin to negatively affect the host bacteria; toxins can work through various mechanisms, and effects can include degradation of RNA and inhibition of proteins essential for the cell life cycle. There is

evidence that the TA module itself does not mechanistically inhibit the phage, but rather it is hypothesized that the negative effects of the TA on the host's cellular stability leads to abortion of the phage life cycle[115].

Phages themselves can be a defense mechanism against phage infection. Prophage integrated in the host genome can express genes that can result in superinfection immunity or superinfection exclusion[60,116]. In superinfection immunity, the prophage expresses repressor proteins (e.g., Sim in P1 phage) to block post-adsorption steps of the phage life cycle, decreasing the frequency of co-infection or co-lysogeny[117]. Superinfection exclusion results in blocking of phage similar to the existing prophage; a classical example in *E. coli* is phage T4, which codes for the proteins Imm and Sp that inhibit injection of DNA by phages of the TF4 family[108].

An important question regarding anti-phage genes in bacteria is how these are maintained and transmitted in a bacteria population[108,118,119]. Ideally, these genes would be maintained in a stable manner even in the absence of selective pressure and could be easily propagated horizontally and vertically. Given these considerations, plasmids are a logical candidate for the maintenance and propagation of anti-phage genes[118,119].

The Plasmidome and the Microbiome

Plasmids are mobile genetics elements (MGE) that can be reservoirs and vectors for genetic content[120,121]. Plasmids can be considered as parasitic invaders and thus plasmids rely on various traits to be stably maintained in the host cell[122]. Plasmids can code for beneficial genetic content that will increase the fitness of the host and thus be maintained under selective pressure (e.g., genes for antibiotic resistance, virulence, or metabolism)[120,123]. Plasmids can also be stably maintained by expressing addiction systems, such as the *pemIK* TA module[124]. Conversely, plasmids can code for traits that will antagonize invasion by other plasmids, such as the protein TraT that blocks conjugation pili from attaching[107,125].

Plasmids utilize specific replication (Rep) and incompatibility (Inc) proteins for stable replication, partitioning, and thus inheritance[121,126]. Plasmids with similar Rep and Inc proteins cannot co-exist, as they will compete for these factors and eventually one of the plasmids will be lost[121]. Multiple plasmids can be retained in the same host, however, if they use different Rep and Inc proteins. Thus, Rep and Inc genes can be used to profile the plasmid content of a genome. Granted cataloguing plasmids in this manner is an approximation, as plasmids are highly heterogeneous and prone to genetic exchange with other plasmids, phage, and the host chromosome[121]. Given the wealth of information that they can encode, plasmid analysis can be a substantial source of information concerning the genetic and biological potential of bacteria.

The plasmidome is the collection of plasmids in a given sample; plasmidome research is a novel field, the term itself was coined less than a decade ago[127]. The plasmidome can be studied at the level of a single strain of a species (e.g., only isolates from *E. coli* O157:H7), multiple strains of a species (UPEC isolates) or an entire biological sample (e.g., urine from a patient)[127–130]. Over the last decade, there have been reports on the plasmidome of the microbiome, predominantly from the human gut, environmental samples, and samples relevant to industry[129,131]. To our knowledge, there are no reports on the plasmidome of the urobiome or even of just the *E. coli* strains of the urobiome. Plasmids are a key component of a bacteria's biological potential, allowing for rapid exchange and persistence of traits that may be essential under selective pressures, such as antibiotic exposure and metabolite availability[123,132]. At least *in vitro*, plasmids have been shown to impact phage-bacteria dynamics and this relationship also could exist in the microbiome[133].

Phage Antagonism by Bacteria Plasmids

An important factor to consider in bacteria-phage dynamics is that bacteria not only need anti-phage genes, but also require a method to quickly transmit and retain anti-phage genetic content in the bacteria population[50]. Plasmids represent a superb vehicle for the transfer, retention, and selection of

anti-phage genetic content[132,134,135]. Phage can evolve and disseminate rapidly, thus imposing a massive selective pressure on bacteria[69,96]. A parallel scenario to phage predation is bacterial exposure to antimicrobials; an ideal defense to antimicrobials would be for a bacterium to possess plasmid-borne antimicrobial resistance genes[123,136].

The relevance of plasmids to phage infection was reported in *in vitro* studies dating to the 1980s, though research interest seemed to have been intermittent until the 2010s; this trend parallels phage biology interest as a whole in these decades[110,119,137]. Plasmid-borne traits may decrease phage infection and likewise phage selective pressure may increase plasmid retention in bacteria[106,119]. In contrast, there is evidence that phage may decrease the retention of some plasmids; this is hypothesized to occur because the plasmid does not provide protection against the phage[138]. Studies indicate that the genes in plasmids that antagonize phage have varied mechanisms, including restriction enzymes, adsorption-blocking proteins (e.g., TraT) or TA systems[105–107]. Plasmids can affect phage predation dynamics and lead to the acquisition of plasmid-linked traits, such as antibiotic resistance[118,133].

While evidence that plasmids can be relevant to phage infection has been known for decades, these studies have been primarily done *in vitro*[105,118,119,133]. Furthermore, these studies do not consider bacteria-phage-plasmid dynamics in the setting of the microbiome; this is understandable as initial studies took place prior to the explosion of microbiome research that began in the 2000s. It remains pertinent to question whether plasmids can influence phage dynamics in a complex setting such as the urobiota, and to identify discrete genetic determinants in plasmids that influence phage infection.

E. coli Plasmids

There is a report on the plasmid content of an individual *E. coli* strain isolated from an individual suffering from a UTI[128], but the plasmidome of urinary *E. coli* as a group of strains has not been studied.

Furthermore, the plasmidome of urinary *E. coli* has not been studied in terms of its relevance to phage infection[128,129]

In general, *E. coli* plasmids are highly heterogeneous in composition with sizes that range from dozens to thousands of kilobases[121]. Despite their variant nature, plasmids in *E. coli* can be profiled in terms of their replicons (Rep and Inc groups), such as the IncFI and IncFII group (F plasmids), Col group (Colicin plasmids), and various others[121,139,140].

Characterizing the plasmidome of urinary *E. coli* would allow us to better understand the types of plasmids present, their associated genes, and their biological potential as it relates to key conditions, such as UTI[139,141]. The genetic content of a plasmid dictates not just effector traits, but also how it is maintained within its host and transferred either horizontally or vertically[89,123,136].

F plasmids in *E. coli* warrant special attention; F plasmids are large genetic elements (~100,000 bp), widespread, genetically stable, and able to transfer genes that confer antibiotic resistance, toxins, and fitness traits[123,134,136]. F plasmids code for transfer genes and can transmitted horizontally via conjugation[141]. F plasmids commonly code for at least one TA plasmid addiction system that increases plasmid retention even in the absence of selection pressure[139]. Finally, F plasmids are known to maintain antibiotic resistance, often for multiple genes, and thus be deemed as Resistance (R) plasmids with substantial clinical relevance[123].

Plasmids in *E. coli* are vectors and reservoirs for antibiotic resistance, toxins, fitness, and prophage genes[136,139]. Because all these genes are linked on the plasmid, selection by one factor could lead to the acquisition of genes outside the selection pressure (e.g., phage predation would result in the acquisition of plasmids with not just anti-phage genes but also antibiotic resistance genes)[142,143]. Phage-driven selection of bacteria could alter microbiome dynamics and urinary health. Furthermore, horizontal plasmid transfer could occur between *E. coli* and other Gram-negative species, affording

further complexity to the acquisition of anti-phage traits and the selective pressures relevant to the urobiome[144–146].

Summary of Introduction

Urinary health can be influenced by the aggregate profile of the urobiota or by individual bacteria species, such as uropathogens [3,17,33]. For example, the bacterium *E. coli* in the urinary tract could be asymptomatic or the etiological agent of infection, depending on factors like the overall composition of the urobiota[147–149]. Evidence exists that the urobiome can be dynamic, drastically changing daily by factors, such as menstruation or vaginal intercourse; there is interest in understanding the range of mechanisms that influence urobiome dynamics[39]. Phages are viruses of bacteria; they are the most abundant life form on the planet[20,150]. Phages shape biomass on a planetary scale daily from marine environments to the human microbiome[150]. Phages interact with bacteria in complex and multifaceted ways; phage can lyse bacteria populations, integrate into the host chromosome and provide traits, move genetic content horizontally within and across bacteria species, and protect bacteria from infection by other phages[52,54–56]. Phages have been identified in the urinary tract and could play a key role in shaping urobiota dynamics[19,63,73]. Phage-related genetic content in bacteria, such as genes that reduce permissibility to phage infection, may affect bacteria-phage interactions and thus the dynamics of the urobiota[52,57].

Very little is known concerning the genetic content of urinary microbiota that modulates phage infection or how this anti-phage genetic content is maintained and mobilized in the microbiota[85,108]. Bacteria can antagonize the phage life cycle by blocking interaction of phage with the host receptor (e.g., TraT), destroying injected phage genetic material (e.g., restriction enzymes), or interrupting the various steps necessary for phage propagation in the host cell (e.g., TA modules)[108]. Genes that disrupt the phage life cycle are not enough, however, as bacteria also require a mechanism to disseminate and maintain this genetic content[146,151]. Anti-phage genes have been identified in bacteria plasmids; bacteria

plasmids have been observed *in vitro* to influence phage predation and bacteria population dynamics[118,119]. Plasmids could be utilized by urinary bacteria to mobilize and maintain anti-phage genes that help modulate phage infection.

Here, I utilize a urinary *E. coli* and lytic coliphage model to study the genetic determinants in the *E. coli* host that modulate phage infection permissibility. Plasmids could be utilized by urinary *E. coli* to maintain and transfer anti-phage genes[106,119]. Because they are widespread, versatile, and heterogeneous, F plasmids are a noteworthy type of *E. coli* plasmids[142]. F plasmids code for their own conjugation machinery and often carry antibiotic resistance and virulence genes[141,142]. Phage predation is a major selective pressure on bacteria; predation by phage could lead to the acquisition of F plasmids with anti-phage content that contain other linked genes, such as antibiotic resistance and virulence genes. The interactions of bacteria, plasmids, and phage could be an important factor that shapes the urobiota and urinary health.

CHAPTER TWO:

METHODS

Urine Collection, Bacterial Culturing, DNA Extraction, and Genome Sequencing

This study uses the urinary *E. coli* isolates and sequence read data first published in Garretto et al. 2020 and genome data is found in NCBI BioProject PRJNA316969 (Table 1)[28]. Urinary isolates were recovered from urine samples in patients during several Institutional Review Board-approved studies at Loyola University Chicago (LU203986, LU205650, LU206449, LU206469, LU207102, and LU207152) and University of California San Diego (170077AW). Urine was collected via transurethral catheter and transferred to a BD Vacutainer Plus C&S preservative tube to be cultured. The culture technique and single colony isolation have been described[28]. Single isolated colonies were verified via Matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrophotometry. Stocks of each isolate were grown from pure colonies and frozen at −80°C in 900 µL 50% glycerol mixed with 90 µL bacteria cells.

For DNA extraction, bacteria were grown as liquid cultures in Lysogeny broth (LB) at 37°C in a shaking incubator for 12 hours. DNA was extracted via the Qiagen DNeasy UltraClean Microbial Kit using the standard manufacturer protocol. Qubit fluorometer was used to quantify the DNA concentration of DNA extractions. Nextera XT DNA library preparation kit was used to make DNA libraries, which were sequenced on the Illumina MiSeq platform using the MiSeq Reagent Kit v2 (500-cycles) in Loyola University Chicago's Genomics Facility (Maywood, IL, United States).

Table 1. Urinary *E. coli* Isolates Used in this Study.

Strain	Taxonomy	Participant condition	Assembly	Level	WGS	BioSample
UMB0103	*Escherichia coli*	UTI asymptomatic	GCA_003892645.1	Contig	RRWT00000000	SAMN09665164
UMB0149	*Escherichia coli*	UTI asymptomatic	GCA_003892555.1	Contig	RRWS00000000	SAMN09665165
UMB0276	*Escherichia coli*	UTI asymptomatic	GCA_003892545.1	Contig	RRWR00000000	SAMN09665166
UMB0527	*Escherichia coli*	UTI asymptomatic	GCA_003892535.1	Contig	RRWQ00000000	SAMN09665167
UMB0731	*Escherichia coli*	UTI asymptomatic	GCA_003892485.1	Contig	RRWP00000000	SAMN09665168
UMB0906	*Escherichia coli*	UTI symptomatic	GCA_003886695.1	Contig	RRWO00000000	SAMN09665169
UMB0923	*Escherichia coli*	UTI symptomatic	GCA_003892635.1	Contig	RRWN00000000	SAMN09665170
UMB0928	*Escherichia coli*	UTI asymptomatic	GCA_003892445.1	Contig	RRWM00000000	SAMN09665171
UMB0931	*Escherichia coli*	UTI symptomatic	GCA_003886495.1	Contig	RRWL00000000	SAMN09665172
UMB0933	*Escherichia coli*	UTI asymptomatic	GCA_003886675.1	Contig	RRWK00000000	SAMN09665173
UMB0934	*Escherichia coli*	UTI symptomatic	GCA_003892475.1	Contig	RRWJ00000000	SAMN09665174
UMB0939	*Escherichia coli*	UTI asymptomatic	GCA_003885295.1	Contig	RRUR00000000	SAMN09665218
UMB0949	*Escherichia coli*	UTI symptomatic	GCA_003892435.1	Contig	RRWI00000000	SAMN09665175
UMB1012	*Escherichia coli*	UTI symptomatic	GCA_003886455.1	Contig	RRWH00000000	SAMN09665176
UMB1091	*Escherichia coli*	UTI symptomatic	GCA_003886445.1	Contig	RRWG00000000	SAMN09665177
UMB1093	*Escherichia coli*	UTI symptomatic	GCA_003885215.1	Contig	RRUQ00000000	SAMN09665219
UMB1160	*Escherichia coli*	UTI symptomatic	GCA_003892605.1	Contig	RRWF00000000	SAMN09665178
UMB1162	*Escherichia coli*	UTI symptomatic	GCA_003892455.1	Contig	RRWE00000000	SAMN09665179
UMB1180	*Escherichia coli*	UTI symptomatic	GCA_008726795.1	Contig	VYWI00000000	SAMN12797014
UMB1193	*Escherichia coli*	UTI symptomatic	GCA_003892595.1	Contig	RRWC00000000	SAMN09665181
UMB1195	*Escherichia coli*	UTI symptomatic	GCA_008726745.1	Contig	VYWF00000000	SAMN12797017
UMB1202	*Escherichia coli*	UTI symptomatic	GCA_003886395.1	Contig	RRWA00000000	SAMN09665183
UMB1220	*Escherichia coli*	UTI symptomatic	GCA_003886385.1	Contig	RRVZ00000000	SAMN09665184
UMB1221	*Escherichia coli*	UTI symptomatic	GCA_003885055.1	Scaffold	RRUG00000000	SAMN10411422
UMB1223	*Escherichia coli*	UTI symptomatic	GCA_003886375.1	Contig	RRVY00000000	SAMN09665185
UMB1225	*Escherichia coli*	UTI symptomatic	GCA_008726695.1	Contig	VYWD00000000	SAMN12797019
UMB1228	*Escherichia coli*	UTI symptomatic	GCA_003886655.1	Contig	RRVW00000000	SAMN09665187

Table 1. Urinary *E. coli* Isolates Used in this Study (continued)

UMB1229	*Escherichia coli*	UTI symptomatic	GCA_003886345.1	Contig	RRVV00000000	SAMN09665188
UMB1284	*Escherichia coli*	UTI symptomatic	GCA_003892355.1	Contig	RRVU00000000	SAMN09665189
UMB1285	*Escherichia coli*	UTI symptomatic	GCA_003886635.1	Contig	RRVT00000000	SAMN09665190
UMB1335	*Escherichia coli*	UTI symptomatic	GCA_003886615.1	Contig	RRVS00000000	SAMN09665191
UMB1337	*Escherichia coli*	UTI symptomatic	GCA_003886325.1	Contig	RRVR00000000	SAMN09665192
UMB1346	*Escherichia coli*	UTI symptomatic	GCA_003886295.1	Contig	RRVQ00000000	SAMN09665193
UMB1347	*Escherichia coli*	UTI symptomatic	GCA_003886285.1	Contig	RRVP00000000	SAMN09665194
UMB1348	*Escherichia coli*	UTI symptomatic	GCA_003886275.1	Contig	RRVO00000000	SAMN09665195
UMB1354	*Escherichia coli*	UTI symptomatic	GCA_003886225.1	Contig	RRVN00000000	SAMN09665196
UMB1356	*Escherichia coli*	UTI symptomatic	GCA_003886245.1	Contig	RRVM00000000	SAMN09665197
UMB1358	*Escherichia coli*	UTI symptomatic	GCA_003886195.1	Contig	RRVL00000000	SAMN09665198
UMB1359	*Escherichia coli*	UTI symptomatic	GCA_003886185.1	Contig	RRVK00000000	SAMN09665199
UMB1360	*Escherichia coli*	UTI symptomatic	GCA_003886565.1	Contig	RRVJ00000000	SAMN09665200
UMB1362	*Escherichia coli*	UTI symptomatic	GCA_003886175.1	Contig	RRVI00000000	SAMN09665201
UMB1526	*Escherichia coli*	UTI symptomatic	GCA_003886105.1	Contig	RRVH00000000	SAMN09665202
UMB1727	*Escherichia coli*	UTI asymptomatic	GCA_003886135.1	Contig	RRVG00000000	SAMN09665203
UMB2019	*Escherichia coli*	UTI asymptomatic	GCA_003886115.1	Contig	RRVF00000000	SAMN09665204
UMB2055	*Escherichia coli*	UTI asymptomatic	GCA_003886095.1	Contig	RRVE00000000	SAMN09665205
UMB2321	*Escherichia coli*	UTI asymptomatic	GCA_003886555.1	Contig	RRVD00000000	SAMN09665206
UMB2328	*Escherichia coli*	UTI asymptomatic	GCA_003886545.1	Contig	RRVC00000000	SAMN09665207
UMB3538	*Escherichia coli*	UTI asymptomatic	GCA_003886535.1	Contig	RRVB00000000	SAMN09665208
UMB3641	*Escherichia coli*	UTI asymptomatic	GCA_003885305.1	Contig	RRUO00000000	SAMN09665221
UMB3643	*Escherichia coli*	UTI asymptomatic	GCA_003885095.1	Scaffold	RRUF00000000	SAMN10411421
UMB4656	*Escherichia coli*	UTI symptomatic	GCA_003886515.1	Contig	RRVA00000000	SAMN09665209
UMB4714	*Escherichia coli*	N/A	N/A	N/A	N/A	N/A
UMB4716	*Escherichia coli*	UTI asymptomatic	GCA_003885995.1	Contig	RRUN00000000	SAMN09665222
UMB4746	*Escherichia coli*	UTI asymptomatic	GCA_003886045.1	Contig	RRUZ00000000	SAMN09665210
UMB5337	*Escherichia coli*	UTI asymptomatic	GCA_003886035.1	Contig	RRUY00000000	SAMN09665211

Table 1. Urinary *E. coli* Isolates Used in this Study (continued).

UMB5814	*Escherichia coli*	UTI symptomatic	GCA_003886015.1	Contig	RRUX00000000	SAMN09665212
UMB5924	*Escherichia coli*	UTI asymptomatic	GCA_003886005.1	Contig	RRUW00000000	SAMN09665213
UMB5978	*Escherichia coli*	UTI asymptomatic	GCA_003885915.1	Contig	RRUV00000000	SAMN09665214
UMB6360	*Escherichia coli*	UTI asymptomatic	N/A	N/A	N/A	N/A
UMB6454	*Escherichia coli*	UTI symptomatic	GCA_003885245.1	Contig	RRUU00000000	SAMN09665215
UMB6611	*Escherichia coli*	UTI asymptomatic	GCA_003885875.1	Contig	RRUT00000000	SAMN09665216
UMB6653	*Escherichia coli*	UTI symptomatic	GCA_003885965.1	Scaffold	RRUS00000000	SAMN09665217
UMB6655	*Escherichia coli*	UTI asymptomatic	GCA_003885255.1	Contig	RRUL00000000	SAMN09665225
UMB6713	*Escherichia coli*	UTI asymptomatic	GCA_003885145.1	Contig	RRUK00000000	SAMN09665226
UMB6721	*Escherichia coli*	UTI symptomatic	GCA_003885125.1	Scaffold	RRUJ00000000	SAMN09665227
UMB6890	*Escherichia coli*	UTI asymptomatic	GCA_003885035.1	Contig	RRUI00000000	SAMN09665228
UMB7431	*Escherichia coli*	UTI symptomatic	GCA_003885225.1	Scaffold	RRUH00000000	SAMN09665229

Note: UTI symptomatic and asymptomatic denote the urinary condition of the participant from which the isolate was isolated.

Whole Genome Sequence Assembly, Gene Homology Scan, and Annotation

Raw sequence reads for the urinary *E. coli* isolates were deposited in NCBI's SRA database as described in Garretto et al. 2020[28]. In this work, raw sequence reads for 67 of the urinary *E. coli* isolates were trimmed using Sickle v1.33 and assembled using SPAdes v3.12 with k values of 55,77,99,127 and the only-assembler parameter[152,153]. Assemblies were renamed via a Bash script and contigs less than 500 bp were removed via bioawk. Assemblies were annotated using Prokka v1.14.5 with standard parameters in addition to parameters –centre XXX and -compliant[154]. Annotation output files were renamed and reorganized using a Bash script.

The literature was reviewed for genes associated with antagonism to the phage life cycle; these genes will hereafter be referred as anti-phage genes. The genes were binned into the following groups: abortive infection system, adsorption blocking, and *E. coli* receptors known to bind phage (Table 2). The amino acid sequence for each gene was download from NCBI's protein database (https://www.ncbi.nlm.nih.gov/nucleotide). Whole genome sequences were downloaded for the *E. coli* laboratory strains K-12 (U00096.2), B (NZ_CP014268.2), C (NZ_CP020543.1), and the UPEC strains CFT073 (AE014075.1), UTI89 (NC_007946.1), and NU14 (CP019777.1). The genome sequences of the *E. coli* urinary isolates, laboratory strains, and UPEC strains were uploaded to Geneious Prime v2019.0 and converted into a custom BLAST database[155]. The protein sequence of each anti-phage gene was queried against each *E. coli* BLAST database via tblastn[156]. Each query generated multiple hits in each database; for each hit, a homology score was generated that weighed the query's sequence identity, query coverage, and E-value. The top hit of each anti-phage query in each database was organized into a matrix. In addition, the amino acid sequence of *traT* and TA genes were queried into the *E. coli* databases (Table 3).

Table 2. Anti-phage Genes Queried in Urinary *E. coli*.

Anti-phage gene	Mechanism	GenBank accession
RNApol	Conserved *E. coli* gene (control)	AKK16086.1
rexA	Abortive infection system	WP_085543222.1
rexB	Abortive infection system	EOW06363.1
lit	Abortive infection system	NP_415657.1
prrC	Abortive infection system	WP_032251799
prrA	Abortive infection system	BAA97910.1
pifA	Abortive infection system	BAA97910.1
hok	Abortive infection system	AVQ79409.1
shok	Abortive infection system	ACM18292.1
mazE	Abortive infection system	P0AE72.1
mazF	Abortive infection system	P0AE70.1
imm	Adsorption blocking	AYH53380.1
sp	Adsorption blocking	NP_049651.1
llp	Adsorption blocking	YP_009031775.1
mccJ25	Adsorption blocking	Q9X2V7.1
traT	Adsorption blocking	ABC42217.1

Table 3. Receptors Known to Bind Coliphage.

Receptor gene	Phage that can bind	GenBank accession
fhuA	T1, T5, and Φ80	OAF98304.1
ompA	M1, Ox2, Tull	AXO86249.1
ompC	434, Me1, Tulb, T4	EGT66362.1
btuB	BF23	EGT70772.1
lamB	K10, Lambda	OMI64668.1
tolC	TLS	EGT68750.1
ompF	Tula, T2	ABJ00346.1
tonB	T1, φ80	ANK06875.1
tsx	T6	ANK05754.1
fadL	T2	ANK02657.1

Table 4. Toxin-Antitoxin Modules Queried in Urinary *E. coli*

TA factor	Operon	ID	Size (AA)	Mechanism for host antagonism
MazF	*mazEmazF*	NP_417262.1	111	Ribosome-independent RNA inteferases
ChpBK	*chpBchpBK*	NP_418646.1	116	Ribosome-independent RNA inteferases
HicA	*hicAhicB*	NP_415954.2	58	Ribosome-independent RNA inteferases
YhaV	*prlFyhaV*	NP_417599.1	154	Ribosome-independent RNA inteferases
MqsR	*mqsRmqsA*	NP_417494.1	98	Ribosome-independent RNA inteferases
RnlA	*rnlArnlB*	NP_417119.1	357	Ribosome-independent RNA inteferases
RelE	*relBrelE*	NP_416081.1	95	Ribosome-dependent RNA interfereses
YoeB	*yefMyoeB*	YP_588458.1	84	Ribosome-dependent RNA interfereses
YafO	*yafNyafO*	NP_414768.1	132	Ribosome-dependent RNA interfereses
YafQ	*dinJyafQ*	NP_414760.1	92	Ribosome-dependent RNA interfereses
HigB	*higBhigA*	NP_417554.1	104	Ribosome-dependent RNA interfereses
RatA	*ratAyfjF*	NP_417109.1	158	Inhibitor of ribosome subunit association
CbtA	*yeeUcbtA*	NP_416509.1	124	Inhibitors of cell division
Ykfl	*yafQykfl*	WP_016243829.1	113	Inhibitors of cell division
YfjF	*yjfZypjF*	SOR07133.1	122	Inhibitors of cell division
GnsA	*gnsAymcE*	CDU39217.1	57	Inhibitor of phospholipid synthesis
HipA	*hipBhipA*	AAA56878.1	440	Unknown
YjhX	*yjhXyjhQ*	OUR50196.1	85	Unknown
YdaS	*ydaSydaT*	NP_415875.1	98	Unknown
PemK	*pemIpemK*	AEE59819.1	133	Endoribonuclease
MazE	*mazEmazF*	NP_417263.1	82	Ribosome-independent RNA inteferases
ChpBI	*chpBchpBK*	NP_418645.2	83	Ribosome-independent RNA inteferases
HicB	*hicAhicB*	NP_415955.2	138	Ribosome-independent RNA inteferases
PrlF	*prlFyhaV*	NP_417598.1	111	Ribosome-independent RNA inteferases
MqsA	*mqsRmqsA*	NP_417493.1	131	Ribosome-independent RNA inteferases
RnlB	*rnlArnlB*	NP_417120.2	123	Ribosome-independent RNA inteferases
RelB	*relBrelE*	AZZ87905.1	79	Ribosome-dependent RNA interfereses
YefM	*yefMyoeB*	NP_416521.2	83	Ribosome-dependent RNA interfereses
YafN	*yafNyafO*	NP_414767.1	97	Ribosome-dependent RNA interfereses
DinJ	*dinJyafQ*	NP_414761.1	86	Ribosome-dependent RNA interfereses
HigA	*higBhigA*	NP_417553.1	138	Ribosome-dependent RNA interfereses
YfjF	*ratAyfjF*	NP_417132.1	105	Inhibitor of ribosome subunit association
YeeU	*yeeUcbtA*	ADD91700.1	122	Inhibitors of cell division
YafW	*yafQykfl*	QBM92663.1	105	Inhibitors of cell division
YpjZ	*yjfZypjF*	NP_417132.1	105	Inhibitors of cell division
YmcE	*gnsAymcE*	CAQ31517.1	76	Inhibitor of phospholipid synthesis
HipB	*hipBhipA*	NP_416025.1	88	Unknown
YjhQ	*yjhXyjhQ*	NP_418727.1	181	Unknown
YdaT	*ydaSydaT*	NP_415876.1	140	Unknown
PemI	*pemIpemK*	EFH4033665.1	85	Endoribonuclease

Plasmidic Assembly, Genomic and Gene Homology Scan, and Annotation

The raw sequence reads of the urinary *E. coli* isolates were assembled using plasmidspades.py of SPAdes v3.12 with k values of 55,77,99,127 and the only-assembler parameter[157]. Assemblies were renamed via a Bash script and contigs less than 500 bp were removed via bioawk. Plasmidic assemblies were BLASTed via NCBI BLAST and hits were binned as either *E. coli* plasmid or chromosome[158]. A homology heatmap of plasmidic assemblies was generated using sourmash v4.0 by generating signature files from the plasmidic assemblies, a signature index, and searching references in the index[159]. Plasmidic assemblies were annotated using Prokka v1.14.5 with standard parameters in addition to the parameters –centre XXX and -compliant[154]. Annotation output files from Prokka were renamed and reorganized using a Bash script. The open reading frames (ORF) in each *.faa file from all annotations were concatenated into a single file and sorted by length via the sortbyname.sh script from bbmap[160]. ORFs were clustered by homology using USEARCH v.11.0 with the -cluster-fast -id 0.8 -clusters parameters[161]. Files were renamed with the name of the top-most ORF in the file using a Bash script. Files were filtered using words related to plasmid content, such as “plasmid”, “toxin”, “resist”, “transfer”, “replicate”, “inc”, etc.

To identify known plasmid Inc and Rep genes, *E. coli* plasmidic assembly FASTA files were scanned using PlasmidFinder v2.1 (https://cge.cbs.dtu.dk/services/PlasmidFinder/) using the Enterobacteriaceae database with a threshold of 95% sequence identity and 60% minimum % coverage[162]. To identify known acquired antibiotic resistance genes, the FASTA files were scanned with ResFinder v4.1 (https://cge.cbs.dtu.dk/services/ResFinder/) using the “acquired antimicrobial resistance genes” option and *Escherichia coli* species database[163]. To identify known virulence genes, the FASTA files were scanned with VirulenceFinder v2.0 (https://cge.cbs.dtu.dk/services/VirulenceFinder/) using the *Escherichia coli* species database, with a sequence identity threshold of 90%, and a minimum sequence length of 60%[164].

Phage Propagation and Phage Infection Phenotype Testing

The lytic phages P1vir, Greed, and Lust were described in previous studies and used in this work[68,98]. The phage host was prepared by inoculating 5 ml of liquid broth media (TBT for P1vir, LB for Greed and Lust) with a single colony of *E. coli* K-12 MG1655 and grown until optical density (OD) ~0.4 in aerated conditions at 37°C (approximately 50 minutes for MG1655). For phage propagation, 100 ul of the respective phage at a titer of 10^{10} particle forming unit (PFU)/ml was pipetted into the culture tube and grown in aerated conditions at 37°C for five hours. The culture was transferred to a 15 ml conical centrifuge tube and 100 ul of chloroform was added. The tubes were centrifuged at 3000 rpm for 15 minutes, then the supernatant transferred to a new tube and centrifuged again. The supernatant was transferred to a final conical tube, wrapped in tin foil, and stored at –4 °C.

To quantify the titers of the phages, the full plate titer technique was used[165]. This method consists of (1) approximating the lowest concentration of phage that results in a cleared spot when the phage is placed on a host lawn and (2) then a second procedure to assess the plaque counts at that concentration. *E. coli* K-12 MG1655 was streaked onto an LB plate and grown overnight at 37°C to verify purity of the culture. Then, single colonies were grown in 5 ml liquid LB overnight in aerated conditions at 37°C. From each overnight culture, 100 ul were transferred to a new 5ml liquid LB tube subculture and grown in aerated conditions at 37°C until early exponential phase at OD ~0.4. From each subculture, 200 ul were transferred into a tube prepared with 0.7% agar LB media pre-heated to 52°C and immediately the tube was mixed and then poured and spread onto the surface of an 1.5% agar LB plate. Phage were titrated by 1:10 serial dilution in LB with an estimated starting concentration of 10^{10} PFU/ml and a final concentration of 10^{2} PFU/ml. Phage spots (10 uL spotted) were allowed to dry for 20 minutes and plates were incubated overnight at 37°C. The following day, the lowest concentration that resulted in clearance was noted. To count the number of plaques more accurately at this given phage titer, the process outlined above (streaking, culturing, subculturing K-12 MG1655) was repeated with

the following modification: after 200 ul of subculture was added to a tube prepared with 0.7% agar LB media pre-heated to 52°C, titrated phage also was added to the tube. The phage dilution used corresponded to the lowest concentration that resulted in clearance plus one phage dilution above and one below (for a total of three dilutions tested per phage titer quantified). Plates were incubated at 37°C; the next day, plaques on the plates were counted. The titer of the phage as PFU per ml was determined by the equation:

PFU/ml = (plaques on plate/volume of phage added)/(dilution factor of phage added)

The phage spot titration assay was used to test permissibility of *E. coli* isolates to P1vir, Greed, and Lust. The urinary, laboratory, and UPEC strains were re-streaked and grown overnight at 37°C to verify pure cultures. Single colonies per isolate were grown in in 5 ml liquid LB overnight in aerated conditions at 37°C. From each overnight culture, 100 ul were transferred to a new 5ml liquid LB tube subculture and grown in aerated conditions at 37°C until early exponential phase at OD ~0.4. From each subculture, 200 ul were transferred into a tube prepared with 0.7% agar LB media pre-heated to 52°C and immediately mixed and plated on a 1.5% agar LB plate. Overlaid plates were allowed to cool for 10 minutes and 10 ul of each phage was spotted with a pipette on the plate surface; spotting with liquid LB was used as a negative control. The phage spots were allowed to dry for 20 minutes at room temperature and the plates were incubated at 37°C overnight. The phage permissibility phenotype was assessed by observing the plates at all titrations and noting if there was a clear spot, turbid spot, or no spot.

Urinary Plasmid Conjugation in *E. coli*

The urinary *E. coli* isolates were grown on various antibiotic plates to assess their selection marker profile. LB plates were made with the following antibiotics: Ampicillin (Amp, 100 ug/ml), Chloramphenicol (Cam, 25 ug/ml), Kanamycin (Kan, 40 ug/ml), Spectinomycin (Spec, 100 ug/ml), and Tetracycline (Tet, 15 ug/ml). The urinary and laboratory *E. coli* isolates were streaked and incubated

overnight at 37°C; growth then was assessed as a binary (yes/no) as to whether colonies were present or not. The plasmidic assemblies of the urinary *E. coli* were reviewed for any ORF annotated as an antibiotic resistance gene and compared to growth on the selection plates.

Conjugation was utilized as the method to transfer plasmids from urinary *E. coli* to a naïve laboratory *E. coli* K-12 strain[166]. Urinary *E. coli* plasmid donor candidates were filtered on the following basis: they could grow on an antibiotic selection marker, they carried a gene that was predicted to encode antibiotic resistance, they had evidence of conjugation genes, and they had not been permissive to phage infection. See Table 5 for conjugation plasmid donors and recipients.

Isolates used in the conjugation assay were streaked on LB plates with the appropriate selection marker and incubated overnight at 37°C, then cultured with the appropriate selection maker in liquid LB media supplemented with an antibiotic (concentration noted above) and incubated overnight in aerated conditions at 37°C. For each overnight culture, 1 ml was transferred to a 1 ml microtube and centrifuged at 13500 rpm for one minute. The supernatant was discarded, and the pellet was resuspended in 1 ml of LB liquid and centrifuged at 13500 rpm for one minute; this was repeated to remove the rest of the antibiotic. To create the conjugation culture, 100 ul from the plasmid donor was mixed with 100 ul of the plasmid recipient; this suspension was vortexed and spotted onto an LB plate with no added selection marker. The unmixed plasmid donor and recipients also were spotted separately onto LB plates, as controls; all plates were incubated overnight at 37°C. Each overnight spot was scrapped from the LB plates and placed in a 1 ml microtube; the microtube was centrifuged at 1300 rpm for 1 minute. The supernatant was removed from the microtube and the pellet resuspended in 1 ml of LB liquid. From each microtube, a volume of 100 ul and 10 ul were plated into a respective LB plate with the appropriate selection marker (plasmid donor selection maker, plasmid recipient selection marker, double selection marker, LB plate). Plates were incubated overnight at 37°C. On the next day, colonies

were counted for all plates.

Table 5. Urinary *E. coli* Plasmid Recipient and Donors.

Strain	Background	Conjugation role	BLAST hit	Conjugation machinery on plasmid	camR on chromosome	tetR on plasmid assembly	Cm plate	Tet plate	Cam/Tet plate	Expected transconjugants on Cam Tet plates
AJW1776	MG1655	Plasmid recipient negative control	Chromosomal	No	No	No	No	No	No	No
AJW4793	MG1655 pCA24n-cm	Plasmid recipient	Chromosomal	No	Yes	No	Yes	No	No	Yes
AJW5116	MG1655 *yfiQ*::Cm, *cobB*::FRT	Plasmid recipient	Chromosomal	No	Yes	No	Yes	No	No	Yes
UMB1284	Urinary	Plasmid donor	Plasmid	Yes	No	Yes	No	Yes	No	Yes
UMB0928	Urinary	Plasmid donor	Plasmid	Yes	No	Yes	No	Yes	No	Yes
UMB1223	Urinary	Plasmid donor	Plasmid	Yes	No	Yes	No	Yes	No	Yes
UMB6721	Urinary	Plasmid donor	Plasmid	Yes	No	Yes	No	Yes	No	Yes
UMB1091	Urinary	Plasmid donor	Plasmid	Yes	No	Yes	No	Yes	No	Yes
UMB0939	Urinary	Plasmid donor negative control	Chromosomal	Yes	No	No	No	Yes	No	No
UMB1362	Urinary	Plasmid donor negative control	Chromosomal	Yes	No	No	No	Yes	No	No

From the double selection marker plates, single colonies were re-streaked onto fresh double selection marker plates and grown overnight at 37°C to verify purity. On the next day, single colonies grown on the second double selection marker plate were tentatively referred to as transconjugants; these were grown overnight in 5 ml LB liquid in the presence of both selection markers. From the overnight transconjugant cultures, 100 ul was transferred to a fresh 5 ml LB liquid supplemented with both selection markers and incubated until early exponential phase (OD ~0.4). From these subcultures, 900 ul was transferred into a cryovial tube with 100 ul dimethyl sulfoxide (DMSO) and placed at -80°C for long-term storage.

E. coli K-12 Constructs for Conjugation and Phage Permissiveness Phenotype Testing

Multiple *E. coli* K-12 MG1655 constructs were conjugated with urinary strains to test the effect of the urinary plasmid on phage infection permissiveness (Table 6). *E. coli* K-12 MG1655 with the empty vector pCA24n was used to select for the chloramphenicol resistance in the plasmid recipient; this construct would test the effect of the urinary plasmid on the phenotype in the absence of chromosomal

mutations. *E. coli* K-12 MG1655 with deletions of the genes *yfiQ* and/or *cobB* were used to select for antibiotic resistance present in the chromosome; these constructs would test the effect of the urinary plasmid in different genetic backgrounds related to protein acetylation. When possible, an existing K-12 construct in the Wolfe Lab collection were used. The pCA24n empty vector was purified from *E. coli* K-12 strain MG1655 (also known as AJW4793). The vector expressing *yfiQ* was purified from *E. coli* K-12 strain JW2568 carrying pCA24n-*yfiQ* of the ASKA collection.

To test the change in phage infection permissiveness of the K-12 transconjugants, the phage spot titration assay was utilized. *E. coli* transconjugants and controls were streaked from frozen stocks on the appropriate selection maker and incubated overnight at 37°C. Single colonies were used to inoculate 5 ml LB with the appropriate selection marker and incubated overnight in aerated conditions at 37°C. From each overnight culture, 100 ul was transferred to 5 ml of LB liquid with the appropriate antibiotic and incubated in aerated conditions at 37°C until early exponential phase (OD ~0.4). From each of these subcultures, 200 ul were transferred to a tube prepared with 0.7% agar LB media pre-heated to 52°C and immediately mixed and plated onto an 1.5% agar LB plate. Plates were allowed to cool for 10 minutes and spotted with 10 ul of diluted phage suspension. Phage was titrated by 1:10 serial dilution in LB with a starting concentration of 10^{10} PFU/ml and a final concentration of 10^{2} PFU/ml. Phage spots were allowed to dry for 20 minutes and plates were grown overnight at 37°C. The following day, phage spots were visualized; the lowest titration that resulted in clearance was noted; an integer was given to each titration based on the number of dilutions it was removed from the starting concentration (the lowest titration being one and the highest being eight). To assess the effect of the urinary plasmid on the phage permissiveness of the transconjugant *E. coli* K-12, growth curves were made as described here.

Table 6. *E. coli* K-12 Constructs Used as Plasmid Recipients.

E. coli strain	Construct	Mutation	ASKA plasmid	Marker	Reason for using
MG1655	AJW1776	WT	None	None	Wild type negative control (no selection marker)
MG1655	AJW4793	WT	pCA24n-Empty	Cm	Plasmid recipient, empty vector
MG1655	AJW5035	*yfiQ*::Kn	None	Kn	*yfiQ* deletion, no vector
MG1655	AJW1776	WT	pCA24n-*yfiQ*	Cm	*yfiQ* vector, overexpression
MG1655	AJW5035	*yfiQ*::Kn	pCA24n-*yfiQ*	Kn, Cm	*yfiQ* complement
MG1655	AJW5035	*yfiQ*::Kn	pCA24n-Empty	Kn, Cm	C*yfiQ* deletion control, empty vector
MG1655	AJW5184	*yfiQ*::Cm	None	Cm	*yfiQ* deletion, different selection marker
BW25113	AJW4688	*yfiQ*::Kn	None	Kn, Tc	*yfiQ* deletion, different K-12 background
MG1655	AJW5037	*cobB*::Cm	None	Cm	Different gene deletion
MG1655	AJW5116	*yfiQ*::Cm, *cobB*::FRT	None	Cm	No acetylation system (*yfiQ* and cob deletion)

Table 7. *E. coli* K-12 Constructs Made Carrying Urinary *E. coli* Plasmids.

E. coli strain	Construct	Mutation	ASKA plasmid	Urinary plasmid	Marker	Reason for using
MG1655	AJW4793	WT	pCA24n-Empty	pU0928	Cm, Tc	Empty vector, bladder plasmid
MG1655	AJW4793	WT	pCA24n-Empty	pU1091	Cm, Tc	Empty vector, bladder plasmid
MG1655	AJW4793	WT	pCA24n-Empty	pU1223	Cm, Tc	Empty vector, bladder plasmid
MG1655	AJW4793	WT	pCA24n-Empty	pU1284	Cm, Tc	Empty vector, bladder plasmid
MG1655	AJW4793	WT	pCA24n-Empty	pU6721	Cm, Tc	Empty vector, bladder plasmid
MG1655	AJW5035	*yfiQ*::Kn	pCA24n-*yfiQ*	pU0928	Cm, Kn, Tc	*yfiQ* complement, bladder plasmid
MG1655	AJW5035	*yfiQ*::Kn	pCA24n-Empty	pU0928	Cm, Kn, Tc	Empty vector, bladder plasmid
MG1655	AJW5184	*yfiQ*::Cm	None	pU0928	Cm, Tc	*yfiQ* deletion, different selection marker, bladder plasmid
BW25113	AJW4688	*yfiQ*::Kn	None	pU0928	Kn, Tc	*yfiQ* deletion, different K-12 background
MG1655	AJW5037	*cobB*::Cm	None	pU0928	Cm, Tc	Different gene deletion, bladder plasmid
MG1655	AJW5116	*yfiQ*::Cm, *cobB*::FRT	None	pU0928	Cm, Tc	No acetylation system, bladder plasmid
MG1655	AJW5116	*yfiQ*::Cm, *cobB*::FRT	None	pU1091	Cm, Tc	No acetylation system, bladder plasmid
MG1655	AJW5116	*yfiQ*::Cm, *cobB*::FRT	None	pU1223	Cm, Tc	No acetylation system, bladder plasmid
MG1655	AJW5116	*yfiQ*::Cm, *cobB*::FRT	None	pU1284	Cm, Tc	No acetylation system, bladder plasmid
MG1655	AJW5116	*yfiQ*::Cm, *cobB*::FRT	None	pU6721	Cm, Tc	No acetylation system, bladder plasmid

E. coli transconjugants and controls were streaked from frozen stocks on the appropriate selection maker and incubated overnight at 37°C. Single colonies were used to inoculate 10 ml LB with

the appropriate selection marker and incubated overnight in aerated conditions at 37°C. From each overnight culture, 1 ml was transferred to 25 ml of LB liquid in a 25 ml flask. The OD was measured to assess whether all cultures were approximately at the same cell density (OD ~0.2) and then subcultures were incubated in aerated conditions at 37°C until early exponential phase (OD ~0.4). Each phage (P1vir, Greed, Lust) was titrated and 0.5 ml was added to the flask to achieve a multiplicity of infection (MOI) of 0.0, 0.01, and 10.0. Cultures were grown in aerated conditions at 37°C for 8 hours with OD measured every hour. Each treatment in the growth curve was repeated in triplicate.

Bladder Plasmid Extraction, Sequencing, and Analysis

E. coli K-12 transconjugants are listed in Table 7 and those picked for sequencing were verified as *E. coli* by MALDI-TOF. For DNA extraction, bacteria were grown as liquid cultures in LB (with tetracycline as selection marker for the urinary plasmids) at 37°C in a shaking incubator for 12 hours. Whole genome DNA was extracted via the Qiagen DNeasy UltraClean Microbial Kit using the standard manufacturer protocol. Plasmid DNA was extracted via the Qiagen Large-Construct Kit using the standard manufacturer protocol. Qubit fluorometer was used to quantify the DNA concentration of DNA extractions. Nextera XT DNA library preparation kit was used to make DNA libraries, which were sequenced on the Illumina MiSeq platform using the MiSeq Reagent Kit v2 (500-cycles) in Loyola University Chicago's Genomics Facility (Maywood, IL, United States).

The raw sequence reads of the urinary *E. coli* isolates were assembled using plasmidspades.py of SPAdes v3.12 with k values of 55,77,99,127 and the only-assembler parameter[157]. Assemblies were renamed via a Bash script and contigs less than 500 bp were removed via bioawk. Plasmidic assemblies were BLASTed via NCBI (web) BLAST and hits were binned as either *E. coli* plasmid or chromosome. Plasmid assembly[158]. Plasmidic assemblies were annotated using Prokka v1.14.5 with standard parameters in addition to –centre XXX and -compliant[154]. Annotation output files from Prokka were renamed and reorganized using a Bash script. The ORF in each *.faa file from all annotations were

concatenated into a single file and sorted by length via the sortbyname.sh script from bbmap. ORFs were clustered by homology using USEARCH v.11.0 with the -cluster-fast -id 0.8 -clusters parameters[161]. The genetic content of the curated plasmid assemblies (e.g. pruned of chromosomal contigs) was scanned with PlamsidFinder, ResFinder, and VirulenceFinder[162–164]. Raw sequence reads were mapped to the curated plasmid assemblies via the BBmap plugin in Geneious[155,160].

Estimates of ORFs assigned a known function or hypothetical function and lists of distinct ORFs in a plasmid were made via Bash scripts. ORF lists were reviewed for ORFs sequences shared by plasmids pU0928, pU1223, and pU1284 (0.8 amino acid sequence identity threshold). The three ORFs shared by pU0928 and pU1284 were scanned via web BLAST, protein domains were reviewed, and aligned to assess ORF homology between plasmids. The three ORFs shared by the anti-phage plasmids (pU0928 and pU1284) were queried in all urinary *E. coli* plasmidic assemblies via Local BLAST to assess presence (sequence identity >90%, query coverage >90%)[156]. Urinary *E. coli* plasmidic assemblies with any of the three ORFs were concatenated into a single file with a Bash script and compared for overall homology via sourmash[159].

Phage-like Genetic Content in Urinary *E. coli* Anti-phage Plasmids

A phylogenetic tree of phage integrase sequences was made to assess the relationship of the urinary *E. coli* phage integrase in pU0928 and pU1284 to phage integrases in the NCBI database. Phage integrase amino acid sequences were obtained via web BLAST hits from the pU1284 phage integrase, random searches for “phage-integrase” in the NCBI gene database (either plasmid, phage, or whole-genome), and homologs of the Lambda phage integrase. Phage integrase sequences of bacteria whole genome origin were EDU65901.1 (*E. coli*), EDV61695.1 (*E. coli*), KMG57351.1 (*K. pneumoniae*), WP_000845048.1 (Multispecies), WP_097473620.1 (*E. coli*), WP_197577570.1 (*Comamonas thiooxydans*), SAY12656.1 (*K. pneumoniae*), EBZ3358802.1 (*Salmonella*), CAA11470.1 (*Pseudomonas aeruginosa*), EFW2126725.1 (*Shigella oydii*), AVV61735.1 (*Serratia proteamaculans*). Phage integrase

sequences of phage origin were P03700 (Lambda phage), A0A346FJ43 (*Enterobacteria* phage), KRM93595.1 (*Lactobacillus senioris* phage), K7P8A1 (*Escherichia* phage), K7PMH8 (*Escherichia* phage), BAB48054.1 (*Mesorhizobium japonicum* phage), AGF84336.1 (*Salmonella enterica* phage), ENO07474.1 (*E. coli* phage), VBI72162.1 (*Burkholderia pseudomallei*), YP_009823978.1 (*Erwinia phage*), P21442.1 (VINT_BPHC1 phage), NP_680502.1 *(Lactobacillus* phage A2), VVL20124.1 (*Vibrio* phage), P08320.2 (VINT_BPP4), P27077.1 (VINT_BPP21), QBQ72150.1 (*Serratia* phage Parlo), CAL11869.1 (*Yersinia enterocolitica* phage). Phage integrase sequences of plasmid origin were pU0928 (*E. coli*), pU1284 (*E. coli*), BAW89208.1 (*E. coli*), ABC42260.1 (*E. coli*), ASI37956.1 (*E. coli*), CAX66675.1 (*Lactobacillus johnsonii*), QJS06527.1 (*Arthrobacter* sp.), ABG11677.1 (*Mycobacterium* sp. MCS), ABV95733.1 (*Dinoroseobacter shibae*), CEK42513.1 (*Pseudomonas fluorescen*), ADY68011.1 (*Agrobacterium* sp), AIW54703.1 (*Clostridium botulinum*), ACF28473.1 (*Azospirillum baldaniorum*), ABG65572.1 (*Chelativorans* sp. BNC1), ABA24901.1 (*Trichormus variabilis*). Phage integrases from pU0928 and pU1284 have a two amino acid difference (P to L at position 228, V to F at position 232).

pU0928, pU1223, and pU1284 were scanned for phage genetic content via PHAST and PHASTER using default settings[167,168]. Phage sequence maps generated by PHASTER were downloaded. Phage-like sequences predicted by PHAST and PHASTER were aligned to assess if there were redundant sequences, which were pruned. The phage-like sequences were compared to one another on overall sequence homology via sourmash[159]. The phage-like sequences from pU0928, pU1223, and pU1284 were annotated using Prokka v1.14.5 with standard parameters in addition to –centre XXX and -compliant[154]. Annotation output files from Prokka were renamed and reorganized using a Bash script. The ORF in each *.faa file from all annotations were concatenated into a single file and sorted by length via the sortbyname.sh script from bbmap[160]. ORFs were clustered by homology using USEARCH v.11.0 with the -cluster-fast -id 0.8 -clusters parameters[161]. Lists of distinct ORFs in a plasmid were made via Bash scripts and reviewed for ORFs sequences shared by plasmids pU0928, pU1223, and pU1284 (0.8 amino acid

sequence identity threshold). Urinary *E. coli* plasmidic assemblies predicted to have the phage integrase ORF shared by pU0928 and pU1284 were scanned via PHASTER[168].

Plasmid Analysis of Urinary Microbiome Species

WGS assemblies were obtained for various urinary species (*Aerococcus urinae, Corynebacterium amycolatum, Enterococcus faecalis, Gardnerella vaginalis, Klebsiella pneumoniae, L. gasseri, Lactobacillus jensenii, Staphylococcus epidermidis, Streptococcus anginosus,* and *Streptococcus mitis*) from NCBI BioProject PRJNA316969 (Table 8). Contigs were scanned with PlasmidFinder to identify *inc* and *rep* genes[162]. For species with evidence of plasmid content like that in urinary *E. coli*, raw sequence reads were assembled using plasmidspades.py of SPAdes v3.12 with k values of 55,77,99,127 and the only-assembler parameter[157]. Assemblies were renamed via a Bash script and contigs less than 500 bp were removed via bioawk. Plasmidic assemblies were BLASTed via NCBI (web) BLAST and hits were binned as either plasmid or chromosome[158]. The curated plasmid assembly consisted of only contigs with homology to plasmids in the NCBI database.

Plasmidic assemblies were annotated using Prokka v1.14.5 with standard parameters in addition to –centre XXX and -compliant[154]. Annotation output files from Prokka were renamed and reorganized using a Bash script. Annotated ORFs were filtered based on being involved in plasmid transfer, plasmid replication, and virulence. Plasmid assemblies were scanned for phage content via PHASTER[168]. Plasmid assemblies were compared to urinary *E. coli* plasmid assemblies via sourmash[159].

Table 8. Urinary Species Analyzed for Plasmid Content.

Strain	Taxonomy	# Assembly	Level	WGS	BioSample
UMB637	*Aerococcus urinae*	GCA_008726885.1	Contig	VYWL00000000	SAMN12797010
UMB970	*Aerococcus urinae*	GCA_008726845.1	Contig	VYWK00000000	SAMN12797012
UMB2126	*Aerococcus urinae*	GCA_008726675.1	Contig	VYWA00000000	SAMN12797023
UMB7049	*Aerococcus urinae*	GCA_008726475.1	Contig	VYVT00000000	SAMN12797031
UMB8614	*Aerococcus urinae*	GCA_008726385.1	Contig	VYVN00000000	SAMN12797038
UMB8662	*Aerococcus urinae*	GCA_008726315.1	Contig	VYVK00000000	SAMN12797041
UMB8711	*Aerococcus urinae*	GCA_008726285.1	Contig	VYVI00000000	SAMN12797043
UMB9184	*Corynebacterium amycolatum*	GCA_008726215.1	Contig	VYVF00000000	SAMN12797046
UMB9256	*Corynebacterium amycolatum*	GCA_008726175.1	Contig	VYVD00000000	SAMN12797048
UMB7760	*Corynebacterium amycolatum*	GCA_008726455.1	Contig	VYVQ00000000	SAMN12797034
UMB1310	*Corynebacterium amycolatum*	GCA_008726645.1	Contig	VYWB00000000	SAMN12797022
UMB1182	*Corynebacterium amycolatum*	GCA_008726785.1	Contig	VYWH00000000	SAMN12797015
UMB7780	*Enterococcus faecalis*	GCA_012030205.1	Contig	JAAUVY000000000	SAMN14478493
UMB0843	*Enterococcus faecalis*	GCA_012030565.1	Contig	JAAUWL000000000	SAMN14478480
UMB1309	*Enterococcus faecalis*	GCA_012030535.1	Contig	JAAUWG000000000	SAMN14478485
UMB0768	*Gardnerella vaginalis*	GCA_013315255.1	Scaffold	JABUGZ000000000	SAMN15064053
UMB1686	*Gardnerella vaginalis*	GCA_013315215.1	Contig	JABUHA000000000	SAMN15064054
UMB0264	*Gardnerella vaginalis*	GCA_013315195.1	Scaffold	JABUHC000000000	SAMN15064056
UMB0170	*Gardnerella vaginalis*	GCA_013315145.1	Scaffold	JABUHD000000000	SAMN15064057
UMB1642	*Gardnerella vaginalis*	GCA_013315135.1	Contig	JABUHE000000000	SAMN15064058
UMB7783	*Klebsiella pneumoniae*	GCA_012030295.1	Contig	JAAUVW000000000	SAMN14478495
UMB8492	*Klebsiella pneumoniae*	GCA_012030275.1	Contig	JAAUVU000000000	SAMN14478498
UMB7779	*Klebsiella pneumoniae*	GCA_012030245.1	Contig	JAAUVZ000000000	SAMN14478492
UMB1399	*Lactobacillus gasseri*	GCA_007826985.1	Scaffold	VNGD00000000	SAMN12277447

Table 8. Urinary Species Analyzed for Plasmid Content (continued).

UMB4205	*Lactobacillus gasseri*	GCA_007786065.1	Contig	VNFS00000000	SAMN12277458
UMB3077	*Lactobacillus gasseri*	GCA_007786195.1	Scaffold	VNGC00000000	SAMN12277448
UMB2965	*Lactobacillus gasseri*	GCA_007785975.1	Contig	VNFT00000000	SAMN12277457
UMB1196	*Lactobacillus gasseri*	GCA_007785965.1	Contig	VNFU00000000	SAMN12277456
UMB0607	*Lactobacillus gasseri*	GCA_007785995.1	Scaffold	VNFY00000000	SAMN12277452
UMB0055	*Lactobacillus jensenii*	GCA_007786035.1	Scaffold	VNFZ00000000	SAMN12277451
UMB1307	*Lactobacillus jensenii*	GCA_007786135.1	Scaffold	VNGG00000000	SAMN12277444
UMB7765	*Staphylococcus epidermidis*	GCA_012030625.1	Contig	JAAUWD000000000	SAMN14478488
UMB8493	*Staphylococcus epidermidis*	GCA_012029805.1	Contig	JAAUVT000000000	SAMN14478499
UMB1227	*Staphylococcus epidermidis*	GCA_008728125.1	Contig	VYYY00000000	SAMN12797020
UMB626	*Staphylococcus epidermidis*	GCA_008726905.1	Contig	VYWN00000000	SAMN12797008
UMB593	*Staphylococcus epidermidis*	GCA_008727015.1	Contig	VYWS00000000	SAMN12797002
UMB1201	*Staphylococcus epidermidis*	GCA_008726735.1	Contig	VYWE00000000	SAMN12797018
UMB1188	*Staphylococcus epidermidis*	GCA_008726775.1	Contig	VYWG00000000	SAMN12797016
UMB7759	*Staphylococcus epidermidis*	GCA_008726465.1	Contig	VYVR00000000	SAMN12797033
UMB9183	*Staphylococcus epidermidis*	GCA_008726255.1	Contig	VYVG00000000	SAMN12797045
UMB7768	*Streptococcus anginosus*	GCA_012030235.1	Contig	JAAUWB000000000	SAMN14478490
UMB0839	*Streptococcus anginosus*	GCA_012030555.1	Contig	JAAUWM000000000	SAMN14478479
UMB1296	*Streptococcus anginosus*	GCA_012030005.1	Contig	JAAUWH000000000	SAMN14478484
UMB248	*Streptococcus anginosus*	GCA_008727075.1	Contig	VYWV00000000	SAMN12796999
UMB567	*Streptococcus anginosus*	GCA_008727045.1	Contig	VYWU00000000	SAMN12797000
UMB8710	*Streptococcus anginosus*	GCA_008726305.1	Contig	VYVJ00000000	SAMN12797042
UMB8390	*Streptococcus anginosus*	GCA_008726415.1	Contig	VYVP00000000	SAMN12797036
UMB8616	*Streptococcus anginosus*	GCA_008726365.1	Contig	VYVM00000000	SAMN12797039
SM50	*Streptococcus mitis*	GCA_009496905.1	Contig	WIJB00000000	SAMN13105985
SM49	*Streptococcus mitis*	GCA_009496895.1	Contig	WIJC00000000	SAMN13105984
SM18	*Streptococcus mitis*	GCA_009496835.1	Contig	WIIZ00000000	SAMN13105987
SM17	*Streptococcus mitis*	GCA_009496825.1	Contig	WIJA00000000	SAMN13105986
SM36	*Streptococcus mitis*	GCA_009496815.1	Contig	WIIW00000000	SAMN13105990
SM19	*Streptococcus mitis*	GCA_009496805.1	Contig	WIIY00000000	SAMN13105988
SM37	*Streptococcus mitis*	GCA_009496795.1	Contig	WIIV00000000	SAMN13105991
SM42	*Streptococcus mitis*	GCA_009496755.1	Contig	WIIU00000000	SAMN13105992

CHAPTER THREE:

EXPERIMENTAL RESULTS

Anti-phage Genes in Urinary *E. coli*

The comparative genomics of the urinary *E. coli* isolates used in this work was reported in Garretto et al. 2020. The current project further characterizes these genomes. To identify genes that may antagonize phage infection, I scanned the contigs of these isolates for sequence homology to 27 genes; as controls, I also scanned the genome of the laboratory strain *E. coli* K-12 and three classical UPEC strains (CFT073, NU14, UTI89) (Figure 1a-c). A homology score of 100 indicates identical sequence identity between the gene query and the scanned genome. A homology score of 0 indicates absence of the query in the scanned genome. A homology score between 100 and 0 for a gene would indicate some conserved regions in the scanned genome (e.g., a homolog or mutant). A large error bar would indicate mixed levels of conservation within a group. Anti-phage genes of interest would be those differentially present in the urinary isolates but not present in *E. coli* K-12, which is phage permissive. Alternatively, such genes could have highly variable scores amongst the urinary isolates, indicating the presence of anti-phage gene homologues within some of those isolates but not others.

RNA polymerase was used as a positive control, as it was expected to be conserved and thus have a homology score near 100 for all genomes. I assessed genes reported to be associated with abortive effects on the phage life cycle (Figure 1a), blocking phage adsorption (Figure 1b), or documented to be receptors used by phage for adsorption (Figure 1c). Amongst the urinary isolates, phage life cycle abortive genes with highly variable scores were *rexA, hok, shok, mazE*, and *mazF*. *rexB* is

a lambda phage gene that aborts lytic growth of bacterial viruses and it presence could denote infection by a lambda-like phage in some urinary isolates. The genes *hok/shok* and *mazEF* are Toxin-Antitoxin (TA) modules associated with plasmid addiction/retention and inhibition of life cycle processes essential to the host cell (Figure 1a). Amongst the urinary isolates, phage adsorption blocking genes with highly variable scores were *imm* and *traT* (Figure 1b). The *imm* gene is a phage T4 gene associated with superinfection immunity and could indicate prophage infection in some isolates. *traT* is a plasmid-borne gene that blocks invading plasmids, but also can block phage from binding to cell surface proteins. Genes for prototypical coliphage receptors had similar homology scores between *E. coli* K-12, the UPEC strains, and the urinary isolates with error bars that indicate no statistical difference, indicating that these genes are conserved in all genomes scanned (Figure 1c).

Of the genes in the urinary isolates with high variance (*rexB, imm, hok/shok, mazEF,* and *traT*), the common denominators were: (1) *rexB* and *imm*, which are associated with prophage, and (2) *hok/shok, mazEF*, and *traT*, which are associated with plasmids. The genes *mazEF* were highly conserved in *E. coli* K-12, which is permissible to phage infection by P1vir, Greed, and Lust (Figure 2), but the large variance for *mazEF* in the UPEC strains and the urinary isolates could indicate that some of these genomes could carry *mazEF* homologues. Different TA modules could result in different phage permissivity phenotypes. To assess if TA genes could be associated with phage permissivity, I first spotted coliphages P1vir, Greed,4 and Lust onto lawns of 67 urinary *E. coli* isolates and 3 laboratory strains (K-12, B, and C). Isolates that had a clear spot when exposed to phage were given a positive result (Table 9). The laboratory strains were permissive for all 3 phages. Most of the urinary isolates were not permissive. The isolates permissive for P1vir differed substantially from the isolates permissive for Greed and Lust, whose susceptibility profiles were quite similar. Since *mazEF* was associated with phage permissive *E. coli* but other TA homologues may not be, I disqualified TAs conserved in phage permissive *E. coli*. I queried the genomes of *E. coli* isolates susceptible to all three phages for TA

modules. The only TA module that was not conserved (i.e., had a homology score under 90%) in all four permissive isolates (*E. coli* K-12, B, C, and UMB2019) was *pemIK* (Table 10).

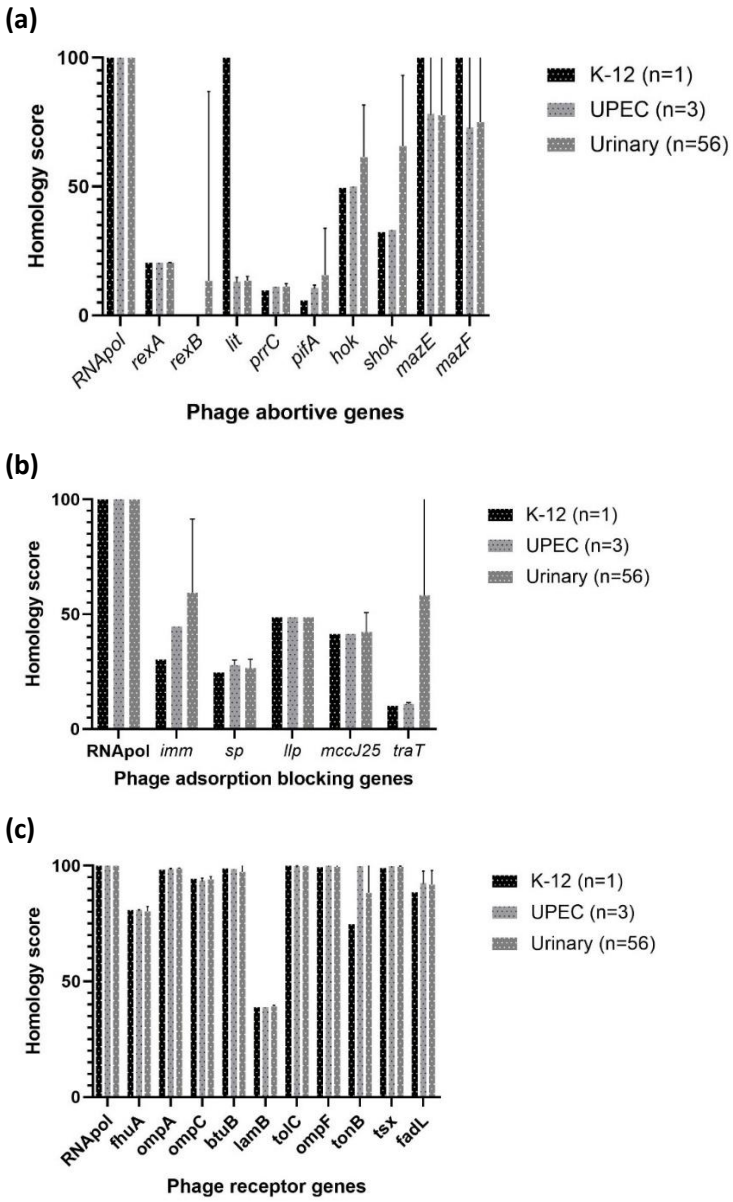


Figure 1. Alignment Score of Anti-phage Genes to Urinary *E. coli* Genomes.
Homology score measures how conserved a query gene is in the given genome (100% is perfectly conserved, 0% is absent, intermediate values denote a mutant or gene homologue compared to the referenced queried). (a) Phage abortive genes abort the phage life cycle. (b) Phage adsorption blocking genes antagonize binding of phage to the cell surface. (c) Classical cell surface proteins used for phage for adsorption and host entry.

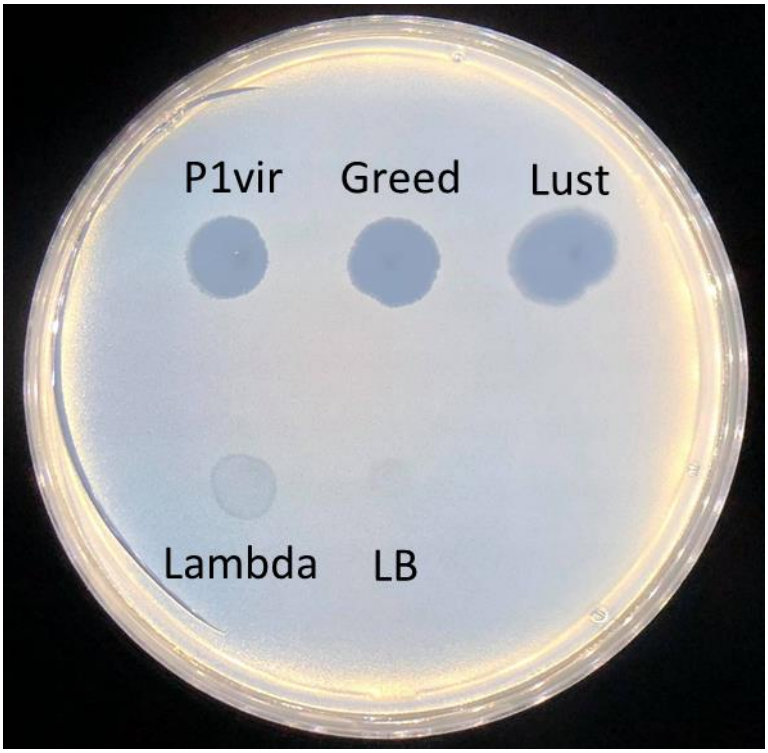


Figure 2. *E. coli* K-12 Lawn Spotted with a Variety of Phages.

Phages P1vir, Greed, and Lust have a clear spot phenotype, correlated with lytic infection. Lambda phage has a turbid spot phenotype, correlated with lysogenic infection. LB spotting (negative control) does not result in a spot, indicative of no visual evidence of infection.

Table 9. Urinary *E. coli* Clear Phage Spot Phenotype.

Strain	Strain type	P1 lysis	Greed lysis	Lust lysis
K-12	Laboratory	+	+	+
B	Laboratory	+	+	+
C	Laboratory	+	+	+
UMB0103	UTI asymptomatic			
UMB0149	UTI asymptomatic			
UMB0276	UTI asymptomatic			
UMB0527	UTI asymptomatic			
UMB0731	UTI asymptomatic		+	+
UMB0906	UTI symptomatic			
UMB0923	UTI symptomatic			
UMB0928	UTI asymptomatic			
UMB0931	UTI symptomatic		+	
UMB0933	UTI asymptomatic			
UMB0934	UTI symptomatic			
UMB0939	UTI asymptomatic			
UMB0949	UTI symptomatic			
UMB1012	UTI symptomatic			
UMB1091	UTI symptomatic			
UMB1093	UTI symptomatic			
UMB1160	UTI symptomatic	+		
UMB1162	UTI symptomatic			
UMB1180	UTI symptomatic			
UMB1193	UTI symptomatic			
UMB1195	UTI symptomatic			
UMB1202	UTI symptomatic			
UMB1220	UTI symptomatic			
UMB1221	UTI symptomatic			
UMB1223	UTI symptomatic			
UMB1225	UTI symptomatic			
UMB1228	UTI symptomatic			
UMB1229	UTI symptomatic			
UMB1284	UTI symptomatic			
UMB1285	UTI symptomatic			
UMB1335	UTI symptomatic			
UMB1337	UTI symptomatic			
UMB1346	UTI symptomatic			
UMB1347	UTI symptomatic			
UMB1348	UTI symptomatic			
UMB1354	UTI symptomatic			

Table 9. Urinary *E. coli* Clear Phage Spot Phenotype (continued).

UMB1356	UTI symptomatic			
UMB1358	UTI symptomatic	+		
UMB1359	UTI symptomatic			
UMB1360	UTI symptomatic			
UMB1362	UTI symptomatic			
UMB1526	UTI symptomatic			
UMB1727	UTI asymptomatic			
UMB2019	UTI asymptomatic	+	+	+
UMB2055	UTI asymptomatic			
UMB2321	UTI asymptomatic			
UMB2328	UTI asymptomatic			
UMB3538	UTI asymptomatic			
UMB3641	UTI asymptomatic			
UMB3643	UTI asymptomatic		+	+
UMB4656	UTI symptomatic	+		
UMB4714	N/A			
UMB4716	UTI asymptomatic			
UMB4746	UTI asymptomatic	+		
UMB5337	UTI asymptomatic			
UMB5814	UTI symptomatic			
UMB5924	UTI asymptomatic	+		
UMB5978	UTI asymptomatic			
UMB6360	UTI asymptomatic			
UMB6454	UTI symptomatic			
UMB6611	UTI asymptomatic	+		
UMB6653	UTI symptomatic			
UMB6655	UTI asymptomatic		+	+
UMB6713	UTI asymptomatic			
UMB6721	UTI symptomatic			
UMB6890	UTI asymptomatic	+		
UMB7431	UTI symptomatic			

Note: UTI symptomatic and asymptomatic denote the urinary condition of the participant from which the isolate was isolated. A clear phage spot phenotype is indicative of lysis.

Table 10. Toxin-Antitoxin Genes in Phage Permissive *E. coli*.

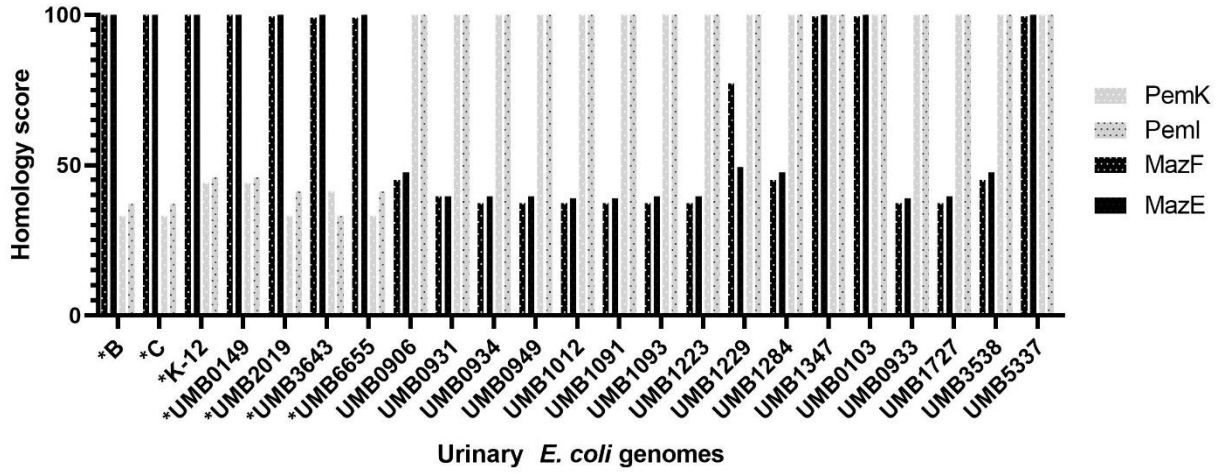
Protein	Operon	Mechanism for host antagonism	B	C	K-12	UMB2019
MazF	*mazEmazF*	Ribosome-independent RNA inteferases	100	100	100	99.5
ChpBK	*chpBchpBK*	Ribosome-independent RNA inteferases	45.3	45.3	100	45.3
HicA	*hicAhicB*	Ribosome-independent RNA inteferases	99.1	99.1	99.1	41.4
YhaV	*prlFyhaV*	Ribosome-independent RNA inteferases	97.7	98.7	100	16.9
MqsR	*mqsRmqsA*	Ribosome-independent RNA inteferases	100	99.5	100	27
RnlA	*rnlArnlB*	Ribosome-independent RNA inteferases	18.9	18.9	100	18.9
RelE	*relBrelE*	Ribosome-dependent RNA interfereses	27.9	100	100	31.6
YoeB	*yefMyoeB*	Ribosome-dependent RNA interfereses	31.5	31.5	99.4	98.8
YafO	*yafNyafO*	Ribosome-dependent RNA interfereses	99.2	67	100	98.9
YafQ	*dinJyafQ*	Ribosome-dependent RNA interfereses	98.9	99.5	100	31
HigB	*higBhigA*	Ribosome-dependent RNA interfereses	100	99.5	100	22.1
RatA	*ratAyfjF*	Inhibitor of ribosome subunit association	100	100	100	99.4
CbtA	*yeeUcbtA*	Inhibitors of cell division	90.3	83	100	100
Ykfl	*yafQykfl*	Inhibitors of cell division	76.9	79	95.1	80.4
YfjF	*yjfZypjF*	Inhibitors of cell division	72.9	77.7	82.5	78.2
GnsA	*gnsAymcE*	Inhibitor of phospholipid synthesis	99.1	98.2	99.1	99.1
HipA	*hipBhipA*	Unknown	99.5	99.5	99.5	98.4
YjhX	*yjhXyjhQ*	Unknown	100	22.4	100	28.2
YdaS	*ydaSydaT*	Unknown	32.7	28.1	100	29.6
PemK	*pemIpemK*	Endoribonuclease	33.1	33.1	44	33.1
MazE	*mazEmazF*	Ribosome-independent RNA inteferases	100	100	100	100
ChpBI	*chpBchpBK*	Ribosome-independent RNA inteferases	48.2	48.2	100	48.2
HicB	*hicAhicB*	Ribosome-independent RNA inteferases	100	100	100	76.1
PrlF	*prlFyhaV*	Ribosome-independent RNA inteferases	98.6	99.1	100	29.3
MqsA	*mqsRmqsA*	Ribosome-independent RNA inteferases	100	100	100	32.1
RnlB	*rnlArnlB*	Ribosome-independent RNA inteferases	27.6	27.6	99.6	20.7

Table 10. Toxin-Antitoxin Genes in Phage Permissive *E. coli* (continued).

RelB	*relBrelE*	Ribosome-dependent RNA interfereses	41.8	100	100	39.9
YefM	*yefMyoeB*	Ribosome-dependent RNA interfereses	39.8	39.8	100	100
YafN	*yafNyafO*	Ribosome-dependent RNA interfereses	99.5	99.5	100	100
DinJ	*dinJyafQ*	Ribosome-dependent RNA interfereses	91.2	91.2	91.9	40.1
HigA	*higBhigA*	Ribosome-dependent RNA interfereses	100	100	100	72.1
YfjF	*ratAyfjF*	Inhibitor of ribosome subunit association	32.2	80	99.1	81
YeeU	*yeeUcbtA*	Inhibitors of cell division	25.8	81.9	94.3	96.3
YafW	*yafQykfI*	Inhibitors of cell division	22.4	82.1	100	80.1
YpjZ	*yjfZypjF*	Inhibitors of cell division	31.4	80.9	100	80.9
YmcE	*gnsAymcE*	Inhibitor of phospholipid synthesis	100	100	100	100
HipB	*hipBhipA*	Unknown	100	100	100	97.2
YjhQ	*yjhXyjhQ*	Unknown	100	33.1	100	33.1
YdaT	*ydaSydaT*	Unknown	16.1	16.1	100	31.8
PemI	*pemIpemK*	Endoribonuclease	37.1	37.1	45.9	41.2

If *pemIK* is linked to reduced permissivity in urinary *E. coli* isolates, we would expect this TA module to be present in less permissive isolates (urinary *E. coli*) and absent in more permissive isolates (K-12 B, C, K and UMB2019). Whereas *pemIK* was absent in all the phage permissive strains/isolates (noted by an asterisk), they all possessed its homolog *mazEF* (Figure 3a). In contrast, *pemIK* was highly conserved in 16 urinary isolates with low permissibility to phage. Because *pemIK* is a plasmid-associated gene, we repeated this analysis for *traT*, the other plasmid-associated gene with high variance in urinary *E. coli*. *traT* was absent in all but one of the permissive strains/isolates but present in 28 of the less permissive isolates (Figure 3b). *pemIK* and *traT* are anti-phage genes with dual function in plasmid maintenance, often associated with *E. coli* F plasmids. Therefore, I hypothesized that plasmids could play a role in reducing phage infection permissivity in urinary *E. coli*.

(a)



(b)

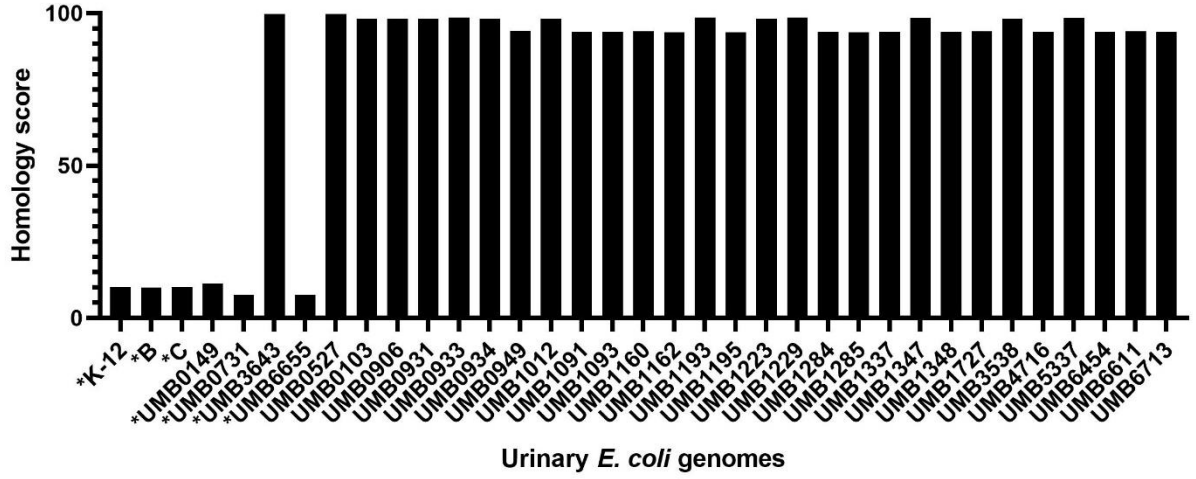


Figure 3. Plasmid-Linked Genes in Phage Permissive vs. Non-Permissive *E. coli*.

E. coli isolates susceptible to P1vir, Greed, or Lust infection are noted by an asterisk. (a) *mazEF* is conserved in all *E. coli* susceptible to phage, while *pemIK* is present only in isolates with no visual evidence of infection. (b) *traT* is absent in all but one of the isolates susceptible to phage infection and present in isolates with no visual evidence of phage infection. *pemIK* and *traT* are plasmid-linked genes with dual anti-phage function.

Types of Plasmids in Urinary *E. coli*

To catalogue the plasmids of the urinary isolates, PlasmidSPAdes was used to build plasmidic assemblies (i.e., contigs) from raw whole genome sequence reads, web BLAST was used to identify contigs with homology to plasmids, and Prokka was used to annotate the curated contigs (Table 11). All but ten genomes generated a plasmidic assembly with a contig average of 10.9. The total nucleotide size ranged from 4,990 bp to 372,241 bp with an average of 156,243 bp (Table 11). These data are generally consistent with the expected *E. coli* plasmid range of 1,000 bp to 200,000 bp. The predicted CDS (i.e., protein coding regions) average was 181.5.

To profile the type of plasmids in each urinary *E. coli* assembly, I scanned the contigs with PlasmidFinder to identify *E. coli* replicon/incompatibility genes, then organize these hits into the major incompatibility groups of Col, IncF, and Inc-various (Table 12). There were six distinct *col* genes with a total of 38 hits in the plasmidic assemblies. There were nine distinct *incF* genes with a total of 96 hits in the plasmidic assemblies. Thirteen distinct *inc* genes grouped into the Inc-various group, with a total of 40 hits in the plasmidic assemblies. In total, there were 27 distinct *inc* gene hits in 57 urinary *E. coli* isolates, which were grouped into the three major groups Col, IncF, and Inc-various (Table 13). For some analyses, the IncF group was divided into the IncFI and IncFII groups (the former having no *incFII* hit).

inc genes were found in approximately 85% of plasmidic assemblies (n=57/67) (Figure 4). The most common types of *inc/rep* gene were those from IncFII (55.22%) and IncFI (13.43%), which are associated with F plasmids. These data indicate that plasmids are common in urinary *E. coli* isolates and that F plasmids predominate. Since the F plasmid-borne genes *pemIK* and *traT* were common in isolates with no visual evidence of phage infection, I analyzed the spot phenotype for association with plasmid type (Table 9). Urinary isolates with the *incFII* gene were statistically more likely to be non-permissive for the urinary coliphages Greed or Lust (Figure 5).

Table 11. Assembly and Annotation Overview of Urinary *E. coli* Plasmidic Assemblies

Strain	Contigs	Bases	ORF
UMB0103	12	231531	251
UMB0149	12	107939	115
UMB0276	0		
UMB0527	20	209793	227
UMB0731	3	120907	149
UMB0906	16	164812	185
UMB0923	3	125070	142
UMB0928	17	183351	202
UMB0931	19	372241	463
UMB0933	13	222613	256
UMB0934	16	166402	202
UMB0939	3	21775	25
UMB0949	12	137266	156
UMB1012	15	156824	184
UMB1091	39	253344	300
UMB1093	23	220386	237
UMB1160	10	140088	172
UMB1162	11	291702	322
UMB1180	1	4990	5
UMB1193	10	151593	186
UMB1195	1	94010	107
UMB1202	2	137792	147
UMB1220	0		
UMB1221	3	139286	151
UMB1223	12	139637	156
UMB1225	0		
UMB1228	3	42562	54
UMB1229	10	152983	189
UMB1284	2	111211	129
UMB1285	4	154985	159
UMB1335	34	200689	233
UMB1337	32	198104	233
UMB1346	2	60482	73
UMB1347	2	60482	73
UMB1348	8	162084	195
UMB1354	0		
UMB1356	0		

Table 11. Assembly and Annotation Overview of Urinary *E. coli* Plasmidic Assemblies (continued).

UMB1358	0		
UMB1359	0		
UMB1360	6	124614	195
UMB1362	12	87221	95
UMB1526	0		
UMB1727	13	239487	296
UMB2019	13	299962	353
UMB2055	3	127588	150
UMB2321	4	100530	115
UMB2328	2	96141	108
UMB3538	15	169304	185
UMB3641	12	93840	98
UMB3643	1	92125	102
UMB4656	28	151223	172
UMB4714	9	239877	287
UMB4716	9	239813	287
UMB4746	0		
UMB5337	1	60192	72
UMB5814	0		
UMB5924	15	112240	123
UMB5978	11	101237	120
UMB6360	9	70556	86
UMB6454	4	112999	131
UMB6611	5	123060	135
UMB6653	34	261301	302
UMB6655	5	182135	222
UMB6713	13	211037	232
UMB6721	17	250312	293
UMB6890	4	181439	223
UMB7431	9	240707	284
Plasmidic contig mean	10.9474	156243	181.474

To assess plasmid sequence identity in the urinary *E. coli* plasmidic assemblies beyond *inc* gene profile, the curated plasmid assemblies were scanned via web BLAST. All urinary *E. coli* plasmidic assemblies had a top BLAST hit to a plasmid in the NCBI database (Table 14). The average size of the plasmid hits in the NCBI database was 105,322 bp. The cover query range was 71-100% and the

sequence identity range was approximately 96-100%. Generally, urinary *E. coli* plasmids had top hits to *E. coli* plasmids in the database, but some isolates had hits to plasmids from *Enterobacter, Klebsiella*, and *Salmonella* species.

Table 12. Inc Genes Identified in Urinary *E. coli* Plasmidic Assemblies.

Plasmid type	Inc gene	Hits in urinary *E. coli* (n=67)
Col	Col(BS512)	7
	Col156	26
	Col(MG828)	1
	ColRNAI	1
	ColpVC	1
	Col440I	2
Total	6	38
IncF	IncFIA	12
	IncFIB(AP001918)	37
	IncFII(pRSB107)	8
	IncFII	10
	IncFIB(pB171)	1
	IncFII(pCoo)	2
	IncFIC(FII)	6
	IncFII(29)	18
	IncFIB(H89-PhagePlasmid)	2
Total	9	96
Inc-various	p0111	5
	IncY	4
	IncI1-I(Gamma)	7
	IncB/O/K/Z	7
	IncQ1	2
	IncP1	1
	IncX4	4
	IncX1	2
	IncN	3
	IncN3	2
	IncB/O/K/Z	1
	IncI2(Delta)	2
Total	12	40

Table 13. Inc Genes in Urinary *E. coli* Plasmidic Assemblies.

Strain	UMB0103	UMB0149	UMB0276	UMB0527	UMB0731	UMB0906	UMB0923	UMB0928	UMB0931	UMB0933
	5	1	0	2	2	4	1	6	5	3
Col(BS512)	+					+		+		
IncFIA	+					+		+	+	
IncFIB(AP001918)	+					+		+	+	+
IncFII(pRSB107)	+					+			+	
p0111	+						+			
IncY		+								
IncFII				+						+
IncI1-I(Gamma)				+				+		
Col156					+			+		
IncB/O/K/Z					+					+
IncQ1								+		
IncP1									+	
IncX4									+	
Col(MG828)										
ColRNAI										
IncFIB(pB171)										
IncFII(pCoo)										
ColpVC										
IncFIC(FII)										
IncFII(29)										
Col440I										
IncX1										
IncX4										
IncN										
IncN3										
IncB/O/K/Z										
IncFIB(H89-PhagePlasmid)										
IncI2(Delta)										

Table 13. Inc *G*enes in *U*rinary *E. coli P*lasmidic *A*ssemblies (continued).

	UMB0934	UMB0939	UMB0949	UMB1012	UMB1091	UMB1093	UMB1160	UMB1162	UMB1180	UMB1193
	3	2	4	5	6	3	3	3	1	3
Col(BS512)										
IncFIA	+		+	+						
IncFIB(AP001918)	+		+		+	+	+	+		+
IncFII(pRSB107)	+			+						
p0111								+		
IncY										
IncFII			+				+			+
IncI1-I(Gamma)										
Col156			+	+	+	+	+			+
IncB/O/K/Z										
IncQ1										
IncP1										
IncX4										
Col(MG828)		+								
ColRNAI		+								
IncFIB(pB171)				+						
IncFII(pCoo)				+	+					
ColpVC					+					
IncFIC(FII)					+			+		
IncFII(29)					+	+				
Col440I									+	
IncX1										
IncX4										
IncN										
IncN3										
IncB/O/K/Z										
IncFIB(H89-PhagePlasmid)										
IncI2(Delta)										

Table 13. Inc *G*enes in *U*rinary *E. coli P*lasmidic *A*ssemblies (continued).

	UMB1195	UMB1202	UMB1220	UMB1221	UMB1223	UMB1225	UMB1228	UMB1229	UMB1284	UMB1285
	3	2	0	4	3	0	1	3	3	2
Col(BS512)										
IncFIA					+				+	
IncFIB(AP001918)	+	+		+	+			+		+
IncFII(pRSB107)					+					
p0111				+						
IncY										
IncFII								+	+	
IncI1-I(Gamma)				+						
Col156								+		
IncB/O/K/Z										
IncQ1										
IncP1										
IncX4									+	
Col(MG828)										
ColRNAI										
IncFIB(pB171)										
IncFII(pCoo)										
ColpVC										
IncFIC(FII)	+	+		+						+
IncFII(29)										
Col440I										
IncX1							+			
IncX4										
IncN										
IncN3										
IncB/O/K/Z	+									
IncFIB(H89-PhagePlasmid)										
IncI2(Delta)										

Table 13. Inc *G*enes in *U*rinary *E. coli P*lasmidic *A*ssemblies (continued).

	UMB1335	UMB1337	UMB1346	UMB1347	UMB1348	UMB1354	UMB1356	UMB1358	UMB1359	UMB1360
	4	4	1	1	4	0	0	0	0	3
Col(BS512)										
IncFIA										
IncFIB(AP001918)	+	+			+					+
IncFII(pRSB107)										
p0111										
IncY										
IncFII										
IncI1-I(Gamma)										
Col156	+	+			+					+
IncB/O/K/Z										
IncQ1										
IncP1										
IncX4										
Col(MG828)										
ColRNAI										
IncFIB(pB171)										
IncFII(pCoo)										
ColpVC										
IncFIC(FII)										
IncFII(29)	+	+	+	+	+					+
Col440I										
IncX1					+					
IncX4										
IncN	+	+								
IncN3										
IncB/O/K/Z										
IncFIB(H89-PhagePlasmid)										
IncI2(Delta)										

Table 13. Inc *G*enes in *U*rinary *E. coli P*lasmidic *A*ssemblies (continued).

	UMB1362	UMB1526	UMB1727	UMB2019	UMB2055	UMB2321	UMB2328	UMB3538	UMB3641	UMB3643
	2	0	4	4	2	1	1	5	3	1
Col(BS512)	+							+		
IncFIA								+	+	
IncFIB(AP001918)	+		+					+	+	
IncFII(pRSB107)								+	+	
p0111										+
IncY				+						
IncFII			+							
IncI1-I(Gamma)				+	+	+	+			
Col156										
IncB/O/K/Z			+							
IncQ1										
IncP1										
IncX4			+							
Col(MG828)										
ColRNAI										
IncFIB(pB171)										
IncFII(pCoo)										
ColpVC										
IncFIC(FII)										
IncFII(29)										
Col440I				+						
IncX1										
IncX4										
IncN								+		
IncN3				+	+					
IncB/O/K/Z										
IncFIB(H89-PhagePlasmid)										
IncI2(Delta)										

Table 13. Inc *G*enes in *U*rinary *E. coli P*lasmidic *A*ssemblies (continued).

	UMB4656	UMB4714	UMB4716	UMB4746	UMB5337	UMB5814	UMB5924	UMB5978	UMB6360	UMB6454
	4	4	4	0	1	0	2	4	2	3
Col(BS512)	0							+	+	
IncFIA							+			
IncFIB(AP001918)	+	+	+					+	+	+
IncFII(pRSB107)										
p0111										
IncY		+	+							
IncFII	+									
IncI1-I(Gamma)										
Col156	+	+	+				+	+		+
IncB/O/K/Z										
IncQ1								+		
IncP1										
IncX4										
Col(MG828)										
ColRNAI										
IncFIB(pB171)										
IncFII(pCoo)										
ColpVC										
IncFIC(FII)										
IncFII(29)	+	+	+		+					+
Col440I										
IncX1										
IncX4										
IncN										
IncN3										
IncB/O/K/Z										
IncFIB(H89-PhagePlasmid)										
IncI2(Delta)										

Table 13. Inc *G*enes in *U*rinary *E. coli P*lasmidic *A*ssemblies (continued).

	UMB6471	UMB6611	UMB6653	UMB6655	UMB6713	UMB6721	UMB6890	UMB7431
	3	3	5	3	4	5	3	3
Col(BS512)								
IncFIA								
IncFIB(AP001918)	+	+	+		+	+		
IncFII(pRSB107)								
p0111								
IncY								
IncFII		+						
IncI1-I(Gamma)								
Col156	+	+	+	+	+	+	+	
IncB/O/K/Z			+		+	+		+
IncQ1								
IncP1								
IncX4			+					
Col(MG828)								
ColRNAI								
IncFIB(pB171)								
IncFII(pCoo)								
ColpVC								
IncFIC(FII)								
IncFII(29)	+		+		+	+		+
Col440I								
IncX1								
IncX4						+		+
IncN								
IncN3								
IncB/O/K/Z								
IncFIB(H89-PhagePlasmid)				+			+	
IncI2(Delta)				+			+	

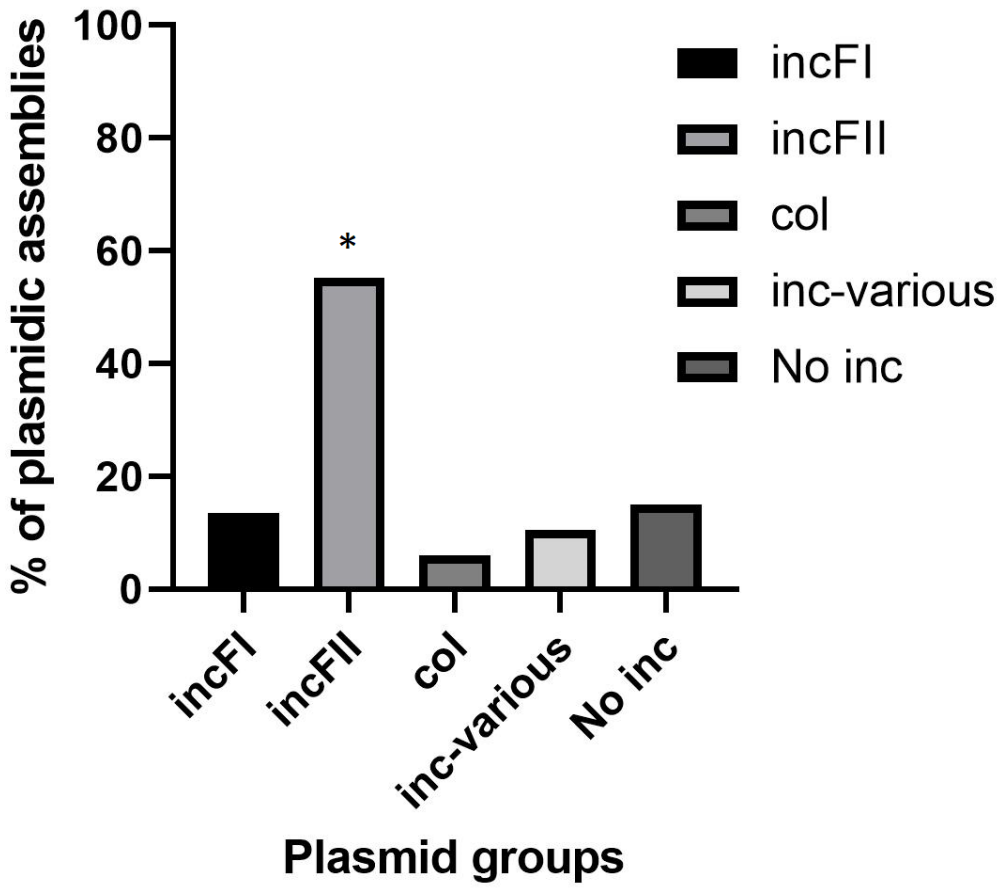


Figure 4. Plasmid Inc Group in Urinary *E. coli* Plasmidome.

Proportion of urinary *E. coli* isolates with *inc* gene hits of a particular group in their plasmidic assembly. The chi-square statistic is 66.903. The p-value is < 0.00001. The result is significant at p < .05. This difference in IncF representation would be greater if IncFII and IncFI were combined into a single F plasmid group.

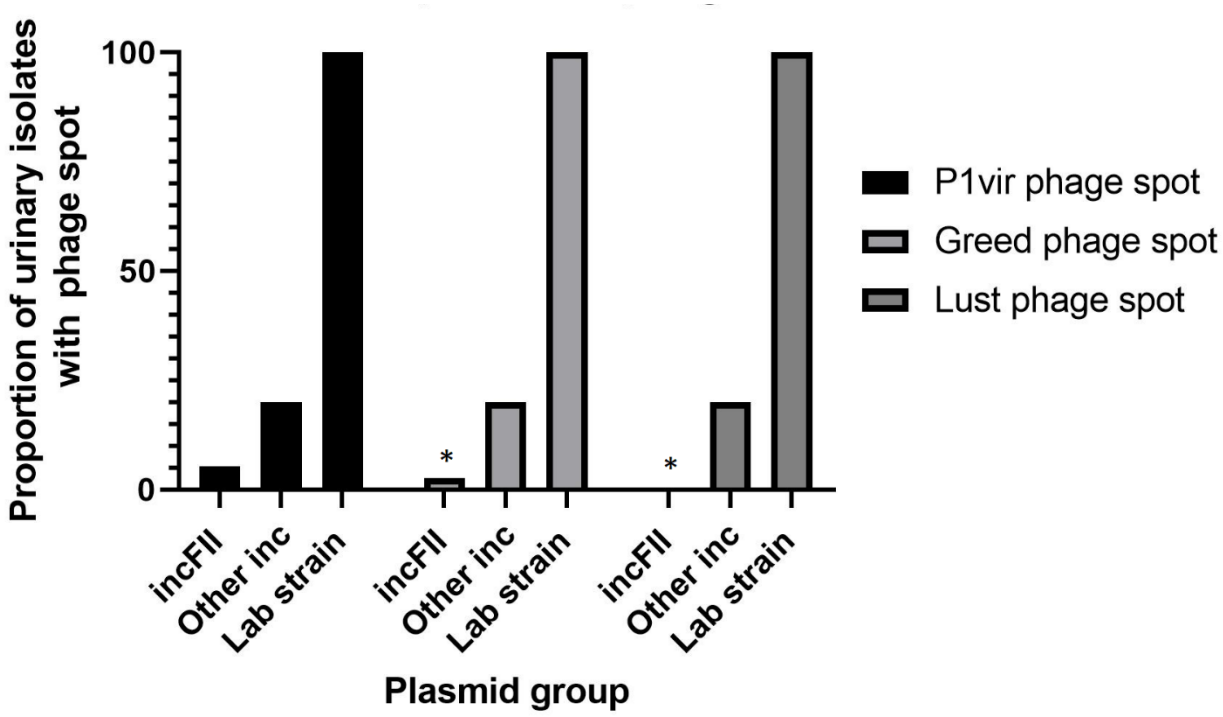


Figure 5. Inc Groups Association with Phage Spot Phenotype.

Urinary *E. coli* with *incFII* genes characteristic of F plasmids are less likely to have evidence of infection when spotted with phage. All laboratory strain of laboratory *E. coli* tested had a spot when exposed to the phages. Bladder *E. coli* isolates with IncFII were statistically more likely to not have spots when exposed to Greed (The Fisher exact test statistic value is 0.0465. The result is significant at $p < .05$) and Lust (The Fisher exact test statistic value is 0.0123. The result is significant at $p < .05$) but not P1vir (The Fisher exact test statistic value is 0.1699. The result is *not* significant at $p < .05$) compared to other Inc groups (IncF, Col, Inc-various).

Table 14. Web BLAST Hit for Urinary *E. coli* Plasmidic Assemblies.

Strain	Inc group	Plasmid	Organism	Bases	Qcov	Eval	Pident	Accession
UMB0103	IncFII	pSCU-120-2	*Escherichia coli*	98630	53.00%	0	99.94%	CP054337.1
UMB0149	IncFII	pZR78	*Escherichia coli*	91281	55.00%	0	98.72%	MF455226.1
UMB0527	IncFII	pMSB1_1D-sc-2280324	*Escherichia coli*	96021	56.00%	0	99.87%	LR898869.1
UMB0731	Col	pECOED	*Escherichia coli*	119594	91.00%	0	98.71%	CU928147.1
UMB0906	IncFII	pMRY09-581ECO_1	*Escherichia coli*	159781	81.00%	0	99.97%	AP018456.1
UMB0923	Inc	pSCU-120-2	*Escherichia coli*	98630	71.00%	0	97.75%	CP054337.1
UMB0928	IncFI	pSH146_87	*Salmonella enterica*	86586	49.00%	0	99.74%	JX445149.1
UMB0931	IncFII	p168-7	*Escherichia coli*	93108	99.00%	0	99.99%	CP041570.1
UMB0933	IncFII	pHUSEC411	*Escherichia coli*	98864	88.00%	0	99.98%	HG428756.1
UMB0934	IncFII	pNMBU-W13E19_01	*Escherichia coli*	122112	99.00%	0	99.99%	CP043407.1
UMB0939	Col	pCRE-085-3	*Klebsiella pneumoniae*	9294	45.00%	0	100.00%	CP061402.1
UMB0949	IncFII	pTHO-003-1	*Escherichia coli*	123467	100.00%	0	99.99%	AP022526.1
UMB1012	IncFII	pRHBSTW-00372_3	*Escherichia coli*	78143	57.00%	0	99.77%	CP056571.1
UMB1091	IncFII	p90-9133_1	*Escherichia coli*	113164	67.00%	0	99.82%	CP042948.1
UMB1093	IncFI	pEC11	*Escherichia coli*	31467	91.00%	0	98.89%	CP027258.1
UMB1160	IncFII	pECO-bc6	*Escherichia coli*	101201	100.00%	0	100.00%	CP014668.1
UMB1162	IncFII	p50579417_1	*Escherichia coli*	96948	86.00%	0	97.22%	CP033882.1
UMB1180	Col	pEcl5-3	*Enterobacter hormaechei*	4863	100.00%	0	99.98%	CP047739.1
UMB1193	IncFII	pXJ-K1	*Klebsiella pneumoniae*	130628	36.00%	0	99.99%	CP032165.1
UMB1195	IncFII	p86	*Escherichia coli*	86147	81.00%	0	99.18%	CP023387.1
UMB1202	IncFII	pRHBSTW-00152_2	*Escherichia coli*	133729	85.00%	0	99.94%	CP056811.1
UMB1221	IncFII	pMSB1_8B-sc-2280300	*Escherichia coli*	97184	93.00%	0	98.37%	LR890538.1
UMB1223	IncFII	pNMBU-W13E19_01	*Escherichia coli*	122112	100.00%	0	99.97%	CP043407.1
UMB1228	Inc	pN18EC0432-4	*Escherichia coli*	47739	75.00%	0	98.05%	CP048291.1
UMB1229	IncFII	pDA61218-116	*Escherichia coli*	116466	100.00%	0	100.00%	CP061207.1
UMB1284	IncFII	p179-1	*Escherichia coli*	122483	58.00%	0	100.00%	CP041560.1
UMB1285	IncFII	pSCU-115-1	*Escherichia coli*	148443	100.00%	0	99.97%	CP054369.1
UMB1335	IncFII	pUTI89	*Escherichia coli* UTI89	114230	100.00%	0	99.99%	CP000244.1
UMB1337	IncFII	pECO-bc6	*Escherichia coli*	101201	49.00%	0	99.99%	CP014668.1
UMB1346	IncFII	p13KWH46-1	*Escherichia coli*	162357	71.00%	0	97.06%	CP019251.1
UMB1347	IncFII	p13KWH46-1	*Escherichia coli*	162357	73%	0	97.06%	CP019251.1

Table 14. Web BLAST Hit for Urinary *E. coli* Plasmidic Assemblies (continued).

Strain	Inc group	Plasmid	Organism	Bases	Qcov	Eval	Pident	Accession
UMB1348	IncFII	pUTI89	*Escherichia coli* UTI89	114230	100.00%	0	99.99%	CP000244.1
UMB1360	IncFII	pEcl4-2	*Enterobacter hormaechei*	114230	100.00%	0	100.00%	CP047742.1
UMB1362	IncFI	p2013C-4465	*Escherichia coli*	66029	69%	0	99.45%	CP015242.1
UMB1727	IncFII	pE16KP0290-2	*Klebsiella pneumoniae*	96685	91%	0	98.93%	CP052260.1
UMB2019	Col	pSRC27-I	*Salmonella enterica*	90483	91%	0	99.72%	CP058811.1
UMB2055	Inc	pSRC27-I	*Salmonella enterica*	90483	91%	0	99.72%	CP058811.1
UMB2321	Inc	p9134dAT	*Salmonella enterica*	109512	97%	0	99.97%	KF705207.1
UMB2328	Inc	p9134dAT	*Salmonella enterica*	109512	97%	0	99.97%	KF705207.1
UMB3538	IncFII	pECAZ161_3	*Escherichia coli*	130905	100%	0	99.96%	CP019009.1
UMB3641	IncFII	pMH13-051M_3 DNA	*Escherichia coli*	74653	94%	0	98.55%	AP018574.2
UMB3643	Inc	pSA20094620.3	*Salmonella enterica*	93719	83%	0	96.49%	CP030188.1
UMB4656	IncFII	pECO-bc6	*Escherichia coli*	101201	100%	0	99.99%	CP014668.1
UMB4714	IncFII	pSF-166-1	*Escherichia coli*	114221	100%	0	99.98%	CP012634.1
UMB4716	IncFII	pSF-166-1	*Escherichia coli*	114221	100%	0	99.98%	CP012634.1
UMB5337	IncFII	p13KWH46-1	*Escherichia coli*	162357	71%	0	97.06%	CP019251.1
UMB5924	IncFI	pSCU-147-1	*Escherichia coli*	104744	100%	0	99.92%	CP054326.1
UMB5978	IncFI	p1-09-02E DNA	*Escherichia coli*	102029	100%	0	99.97%	AP022651.1
UMB6360	IncFI	p14EC007b	*Escherichia coli*	190293	72%	0	99.48%	CP024133.1
UMB6454	IncFII	pSF-166-1	*Escherichia coli*	114221	100%	0	99.95%	CP012634.1
UMB6611	IncFI	pDA61218-116	*Escherichia coli*	116466	89%	0	99.82%	CP061207.1
UMB6653	IncFII	pSCU-109-1	*Escherichia coli*	110400	100%	0	99.97%	CP051734.1
UMB6655	IncFI	pPSUO78_2	*Escherichia coli*	109613	96%	0	98.59%	CP012114.1
UMB6713	IncFII	pHUSEC41-1	*Escherichia coli*	91942	96%	0	99.58%	HE603110.1
UMB6721	IncFII	pSCU-109-1	*Escherichia coli*	110400	100%	0	99.94%	CP051734.1
UMB6890	IncFI	pPSUO78_2	*Escherichia coli*	109613	96%	0	98.59%	CP012114.1
UMB7431	IncFII	pRCS90_pI	*Escherichia coli*	92968	99%	0	99.59%	LT985300.1
			Average plasmid size (bp)	105322				

Genes in Urinary *E. coli* Plasmids

If plasmids are present in urinary *E. coli*, it follows that the plasmid gene content should be profiled to understand its potential contributions to the host. The urinary *E. coli* plasmidic assemblies were scanned with the annotation tool Prokka to generate predicted coding regions with an associated function, if available. The ORFs were sorted via Usearch with a sequence identity threshold of 0.80 and genes organized based on having an assigned function or being hypothetical. The incF plasmid type had the largest number of predicted ORFs (2060), with 26.75% annotated with a function (Table 15). For the col plasmid group, only 18.8% of ORFs were estimated to have an assigned function. For the inc-various group, only 16.96% of ORFs were estimated to have an assigned function. For all distinct ORFs in the urinary *E. coli* plasmidic assemblies, only 24.09% were estimated to have an assigned function.

Table 15. Proportion of ORFs in Annotated Plasmid Type.

	IncF	Col	Inc-various	Total ORFs	% total annotated with known function
Function assigned	551	94	67	712	24.09
Hypothetical	1509	406	328	2243	75.91
Total	2060	500	395	2955	
% annotated with known function	26.75	18.8	16.96		

Given that the TA *pemIK* and transfer protein *traT* are genes of interest in F plasmids, the annotated ORFs urinary *E. coli* F plasmids were reviewed, primarily for plasmid conjugation genes, plasmid retention systems, and toxin secretion system. I identified 15 plasmid transfer genes in the urinary *E. coli* F plasmids (Figure 6). The transfer genes *traC, traD, traI, traJ, traM, traN, traQ, traR, traV,* and *traY* were in more than half of F plasmid assemblies, whereas *traB, traG, traL,* and *traS* were only found in a minority of F plasmid assemblies. The gene *fim*, which blocks the fertility/transfer of invading plasmids, was in more than 80% of F plasmid assemblies. I identified 16 plasmid retention (TA) genes,

specifically 9 toxin and 7 antitoxin genes (Figure 7). Complete TA pairs were identified for *ccdAB, isoAB, mazEF, parDE,* and *pemIK*. The most frequent TA modules were *ccdAB* and *pemIK*. Genes for a Type II secretion system were identified in two urinary *E. coli* F plasmids and genes for a Type IV secretion system were identified in 16 urinary *E. coli* F plasmids (Table 16).

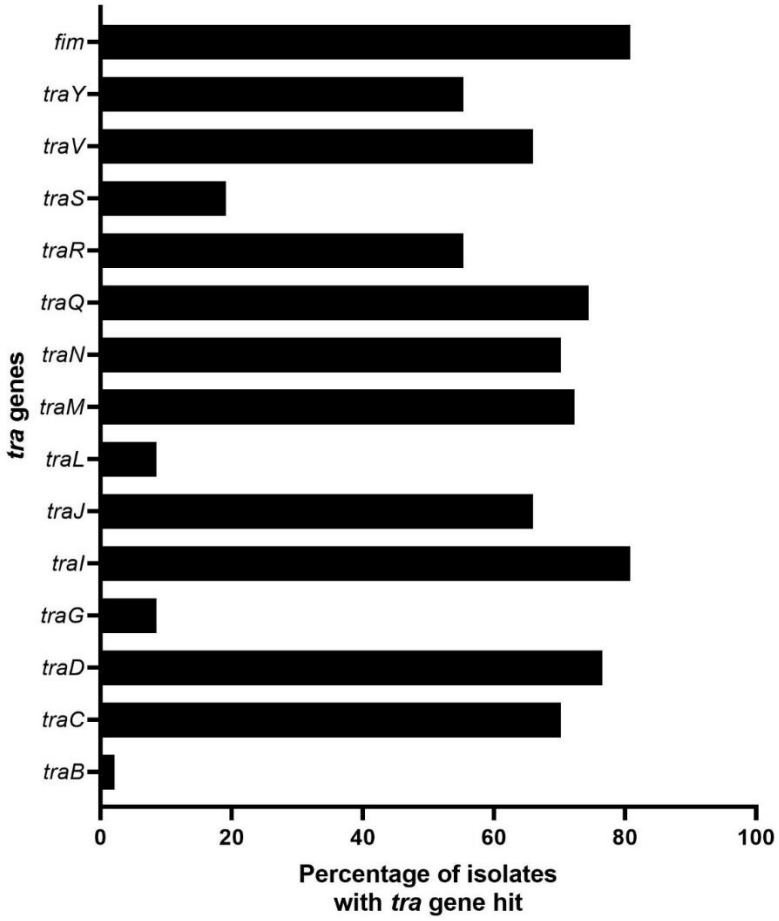


Figure 6. Plasmid Transfer (tra) Genes in Urinary *E. coli* F Plasmids.

Urinary *E. coli* F plasmid assemblies have a variety of transfer genes which are associated with conjugation.

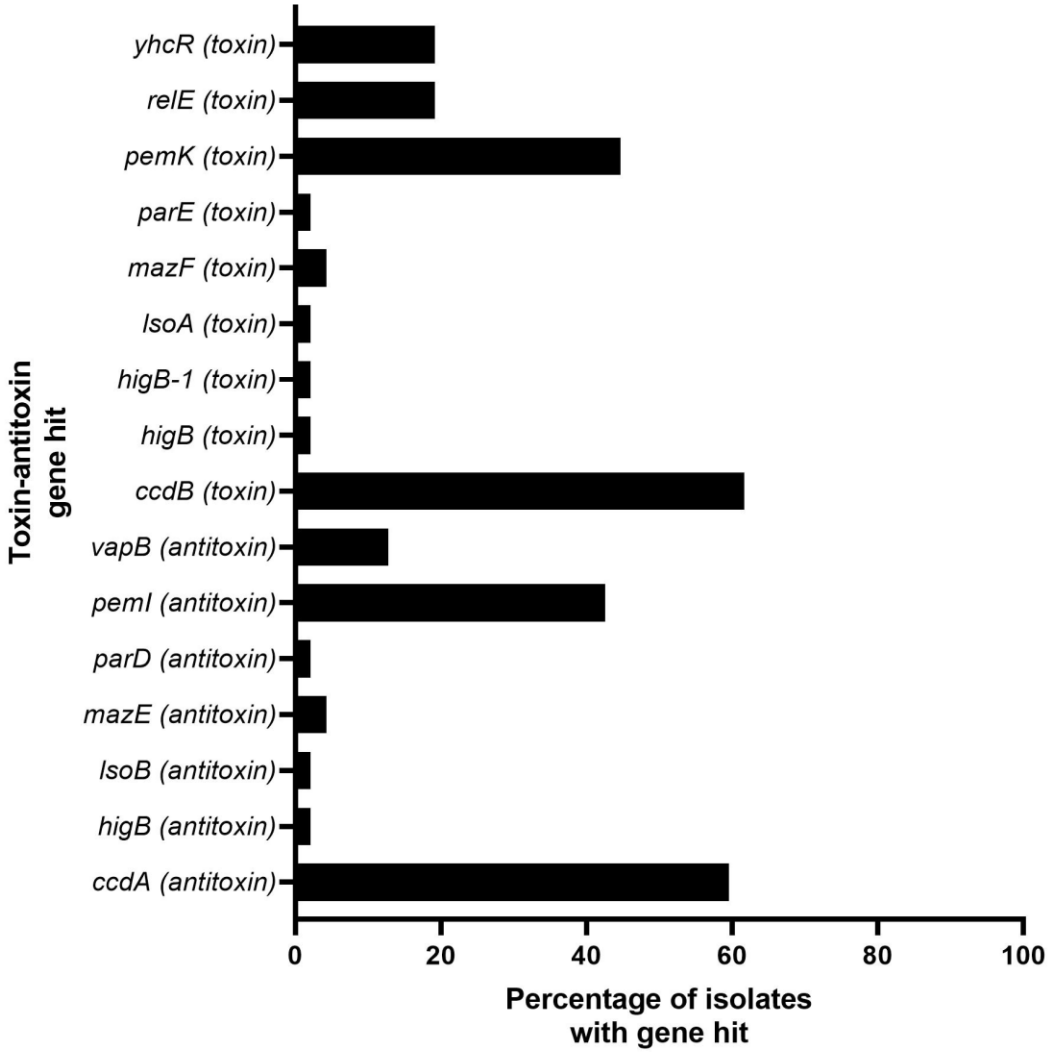


Figure 7. Plasmid Retention (Toxin-Antitoxin) Systems in Urinary *E. coli* F Plasmids.

Urinary *E. coli* F plasmid assemblies have a variety of TA genes which are associated with plasmid retention.

Table 16. Type II and IV Secretion Systems in Urinary *E. coli* Plasmids.

Isolate	Plasmid type	Type II secretion system protein C	Type II secretion system protein E	Type II secretion system protein F	Type II secretion system protein G	Type II secretion system protein J	Type II secretion system protein L	Type II secretion system protein M	Type IV secretion system protein PtlE	Type IV secretion system protein PtlH	Type IV secretion system protein virB1	Type IV secretion system protein VirB11	Type IV secretion system protein virB4	Type IV secretion system protein virB5	Type IV secretion system protein VirB6	Type IV secretion system protein virB8	Type IV secretion system protein virB9	Total per isolate
UMB0906	IncFII												+	+		+	+	4
UMB0931	IncFII								+	+		+	+	+		+	+	7
UMB1193	IncFII											+	+	+	+	+	+	6
UMB1229	IncFII											+	+	+	+	+	+	6
UMB1284	IncFII											+	+			+	+	4
UMB1335	IncFII											+	+	+		+	+	5
UMB1337	IncFII											+	+	+		+	+	5
UMB1348	IncFII								+			+	+				+	4
UMB1362	IncFI	+	+	+	+	+	+	+										7
UMB4714	IncFII											+						1
UMB4716	IncFII											+						1
UMB6360	IncFI	+	+	+	+	+	+	+										7
UMB6653	IncFII											+	+			+	+	4
UMB6655	IncFI										+	+	+			+		4
UMB6721	IncFII											+	+			+	+	4
UMB6890	IncFI										+	+	+			+		4
UMB7431	IncFII											+	+			+	+	4
UMB3538	IncFII											+	+	+		+	+	5

All urinary *E. coli* plasmidic assemblies were scanned for antibiotic resistance genes via ResFinder. There were hits for genes predicted to confer resistance to the following antibiotic classes: aminoglycoside, fluoroquinolone, macrolide, streptomycin, sulfonamide, tetracycline, and trimethoprim (Table 17). Overall, penicillin resistance was the most common antibiotic resistance (32.84%), followed by sulfonamide (26.87%), streptomycin (23.88%), trimethoprim (20.90%), tetracycline (20.90%), and macrolide (16.42%). Predicted resistances to aminoglycoside (8.96%) and fluoroquinolone (1.49%) were the rarest. Some urinary *E. coli* plasmids were predicted to have no hits for antibiotic resistance genes; in contrast, four plasmids were predicted to have seven hits. All four of these plasmids were of the F plasmid group. Antibiotic resistance gene hits were analyzed in the context of plasmid groups IncF, Col, and Inc-various. Hits for all seven antibiotic resistances were present in IncF plasmids, only hits for penicillin resistance were in Col plasmids, and no hits were in Inc-various plasmids (Figure 8a). On average, incF plasmids had 2.15 hits, while col plasmids had an average of 0.25 hits and Inc-various had zero (Figure 8b). The average number of hits in these plasmid groups was not statistically significant given the large variance in the number of hits from IncF plasmids. Antibiotic resistance hits were analyzed in terms of multiple distinct antibiotic resistance hits in each plasmid group (Figure 8c).

Table 17. Types of Antibiotic Resistance Genes in Plasmid Assemblies.

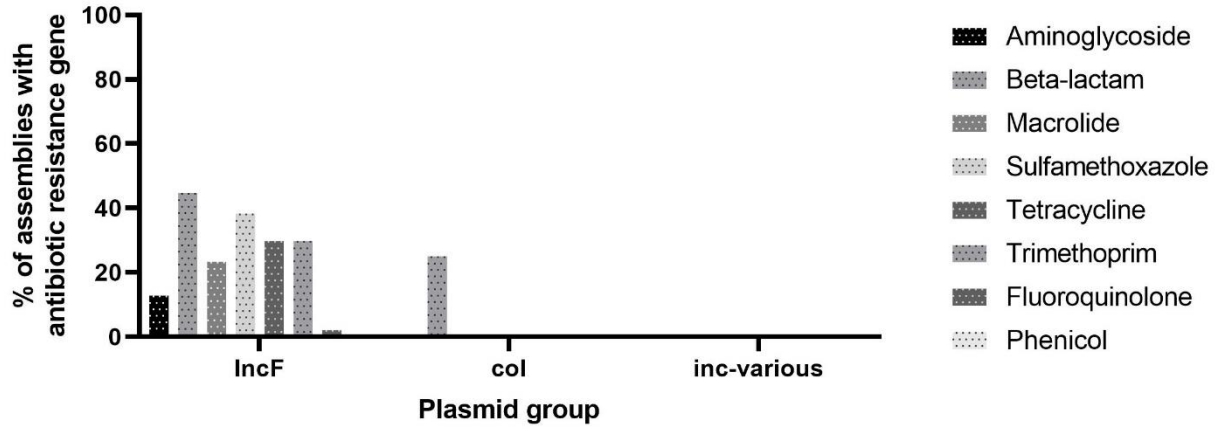
Strain	Plasmid type	Am	Fl	Ma	Pe	St	Su	Tc	Tr	
	% per antibiotic	8.96	1.49	16.42	32.84	23.88	26.87	20.90	20.90	Total per plasmid
UMB0103	IncFII	+			+			+		3
UMB0149	IncFII									0
UMB0276	None									0
UMB0527	IncFII									0
UMB0731	Col									0
UMB0906	IncFII	+		+	+	+	+	+	+	7
UMB0923	Inc									0
UMB0928	IncFI				+	+	+	+	+	5
UMB0931	IncFII			+	+	+	+		+	5
UMB0933	IncFII			+		+	+		+	4
UMB0934	IncFII			+	+	+	+	+	+	6
UMB0939	Col				+					1
UMB0949	IncFII	+		+	+	+	+	+	+	7
UMB1012	IncFII				+					1
UMB1091	IncFII	+		+			+	+	+	5
UMB1093	IncFI				+					1
UMB1160	IncFII			+		+	+		+	4
UMB1162	IncFII									0
UMB1180	Col									0
UMB1193	IncFII				+	+	+	+		4
UMB1195	IncFII									0
UMB1202	IncFII									0
UMB1220	None									0
UMB1221	IncFII									0
UMB1223	IncFII			+	+	+	+	+	+	6
UMB1225	None									0
UMB1228	Inc									0
UMB1229	IncFII				+	+	+	+		4
UMB1284	IncFII							+		1
UMB1285	IncFII									0
UMB1335	IncFII				+					1
UMB1337	IncFII				+					1
UMB1346	IncFII									0

Table 18. Types of Antibiotic Resistance Genes in Plasmid Assemblies (continued).

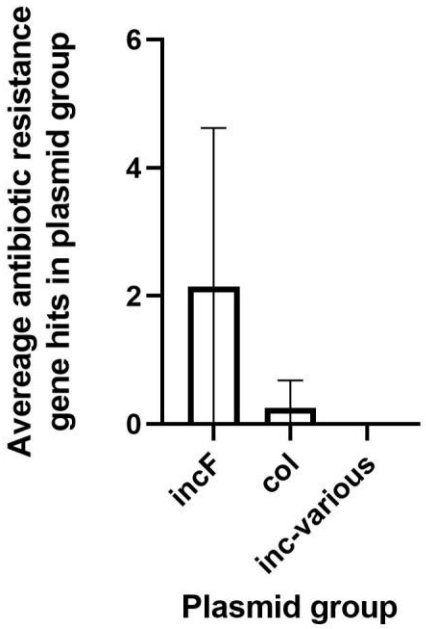
Strain	Plasmid type	Am	Fl	Ma	Pe	St	Su	Tc	Tr	Total per plasmid
UMB1347	IncFII									0
UMB1348	IncFII				+		+			2
UMB1354	None									0
UMB1356	None									0
UMB1358	None									0
UMB1359	None									0
UMB1360	IncFII				+					1
UMB1362	IncFI									0
UMB1526	None									0
UMB1727	IncFII				+					1
UMB2019	Col									0
UMB2055	Inc									0
UMB2321	Inc									0
UMB2328	Inc									0
UMB3538	IncFII	+	+	+	+	+	+		+	7
UMB3641	IncFII									0
UMB3643	Inc									0
UMB4656	IncFII				+					1
UMB4714	IncFII									0
UMB4716	IncFII									0
UMB4746	None									0
UMB5337	IncFII									0
UMB5814	None									0
UMB5924	IncFI	+		+	+	+	+	+	+	7
UMB5978	IncFI			+	+	+	+	+	+	6
UMB6360	IncFI									0
UMB6454	IncFII									0
UMB6611	IncFI									0
UMB6653	IncFII					+	+	+	+	4
UMB6655	IncFI									0
UMB6713	IncFII				+	+	+			3
UMB6721	IncFII					+	+	+	+	4
UMB6890	IncFI									0
UMB7431	IncFII									0

Note: Am=Aminoglycoside, Fl=Fluoroquinolone, Ma=Macrolide, Pe=Penicillin, St=Streptomycin, Su=Sulfonamide, Tc=Tetracycline, Tr=Trimethoprim

(a)


% of assemblies with antibiotic resistance gene
100
80
60
40
20
0
IncF
col
inc-various
Plasmid group
Aminoglycoside
Beta-lactam
Macrolide
Sulfamethoxazole
Tetracycline
Trimethoprim
Fluoroquinolone
Phenicol

(b)


Avereage antibiotic resistance gene hits in plasmid group
6
4
2
0
incF
col
inc-various
Plasmid group

(c)

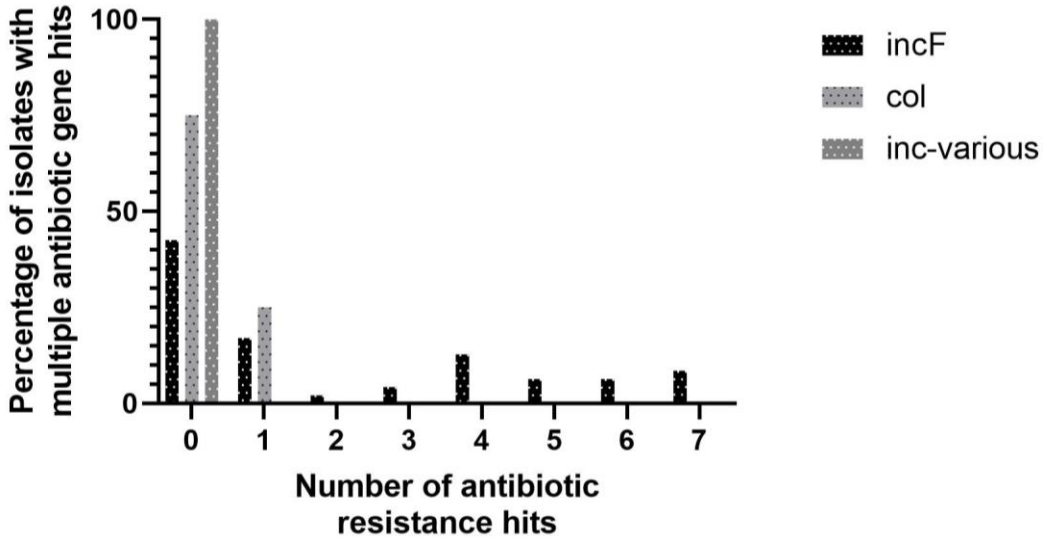


Figure 8. Predicted Antibiotic Resistance Genes in Urinary *E. coli* Plasmids.

(a) Types of antibiotic resistances predicted in plasmids types in urinary *E. coli*. (b) Average number of antibiotic resistance genes in plasmid groups. (c) Percentage of antibiotic resistance hits in isolates from plasmid groups, predictive of multiple antibiotic resistances.

Urinary *E. coli* plasmid assemblies were scanned for virulence genes via VirulenceFinder. There were 30 distinct virulence genes predicted in the plasmid assemblies (Figure 9). Taking into account all the plasmid assemblies, the most common virulence genes were *traT* and *senB*, present in 66.67% and 45.61% of plasmid assemblies, respectively. This also was true for F plasmid assemblies (n=47): *traT* (78.72%) and *senB* (53.19%). One Col plasmid had a hit for *traT* and *senB*, but all the Col plasmids (n=4) had hits to colicin-related virulence genes (*ccl, celb, cib, cia*). The Inc-various plasmids (n=6) had hits to the colicin-related virulence genes *cia* and *cib*. On average, IncF plasmids had 2.4 hits compared to Col plasmids with an average of 1.25 hits and Inc-various with 0.5 hits (Figure 10a). The difference was not statistically significant likely due to the high variance in hits in the IncF plasmid group. In terms of

multiple virulence gene hits, col had a range of zero to three hits, whereas IncF plasmids had a range of zero to four hits (Figure 10b).

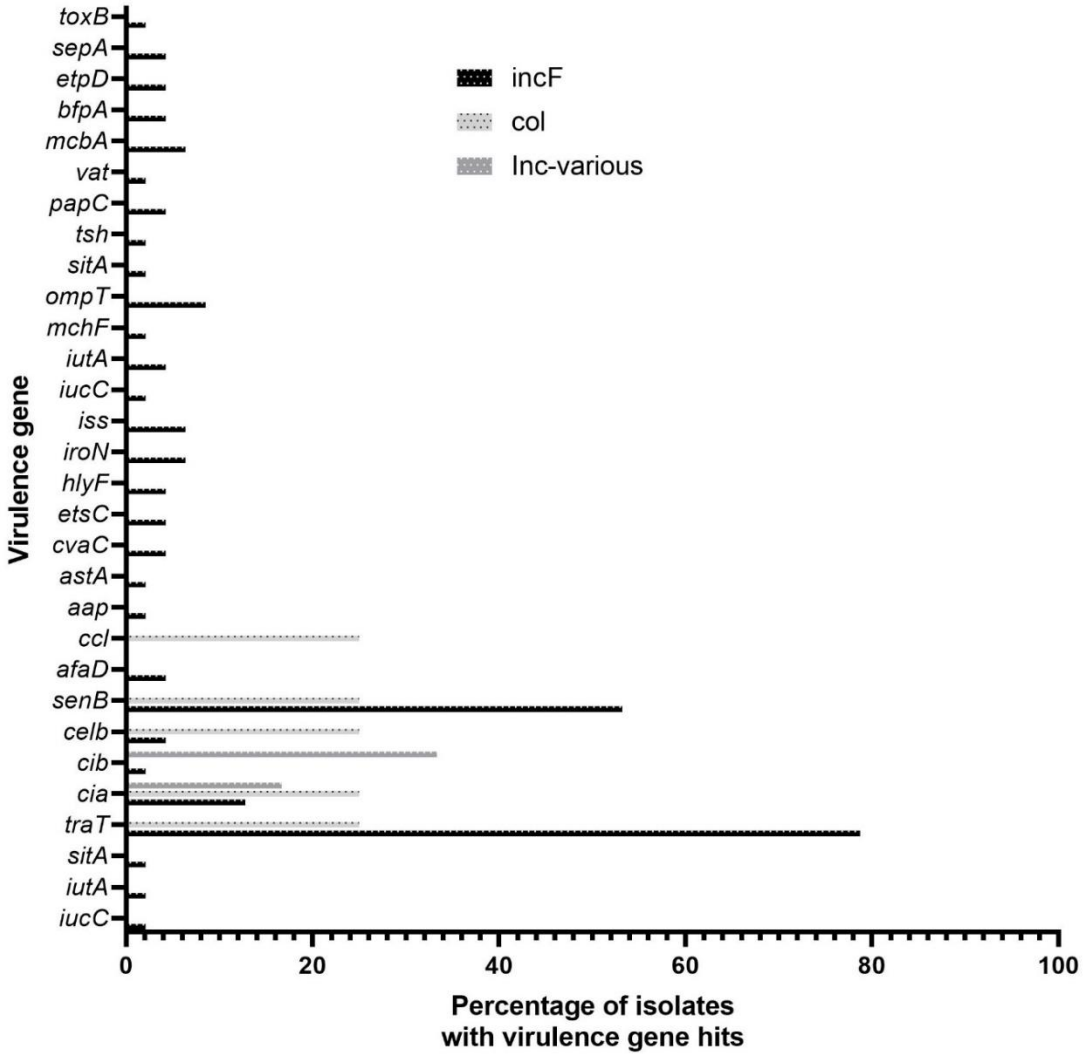
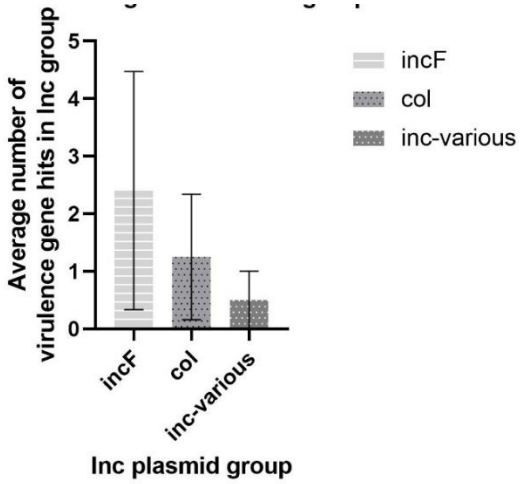


Figure 9. Virulence Genes in Urinary Plasmidic Assemblies.

Percentage of isolates from plasmid group predicted to have a given virulence gene. F plasmids had the largest variety and proportion of virulence gene hits. The most common virulence genes were traT (blocks invading plasmids) and senB (F plasmid-linked enterotoxin). Virulence genes predicted in col plasmids are col genes (*ccl, celb, cib, cia*).

(a)



(b)

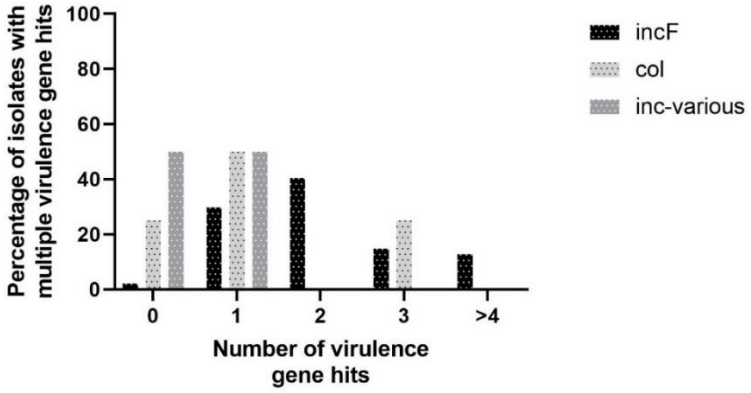


Figure 10. Virulence Genes in Urinary *E. coli* Plasmids.

(a) Average number of virulence genes in plasmids from each plasmid groups. The difference between plasmid groups is not significant due to the high variance in IncF plasmids (b) Percentage of multiple virulence gene hits in isolates from plasmid groups. IncF plasmids have a higher proportion of plasmids with multiple virulence gene hits.

To assess miscellaneous genes of interest outside of those scanned above, the ORFs analyzed by Prokka and Usearch were processed into a list of distinct ORFs in each plasmid group. F plasmids have

the longest list of distinct ORFs with an assigned function, including resistance, fitness, and virulence factors distinct from those identified by ResFinder and VirulenceFinder (Table 18). These genes include plasmid replication machinery, metal transport and resistance genes, leukotoxin genes, multi-drug transporters, phage genes, and virulence regulators. Col plasmids have the second highest count of distinct ORFs, including plasmid replication and persistence genes, metal transport mechanism, and virulence genes consisting of not just those associated with colicin but also regulators of virulence genes (Table 19). Despite having the largest count of distinct *inc* genes, the inc-various plasmid group had the lowest count of distinct ORFs (Table 20). There is evidence of plasmid replication and retention machinery, virulence genes such as those that encode colicin, the adhesin *yadA*, and the virulence regulator *virB*.

Table 19. ORFs Annotated with a Function in All Urinary *E. coli* F Plasmids.

2-keto-3-deoxygluconate permease	Antitoxin PemI	Carboxy-S-adenosyl-L-methionine synthase	Diacetylchitobiose uptake system permease protein NgcG	galactopyranose 3-N-acetyltransferase
34 kDa membrane antigen	Antitoxin VapB	Cardiolipin synthase	Dihydrofolate reductase type 1	Elloramycin glycosyltransferase ElmGT
3',5'-cyclic adenosine monophosphate phosphodiesterase CpdA	Arginine deiminase	Chaperone protein caf1M	Dihydropteroate synthase	Endonuclease YhcR
3'-5' exonuclease DinG	Arginine repressor	Chaperone protein caf1M	Dihydropteroate synthase	Endonuclease YhcR
3-oxo-glucose-6-phosphate:glutamate aminotransferase	Arsenate reductase	Chaperone protein DnaJ	DNA adenine methylase	Endonuclease YhcR
6-hydroxy-3-succinoylpyridine 3-monooxygenase HspA	Arsenate reductase	Chaperone protein PapD	DNA adenine methyltransferase YhdJ	Endonuclease YhcR
Adaptive-response sensory-kinase SasA	Arsenical pump-driving ATPase	Chaperone protein PapD	DNA-binding protein H-NS	Endoribonuclease HigB
Adhesin YadA	Arsenical pump membrane protein	Chaperone protein SicA	DNA-binding protein StpA	Endoribonuclease PemK
Aerobactin synthase	Arsenical resistance operon trans-acting repressor ArsD	Chromosome partition protein Smc	DNA-binding protein StpA	Endoribonuclease toxin MazF
Aldehyde reductase Ahr	Arsenical-resistance protein Acr3	Chromosome partition protein Smc	DNA-cytosine methyltransferase	Endoribonuclease VapD
Allophanate hydrolase	ATM1-type heavy metal exporter	Colicin-A immunity protein	DNA-invertase hin	Enolase
Alpha-D-ribose 1-methylphosphonate 5-phosphate C-P lyase	ATP-dependent Clp protease ATP-binding subunit ClpX	Colicin-E2	DNA-invertase hin	Enterochelin esterase
Alpha-D-ribose 1-methylphosphonate 5-triphosphate synthase subunit PhnI	ATP-dependent Clp protease ATP-binding subunit ClpX	Colicin-E2 immunity protein	DNA-invertase hin	F1 capsule-anchoring protein
Amino-acid permease RocC	ATP-dependent RecD-like DNA helicase	Colicin-E7	DNA-invertase hin	F1 capsule-anchoring protein
Aminoglycoside 3'-phosphotransferase	ATP-dependent RecD-like DNA helicase	Colicin-Ia	DNA-invertase hin	Fe(3+) dicitrate transport ATP-binding protein FecE
Antirestriction protein KlcA	ATP-dependent RNA helicase DbpA	Colicin-Ia	DNA-invertase hin	Ferredoxin--NADP reductase
Antirestriction protein KlcA	Autotransporter adhesin BpaC	Colicin-V	DNA polymerase III subunit alpha	Ferric aerobactin receptor
Antirestriction protein KlcA	Bacteriocin microcin B17	Colicin-V	DNA polymerase III subunit epsilon	Ferric enterobactin receptor
Antitoxin CcdA	Beta-lactamase CTX-M-1	Colicin V secretion protein CvaA	DNA polymerase III subunit theta	Fertility inhibition protein
Antitoxin HicB	Beta-lactamase SHV-1	Conjugal transfer protein TraG	DNA relaxase MbeA	Fimbrial adapter PapF
Antitoxin LsoB	Beta-lactamase TEM	Conjugal transfer protein TraG	DNA relaxase MbeA	Fimbrial adapter PapK
Antitoxin MazE	Carbamate kinase 1	Coupling protein TraD	DNA topoisomerase 3	Fimbrial protein PapE
Antitoxin ParD	Carbamate kinase 1	Coupling protein TraD	DNA topoisomerase 3	Flap endonuclease Xni
	Carbamate kinase 2	Coupling protein TraD	DNA topoisomerase 3	Fumarate reductase flavoprotein subunit
	Carboxylesterase B	Coupling protein TraD	Double-strand break reduction protein	
		Cytosine permease	dTDP-3-amino-3,6-dideoxy-alpha-D-	
		Deoxyuridine 5'-triphosphate nucleotidohydrolase		

Table 19. ORFs Annotated with a Function in All Urinary *E. coli* F Plasmids (continued).

Glucose-1-phosphatase	Inner membrane protein YihN	IS3 family transposase IS103	IS481 family transposase ISKpn28	ISL3 family transposase ISKox3
Glucose-6-phosphate 3-dehydrogenase	Inner membrane protein YqiJ	IS3 family transposase IS1133	IS4 family transposase IS4	ISNCY family transposase ISBcen27
Glutaredoxin 1	Inner membrane protein YqiK	IS3 family transposase IS1133	IS4 family transposase IS421	ISNCY family transposase ISBcen27
Glycerol-3-phosphate regulon repressor	Inner membrane protein YqjE	IS3 family transposase IS1203	IS4 family transposase ISVsa5	ISNCY family transposase ISLad2
Glycine betaine transporter	Iron-sulfur cluster carrier protein	IS3 family transposase IS150	IS4 family transposase ISVsa5	ISNCY family transposase ISRor2
Heat shock protein C	Iron-sulfur cluster carrier protein	IS3 family transposase IS2	IS5 family transposase IS5	Kanosamine-6-phosphate phosphatase
Hemin receptor	IS110 family transposase IS1663	IS3 family transposase IS3	IS5 family transposase IS903	Lactose permease
Hemin transport system permease protein HmuU	IS110 family transposase IS5075	IS3 family transposase IS3	IS630 family transposase ISEc40	Leukotoxin
Hemoglobin-binding protease hbp autotransporter	IS110 family transposase ISEc21	IS3 family transposase IS629	IS66 family transposase ISCro1	Leukotoxin-activating lysine-acyltransferase LtxC
Hemoglobin-binding protease hbp autotransporter	IS110 family transposase ISEc76	IS3 family transposase ISEc16	IS66 family transposase ISEc22	Leukotoxin export ATP-binding protein LtxB
Hemolysin expression-modulating protein Hha	IS110 family transposase ISShdy1	IS3 family transposase ISEc17	IS66 family transposase ISEc23	Leukotoxin export protein LtxD
Hemolysin expression-modulating protein Hha	IS1182 family transposase ISCfr1	IS3 family transposase ISEc31	IS66 family transposase ISEc49	Lipopolysaccharide core heptose(II)-phosphate phosphatase
High-affinity zinc uptake system ATP-binding protein ZnuC	IS1380 family transposase ISEcp1	IS3 family transposase ISEc31	IS66 family transposase ISEc8	Lipoprotein PrgK
Homocysteine S-methyltransferase	IS1595 family transposase ISSsu9	IS3 family transposase ISEc48	IS66 family transposase ISSgsp1	Lipoprotein-releasing system ATP-binding protein LolD
HTH-type transcriptional regulator AppY	IS1 family transposase IS1D	IS3 family transposase ISEc48	IS6 family transposase IS26	Lipoprotein-releasing system transmembrane protein LolC
HTH-type transcriptional regulator CatM	IS21 family transposase IS100kyp	IS3 family transposase ISEcl1	IS6 family transposase IS6100	Lipoprotein YlpA
HTH-type transcriptional regulator PgrR	IS21 family transposase IS1326	IS3 family transposase ISEcl1	IS91 family transposase ISEc37	Lipoprotein YlpA
HTH-type transcriptional regulator YdeO	IS21 family transposase IS21	IS3 family transposase ISKpn37	IS91 family transposase ISEc37	L-lysine N6-monooxygenase
HTH-type transcriptional repressor ComR	IS21 family transposase ISEc10	IS3 family transposase ISKpn8	IS91 family transposase ISSbo1	Lysis protein for colicin N
Ice-binding protein 1	IS256 family transposase IS1414	IS3 family transposase ISLad1	ISKra4 family transposase ISCep1	Lysis protein for colicin N
Inhibitor of g-type lysozyme	IS256 family transposase IS1414	IS3 family transposase ISSty2	ISKra4 family transposase ISEc51	Macrolide export protein MacA
Inner membrane protein YbbJ	IS256 family transposase ISEc39	IS481 family transposase ISAzs36	ISL3 family transposase ISEc53	
	IS30 family transposase IS30	IS481 family transposase ISKpn28	ISL3 family transposase ISKox3	

Table 19. ORFs Annotated with a Function in All Urinary *E. coli* F Plasmids (continued).

Major pilu subunit operon regulatory protein PapB	Modification methylase HpaII	Outer membrane usher protein PapC	Plasmid partition protein A	Protein PsiB
Major structural subunit of bundle-forming pilus	Modulating protein YmoA	Outer membrane usher protein SfmD	Plasmid replication initiator protein TrfA	Protein QmcA
Manganese transport system ATP-binding protein MntB	Modulating protein YmoA	Oxygen-regulated invasion protein OrgA	Plasmid segregation protein ParM	Protein RecA
Manganese transport system membrane protein MntB	mRNA endoribonuclease LsoA	Pap fimbrial major pilin protein	Plasmid segregation protein ParM	Protein SopB
Manganese transport system membrane protein MntB	mRNA interferase toxin RelE	Pap fimbrial major pilin protein	Plasmid segregation protein ParM	Protein StbB
Mating pair stabilization protein TraN	mRNA interferase toxin RelE	Pap fimbrial major pilin protein	Plasmid segregation protein ParM	Protein TraB
Mating pair stabilization protein TraN	Multidrug efflux pump Tap	PAP fimbrial minor pilin protein	Plasmid segregation protein ParM	Protein TraC
Mating pair stabilization protein TraN	Multidrug resistance protein MdtH	PAP fimbrial minor pilin protein	Plasmid segregation protein ParM	Protein TraC
Membrane-bound lytic murein transglycosylase C	Multidrug transporter EmrE	Pentapeptide repeat protein	Prophage tail fiber assembly protein TfaE	Protein TraJ
Mercuric reductase	Multifunctional conjugation protein TraI	Peptide deformylase	Prophage tail fiber assembly protein TfaE	Protein TraJ
Mercuric resistance operon regulatory protein	Multifunctional conjugation protein TraI	Peptide deformylase	Prophage tail fiber assembly protein TfaE	Protein TraJ
Mercuric transport protein MerC	Multifunctional conjugation protein TraI	Peptide deformylase	Protease 7	Protein TraJ
Mercuric transport protein MerT	N(2)-citryl-N(6)-acetyl-N(6)-hydroxylysine synthase	Phospho-2-dehydro-3-deoxyheptonate aldolase, Trp-sensitive	Protease 7	Protein TraJ
Mercuric transport protein periplasmic component	N(6)-hydroxylysine O-acetyltransferase	Phosphoglucosamine mutase	Protease 7	Protein TraJ
Metalloprotease StcE	NADPH-dependent FMN reductase ArsH	Phospholipase D	Protease 7	Protein TraL
Methionine--tRNA ligase	Na(+)-translocating NADH-quinone reductase subunit C	Pilin	Proteasome-associated ATPase	Protein TraL
Microcin B17-processing protein McbB	Nucleoid occlusion protein	Pilin	Protein AfaD	Protein TraM
Microcin B17-processing protein McbC	Nucleoid occlusion protein	Pilin	Protein AfaD	Protein TraQ
Microcin B17-processing protein McbD	Nucleoid occlusion protein	PI protein	Protein FdrA	Protein TraR
Minor fimbrial protein PrsF	Ornithine carbamoyltransferase	PI protein	Protein FdrA	Protein TraR
Mobilization protein MbeC	Outer membrane lipoprotein BfpB	PI protein	Protein FlmC	Protein TraS
	Outer membrane lipoprotein BfpB	Plasmid-derived single-stranded DNA-binding protein	Protein FlmC	Protein TraS
	Outer membrane porin C	Plasmid-derived single-stranded DNA-binding protein	Protein FlmC	Protein TraV
		Plasmid partition protein A	Protein KlcB	Protein TraV
		Plasmid partition protein A	Protein PndA	Protein UmuC
			Protein PrgH	Protein UmuC
			Protein PrgI	Protein UmuC
			Protein PsiB	Protein UmuC
				Protein UmuC
				Protein UmuC
				Protein UmuC
				Protein UmuC
				Protein UmuC
				Protein UmuD

Table 19. ORFs Annotated with a Function in All Urinary *E. coli* F Plasmids (continued).

Protein UmuD	putative mRNA interferase toxin HicA	Relaxosome protein TraY	Serine recombinase PinR	Tetracycline repressor protein class B from transposon Tn10
Protein UmuD	putative periplasmic iron-binding protein	Relaxosome protein TraY	S-fimbrial protein subunit SfaA	Tetracycline repressor protein class H
Protein UmuD	Putative phosphonates utilization ATP-binding protein PhnK	Relaxosome protein TraY	S-formylglutathione hydrolase FrmB	Tetracycline resistance protein, class B
Protein UmuD	putative protein	RepFIB replication protein A	S-(hydroxymethyl)glutathione dehydrogenase	Tetracycline resistance protein, class B
Protein virB10	Putative protein-methionine-sulfoxide reductase subunit YedZ1	RepFIB replication protein A	Signal recognition particle 54 kDa protein	Tetracycline resistance protein, class C
Protein VirD4	Putative protein-methionine-sulfoxide reductase subunit YedZ1	RepFIB replication protein A	Single-stranded DNA-binding protein	Thermonuclease
Pullulanase secretion protein PulS	putative protein YdfB	Replication initiation protein	Single-stranded DNA-binding protein	Thiamine import ATP-binding protein ThiQ
putative 2-dehydro-3-deoxy-D-pentonate aldolase YjhH	putative protein YggR	Replication initiation protein	Single-stranded DNA-binding protein	Thiol-disulfide oxidoreductase ResA
putative ABC transporter ATP-binding protein	putative protein YjiK	Replication initiation protein	Single-stranded DNA-binding protein	Thiol-disulfide oxidoreductase ResA
putative ABC transporter permease protein	putative protein YncE	Replication initiation protein	Single-stranded DNA-binding protein	Thymidylate synthase 2
Putative anti-FlhC(2)FlhD(4) factor YdiV	putative signaling protein	Replication initiation protein	sn-glycerol-3-phosphate-binding periplasmic protein UgpB	Tn3 family transposase
putative chromate transport protein	putative TonB-dependent receptor	Replication initiation protein	sn-glycerol-3-phosphate import ATP-binding protein UgpC	Tn3 family transposase ISEc63
putative fimbrial chaperone LpfB	Putative transposon Tn552 DNA-invertase bin3	Replication protein RepA	sn-glycerol-3-phosphate transport system permease protein UgpA	Tn3 family transposase ISPa38
Putative fluoride ion transporter CrcB	Putative transposon Tn552 DNA-invertase bin3	Replicative DNA helicase	SPbeta prophage-derived aminoglycoside N(3')-acetyltransferase-like protein YokD	Tn3 family transposase ISSba14
putative HTH-type transcriptional regulator RhmR	Quorum-quenching protein AidA	Ribonuclease H	Stable plasmid inheritance protein	Tn3 family transposase ISYps3
putative HTH-type transcriptional regulator YahB	Recombination-associated protein RdgC	Ribonucleoside-diphosphate reductase 1 subunit alpha	Streptomycin 3''-adenylyltransferase	Tn3 family transposase Tn2
putative HTH-type transcriptional regulator YahB	Recombination-associated protein RdgC	Ribonucleoside-diphosphate reductase 1 subunit beta	Streptomycin 3''-adenylyltransferase	Tn3 family transposase TnAs1
putative HTH-type transcriptional regulator YbaQ	Regulatory protein RepA	RNA chaperone ProQ	Sugar transporter SemiSWEET	Tn3 family transposase TnAs1
putative HTH-type transcriptional regulator YbaQ	Regulatory protein rop	RNA chaperone ProQ	Sulfoacetaldehyde reductase	Toxin and drug export protein A
putative HTH-type transcriptional regulator YddM	Regulatory protein rop	Secreted chorismate mutase	TelA-like protein	Toxin CcdB
Putative metal chaperone YciC	Relaxosome protein TraM	Secreted effector protein EspF(U)	Tetracycline repressor protein class A from transposon 1721	Toxin HigB-1
	Relaxosome protein TraM	Secretin PulD		Toxin HigB-1
	Relaxosome protein TraM	Serine protease EspC		Toxin ParE
		Serine protease SepA autotransporter		Toxin RelE2
				Toxin RTX-I translocation ATP-binding protein

Table 19. ORFs Annotated with a Function in All Urinary *E. coli* F Plasmids (continued).

Transcriptional regulator	TrfB transcriptional repressor protein	Type IV secretion system protein PtlE	Type IV secretion system protein virB5	Tyrosine recombinase XerC
Transcriptional regulator HilA	tRNA(fMet)-specific endonuclease VapC	Type IV secretion system protein PtlH	Type IV secretion system protein VirB6	Tyrosine recombinase XerC
Transcriptional regulatory protein RcsB	tRNA(fMet)-specific endonuclease VapC	Type IV secretion system protein virB1	Type IV secretion system protein virB8	Tyrosine recombinase XerD
Transcriptional regulatory protein WalR	tRNA nuclease CdiA	Type IV secretion system protein VirB11	Type IV secretion system protein virB8	UDP-glucose:protein N-beta-glucosyltransferase
Transcriptional repressor PifC	Type-2 restriction enzyme EcoRI	Type IV secretion system protein VirB11	Type IV secretion system protein virB8	Virulence regulon transcriptional activator VirB
Transcriptional repressor protein KorB	Type-2 restriction enzyme EcoRII	Type IV secretion system protein VirB11	Type IV secretion system protein virB8	Virulence regulon transcriptional activator VirB
Transcription antitermination protein RfaH	Type 3 secretion system secretin	Type IV secretion system protein VirB11	Type IV secretion system protein virB9	Virulence regulon transcriptional activator VirB
Transcription antitermination protein RfaH	Type 4 prepilin-like proteins leader peptide-processing enzyme	Type IV secretion system protein VirB11	Type IV secretion system protein virB9	Virulence regulon transcriptional activator VirB
Transcription antitermination protein RfaH	Type II secretion system protein C	Type IV secretion system protein VirB11	Type IV secretion system protein virB9	Virulence regulon transcriptional activator VirF
Transposase for transposon Tn5	Type II secretion system protein E	Type IV secretion system protein virB4	Type IV secretion system protein virB9	Vitamin B12-binding protein
Transposon gamma-delta resolvase	Type II secretion system protein F	Type IV secretion system protein virB4	Tyrosine recombinase XerC	
Transposon Tn10 TetC protein	Type II secretion system protein G	Type IV secretion system protein virB4	Tyrosine recombinase XerC	
Transposon Tn10 TetD protein	Type II secretion system protein J	Type IV secretion system protein virB4	Tyrosine recombinase XerC	
Transposon Tn3 resolvase	Type II secretion system protein L	Type IV secretion system protein virB4	Tyrosine recombinase XerC	
Transposon Tn3 resolvase	Type II secretion system protein M	Type IV secretion system protein virB4	Tyrosine recombinase XerC	
	Type I restriction enzyme EcoR124II R protein	Type IV secretion system protein virB5	Tyrosine recombinase XerC	

Table 20. ORFs Annotated with a Function in All Urinary *E. coli* Col Plasmids.

Antirestriction protein KlcA	galactopyranose 3-N-acetyltransferase	Magnesium transport protein CorA	putative protein YggR	Tn3 family transposase ISYps3
Antitoxin CcdA	Glutamine--tRNA ligase	Mobilization protein MbeC	putative protein YncE	Tn3 family transposase Tn3
Arsenate reductase	Glycine betaine transporter	mRNA interferase toxin RelE	putative TonB-dependent receptor	Toxin CcdB
ATP-dependent Clp protease ATP-binding subunit ClpX	Inner membrane protein YqiJ	mRNA interferase toxin RelE	Putative transposon Tn552 DNA-invertase bin3	Toxin HigB-1
ATP-dependent RecD-like DNA helicase	Inner membrane protein YqiK	Nucleoid occlusion protein	Recombination-associated protein RdgC	Toxin RelE2
Beta-lactamase TEM	Iron-sulfur cluster carrier protein	Outer membrane lipoprotein BfpB	Recombination-associated protein RdgC	Transcription antitermination protein RfaH
Chromosome partition protein Smc	IS110 family transposase ISEsa2	Phosphoethanolamine transferase OpgE	Regulatory protein rop	Transcription antitermination protein RfaH
Cloacin immunity protein	IS3 family transposase IS2	Phospholipase D	Regulatory protein rop	Transposon gamma-delta resolvase
Colicin-E2 immunity protein	IS3 family transposase IS3411	Phospholipase D	Regulatory protein rop	Transposon Tn3 resolvase
Colicin-E3	IS3 family transposase IS629	Plasmid-derived single-stranded DNA-binding protein	Regulatory protein rop	Type I restriction enzyme EcoR124II R protein
Colicin-E7	IS3 family transposase IS629	Plasmid partition protein A	Replication initiation protein	Type IV secretion system protein VirB11
Colicin-Ia	IS3 family transposase ISSd1	Plasmid segregation protein ParM	Replication initiation protein	Type IV secretion system protein virB4
Colicin-Ia	IS3 family transposase ISSfl10	Plasmid segregation protein ParM	Replication protein RepA	Type IV secretion system protein virB8
Coupling protein TraD	IS5 family transposase ISKpn26	Protein PndA	Replicative DNA helicase	Type IV secretion system protein virB9
Deoxyuridine 5'-triphosphate nucleotidohydrolase	IS6 family transposase IS15DII	Protein PsiB	RNA chaperone ProQ	Tyrosine recombinase XerC
DNA adenine methylase	ISNCY family transposase ISRor2	Protein UmuC	Single-stranded DNA-binding protein	Tyrosine recombinase XerC
DNA adenine methyltransferase YhdJ	Lipopolysaccharide core heptose(II)-phosphate phosphatase	Protein UmuD	Thermonuclease	Virulence regulon transcriptional activator VirB
DNA-invertase hin	Lysis protein for colicin N	Protein UmuD	Tn3 family transposase	
DNA relaxase MbeA		putative HTH-type transcriptional regulator YbaQ	Tn3 family transposase ISEc63	
dTDP-3-amino-3,6-dideoxy-alpha-D-			Tn3 family transposase ISSba14	
			Tn3 family transposase ISSba14	

Table 21. ORFs Annotated with a Function in All Urinary *E. coli* Inc-various Plasmids.

3'-5' exonuclease DinG	Plasmid partition protein A	Type IV secretion system protein virB9
Adhesin YadA	Plasmid segregation protein ParM	Type IV secretion system protein virB9
Adhesin YadA	Proteasome-associated ATPase	Type IV secretion system protein virB9
Antirestriction protein KlcA	Protein PndA	Tyrosine recombinase XerC
ATP-dependent RecD-like DNA helicase	Protein PsiB	Tyrosine recombinase XerC
Colicin-Ia	Protein UmuC	Virulence regulon transcriptional activator VirB
Coupling protein TraD	Protein UmuD	
DNA adenine methyltransferase YhdJ	Protein UmuD	
DNA-binding protein StpA	Protein VirD4	
DNA-invertase hin	putative protein YggR	
DNA-invertase hin	Putative transposon Tn552 DNA-invertase bin3	
DNA polymerase III subunit theta	Putative transposon Tn552 DNA-invertase bin3	
DNA primase TraC	Recombination-associated protein RdgC	
DNA topoisomerase 3	RepFIB replication protein A	
DNA topoisomerase 3	Replication initiation protein	
Endonuclease YhcR	Replicative DNA helicase	
IS110 family transposase ISEsa2	Single-stranded DNA-binding protein	
IS3 family transposase IS3	Single-stranded DNA-binding protein	
IS5 family transposase IS903	Single-stranded DNA-binding protein	
IS6 family transposase IS15	Transcription antitermination protein RfaH	
ISNCY family transposase ISRor2	Type IV secretion system protein PtlE	
ISNCY family transposase ISSen7	Type IV secretion system protein PtlE	
Lipopolysaccharide core heptose(II)-phosphate phosphatase	Type IV secretion system protein VirB11	
Modulating protein YmoA	Type IV secretion system protein VirB11	
Modulating protein YmoA	Type IV secretion system protein VirB11	
mRNA interferase toxin RelE	Type IV secretion system protein virB4	
Nucleoid occlusion protein	Type IV secretion system protein virB4	
Phospholipase D	Type IV secretion system protein virB4	
Phospholipase D	Type IV secretion system protein virB8	
Phospholipase D		
PI protein		
Plasmid-derived single-stranded DNA-binding protein		

Urinary Plasmid Conjugation to *E. coli* K-12

If urinary plasmids reduce the permissivity to phage infection in *E. coli*, then we expect that acquisition of these plasmids by phage-susceptible *E. coli* K-12 would decrease permissivity. Given the data thus far (e.g., presence of *pemIK* and *traT* genes in less phage-susceptible isolates), the primary plasmid target for having an anti-phage effect are the F plasmids, which often code for conjugation machinery. To move plasmids from a urinary background to a control background, the plasmid donor and recipient must have different selection markers. The urinary *E. coli* isolates were grown on plates containing antibiotics that can be used as selection markers (Table 21). Overall, the most common selection marker that urinary isolates could grow on was ampicillin (41.2%) followed by tetracycline (23.5%). The growth on selection markers was analyzed based on incompatibility groups, with growth on ampicillin being the most common in all groups, followed by tetracycline (Figure 11a). The IncF group had isolates that could grow on all antibiotic plates tested. Isolates of the IncF group were most likely to grow on more than one antibiotic, with isolates UMB5924 and UMB1284 growing on four out of five antibiotics (Figure 11b, Table 21).

The annotations for the urinary *E. coli* plasmids for isolates that could grow on tetracycline were reviewed for the presence of tetracycline (Tc) resistance genes. From these, the annotations were reviewed for evidence of plasmid transfer genes (i.e., *tra* genes). Five isolates were identified with these criteria and became plasmid donor candidates; none of these isolates grew on chloramphenicol (Cm), the marker used to select *E. coli* K-12 conjugation recipients (Table 22). Two main constructs of *E. coli* K-12 were utilized, one with the selection marker on an empty vector plasmid (pCA14n-Cm) and the other with the selection marker cassette in the chromosome (i.e., *yfiQ::Cm*). Both plasmid recipients generated transconjugants on double antibiotic (Cm and Tc) plates following conjugation. As a negative control for the plasmid recipient, a WT *E. coli* K-12 strain (MG1655) was used since it does not carry a selection marker and did not generate transconjugants under selection. For the negative control plasmid

donors, two urinary isolates were used that met all the criteria for conjugation except that they were not predicted to carry a plasmid, and these isolates did not generate transconjugants under selection. Table 6 and Table 7 lists all the plasmid recipients and constructs tested, respectively. Multiple manipulations of the *yfiQ cobB* acetylation system were tested, in addition to the *E. coli* K-12 BW25113 strain background, and the kanamycin selection marker.

Table 22. Growth of Urinary *E. coli* on Antibiotic Selection Plates.

		Total isolates	68	28	1	3	8	16
		Percentage	100.00	41.18	1.47	4.41	11.76	23.53
Strain	Inc group	# abx that it grew on	LB	Am	Cm	Kn	Spc	Tc
B	Control	0	Yes	No	No	No	No	No
C	Control	0	Yes	No	No	No	No	No
K-12	Control	0	Yes	No	No	No	No	No
UMB0103	IncFII	3	Yes	Yes	Yes	No	No	Yes
UMB0149	Inc	0	Yes	No	No	No	No	No
UMB0276	None	1	Yes	Yes	No	No	No	No
UMB0527	IncFII	0	Yes	No	No	No	No	No
UMB0731	Col	0	Yes	No	No	No	No	No
UMB0906	IncFII	1	Yes	Yes	No	No	No	No
UMB0923	Inc	0	Yes	No	No	No	No	No
UMB0928	IncFI	2	Yes	Yes	No	No	No	Yes
UMB0931	IncFII	2	Yes	Yes	No	No	No	Yes
UMB0933	IncFII	1	Yes	No	No	No	Yes	No
UMB0934	IncFII	2	Yes	Yes	No	No	No	Yes
UMB0939	Col	2	Yes	Yes	No	No	No	Yes
UMB0949	IncFII	2	Yes	Yes	No	No	No	Yes
UMB1012	IncFII	1	Yes	Yes	No	No	No	No
UMB1091	IncFII	3	Yes	Yes	No	No	Yes	Yes
UMB1093	IncFI	1	Yes	Yes	No	No	No	No
UMB1160	IncFII	1	Yes	No	No	No	Yes	No
UMB1162	IncFII	1	Yes	No	No	No	No	Yes
UMB1180	Col	0	Yes	No	No	No	No	No
UMB1193	IncFII	3	Yes	Yes	No	No	Yes	Yes
UMB1195	IncFII	0	Yes	No	No	No	No	No
UMB1202	IncFII	0	Yes	No	No	No	No	No
UMB1220	None	0	Yes	No	No	No	No	No
UMB1221	IncFII	1	Yes	No	No	No	No	Yes
UMB1223	IncFII	2	Yes	Yes	No	No	No	Yes
UMB1225	None	0	Yes	No	No	No	No	No
UMB1228	Inc	0	Yes	No	No	No	No	No
UMB1229	IncFII	3	Yes	Yes	No	No	Yes	Yes
UMB1284	IncFII	4	Yes	Yes	No	Yes	Yes	Yes
UMB1285	IncFII	0	Yes	No	No	No	No	No
UMB1335	IncFII	1	Yes	Yes	No	No	No	No

Table 22. Growth of Urinary *E. coli* on Antibiotic Selection Plates (Continued).

UMB1337	IncFII	1	Yes	Yes	No	No	No	No
UMB1346	IncFII	0	Yes	No	No	No	No	No
UMB1347	IncFII	0	Yes	No	No	No	No	No
UMB1348	IncFII	1	Yes	Yes	No	No	No	No
UMB1354	None	0	Yes	No	No	No	No	No
UMB1356	None	0	Yes	No	No	No	No	No
UMB1358	None	0	Yes	No	No	No	No	No
UMB1359	None	0	Yes	No	No	No	No	No
UMB1360	IncFII	1	Yes	Yes	No	No	No	No
UMB1362	IncFI	2	Yes	Yes	No	No	No	Yes
UMB1526	None	1	Yes	Yes	No	No	No	No
UMB1727	IncFII	1	Yes	Yes	No	No	No	No
UMB2019	Col	0	Yes	No	No	No	No	No
UMB2055	Inc	0	Yes	No	No	No	No	No
UMB2321	Inc	0	Yes	No	No	No	No	No
UMB2328	Inc	0	Yes	No	No	No	No	No
UMB3538	IncFII	2	Yes	Yes	No	Yes	No	No
UMB3641	IncFII	2	Yes	Yes	No	No	Yes	No
UMB3643	Inc	1	Yes	Yes	No	No	No	No
UMB4656	IncFII	1	Yes	Yes	No	No	No	No
UMB4716	IncFII	0	Yes	No	No	No	No	No
UMB4746	IncFII	0	Yes	No	No	No	No	No
UMB5337	None	0	Yes	No	No	No	No	No
UMB5814	IncFII	0	Yes	No	No	No	No	No
UMB5924	None	4	Yes	Yes	No	Yes	Yes	Yes
UMB5978	IncFI	0	Yes	No	No	No	No	No
UMB6454	IncFI	0	Yes	No	No	No	No	No
UMB6471	IncFII	0	Yes	No	No	No	No	No
UMB6611	IncFI	0	Yes	No	No	No	No	No
UMB6655	IncFI	0	Yes	No	No	No	No	No
UMB6713	IncFII	1	Yes	Yes	No	No	No	No
UMB6721	IncFII	1	Yes	No	No	No	No	Yes
UMB6890	IncFI	0	Yes	No	No	No	No	No
UMB7431	IncFII	0	Yes	No	No	No	No	No

Note: Ampicillin (100 ug/ml), Chloramphenicol (25 ug/ml), Kanamycin (40 ug/ml), Spectinomycin (100 ug/ml), Tetracycline (15 ug/ml).

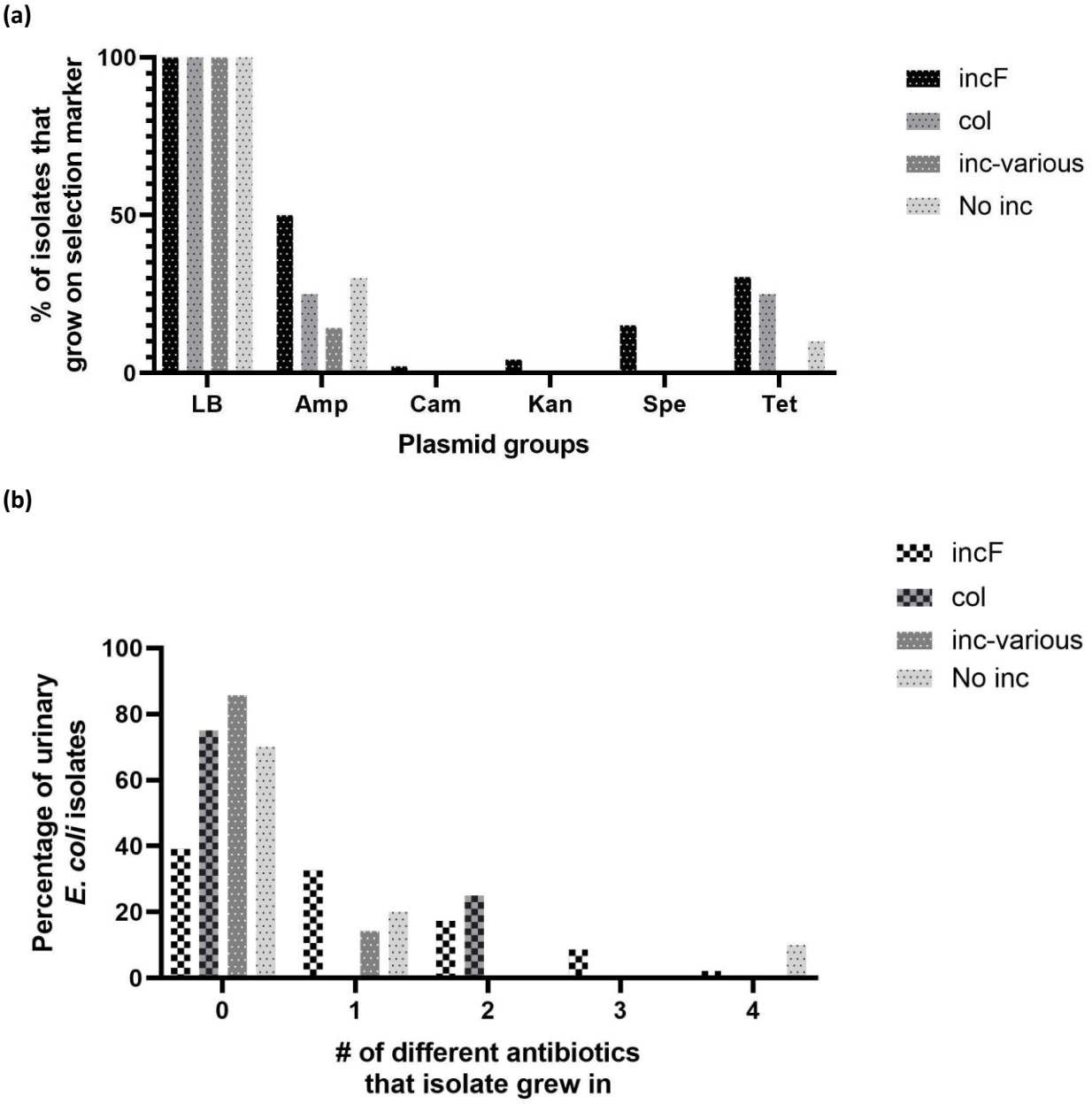


Figure 11. Inc Group and Growth on Antibiotic Plates.

(a) Percentage of urinary *E. coli* isolates from a plasmid group that grew on a given antibiotic plate. (b) Percentage of urinary *E. coli* isolates from a plasmid group that grew on multiple antibiotic plates.

Table 23. Urinary *E. coli* Plasmid Donors, Recipients, and Transconjugants.

Strain	Background	Conjugation role	BLAST hit	Conjugation machinery	Cm resistance on chromosome	Tc resistance on assembly	Cm plate	Tc plate	Cm/Tc plate	Generated transconjugants on Cm/Tc plates
AJW1776	MG1655	Plasmid recipient negative control	Chromosomal	No	No	No	No	No	No	No
AJW4793	MG1655 pCA24n-Cm	Plasmid recipient	Chromosomal	No	Yes	No	Yes	No	No	Yes
AJW5116	MG1655 *yfiQ*::Cm, *cobB*::FRT	Plasmid recipient	Chromosomal	No	Yes	No	Yes	No	No	Yes
UMB1284	Urinary	Plasmid donor	Plasmid	Yes	No	Yes	No	Yes	No	Yes
UMB0928	Urinary	Plasmid donor	Plasmid	Yes	No	Yes	No	Yes	No	Yes
UMB1223	Urinary	Plasmid donor	Plasmid	Yes	No	Yes	No	Yes	No	Yes
UMB6721	Urinary	Plasmid donor	Plasmid	Yes	No	Yes	No	Yes	No	Yes
UMB1091	Urinary	Plasmid donor	Plasmid	Yes	No	Yes	No	Yes	No	Yes
UMB0939	Urinary	Donor negative control	Chromosomal	Yes	No	No	No	Yes	No	No
UMB1362	Urinary	Plasmid donor negative control	Chromosomal	Yes	No	No	No	Yes	No	No

Urinary *E. coli* Plasmids Reduce Phage Infection Permissivity

The phage spot titration assay was used to assess permissibility of phage infection by P1vir, Greed and Lust) of the *E. coli* K-12 transconjugants from Table 22 (Figure 12). The urinary *E. coli* isolates used as plasmid donors were not permissive to the phage at any concentrations tested (Table 23). In contrast, the *E. coli* K-12 recipients (MG1655WT, MG1655 pCA24n-Cm, and MG1655 *yfiQ::*Cm *cobB::*FRT) were susceptible at every concentration tested, including dilution by eight orders of magnitude to 10^2 pfu/mL (Table 23). Urinary plasmids from five isolates were conjugated into the MG1655 pCA24n background; only the pU0928 and pU1284 transconjugants exhibited a permissivity profile that differed from that of the recipient control (Table 24). These transconjugants were only permissive until the third titration (10^6 pfu/mL), indicating a decrease in phage infection permissibility. Similar results were observed in the MG1655 *yfiQ*::Cm *cobB*::FRT background on acquisition of the same two plasmids: pU0928 and pU1284 (Table 25). Constructs were made to assess if complementation of *yfiQ*, a different *E. coli* K-12 background, or a different gene deletion influenced permissivity. None of these constructs by themselves changed permissivity to infection, only decreasing when the pU0928 was conjugated into construct (Table 26).

Growth curves were used to assess the effect of the urinary *E. coli* plasmids on *E. coli* growth during phage infection at a multiplicity of infection (MOI) of 0.0, 0.01, and 10.0. Infection of an *E. coli* K-12 pU0928 transconjugant (either MG1655 pCA24n-Cm or MG1655 *yfiQ*::Cm) with P1vir resulted in comparable optical density (OD_{600}) to uninfected control at all time points (Figure 13a). Infection of the *E. coli* K-12 pU0928 transconjugant with Greed at an MOI of 0.01 resulted in growth characteristics similar to the uninfected control (Figure 13b). Increasing the MOI of Greed to 10.0 resulted in growth characteristics comparable to the control infected with an MOI of 0.01. The 10^3 difference in MOI indicates that acquisition of pU0928 decreases *E. coli* K-12's permissibility to phage infection. Infection with Lust resulted in growth characteristics similar to those of the P1vir infection, with transconjugants

infected with phage at MOI of 0.01 and 10.0 exhibiting comparable growth to the uninfected control (Figure 13c). This experiment was repeated with the *E. coli* K-12 pU1284 and *E. coli* K-12 pU1223 transconjugants. The *E. coli* K-12 pU1284 transconjugant had results comparable to those of the *E. coli* K-12 pU0928 transconjugant (Figure 14a-c). In contrast, the *E. coli* K-12 pU1223 transconjugant had results comparable to the control, indicating that pU1223 does not confer a protective effect that reduces phage infection permissibility (Figure 15a-c).

Acquisition of urinary plasmids by *E. coli* K-12 AJW5116 led to growth on antibiotic plates, like the pattern of the plasmid donor parent (Table 27). Given that pU0928 and pU1284 reduced phage permissibility in *E. coli* K-12, the stability of pU0928 was tested in the absence of selection after passaging for multiple days to assess if the plasmid could be retained over an extended time frame. The pU0928 plasmid was tracked using the tetracycline resistance selection marker in UMB0928 (i.e., plasmid donor), two *E. coli* K-12 pU0928 transconjugants, and the negative control K-12 without pU0928. Even in the absence of selection during passaging, colony growth in tetracycline plates was comparable to growth in LB plates up to ten days (Figure 16). To identify a mechanism that could explain this long-term stability, the plasmidic assembly for UMB0928 was reviewed for plasmid retention genes; UMB0928 is predicted to code for two TA systems, *ccdAB* and *pemIK*.

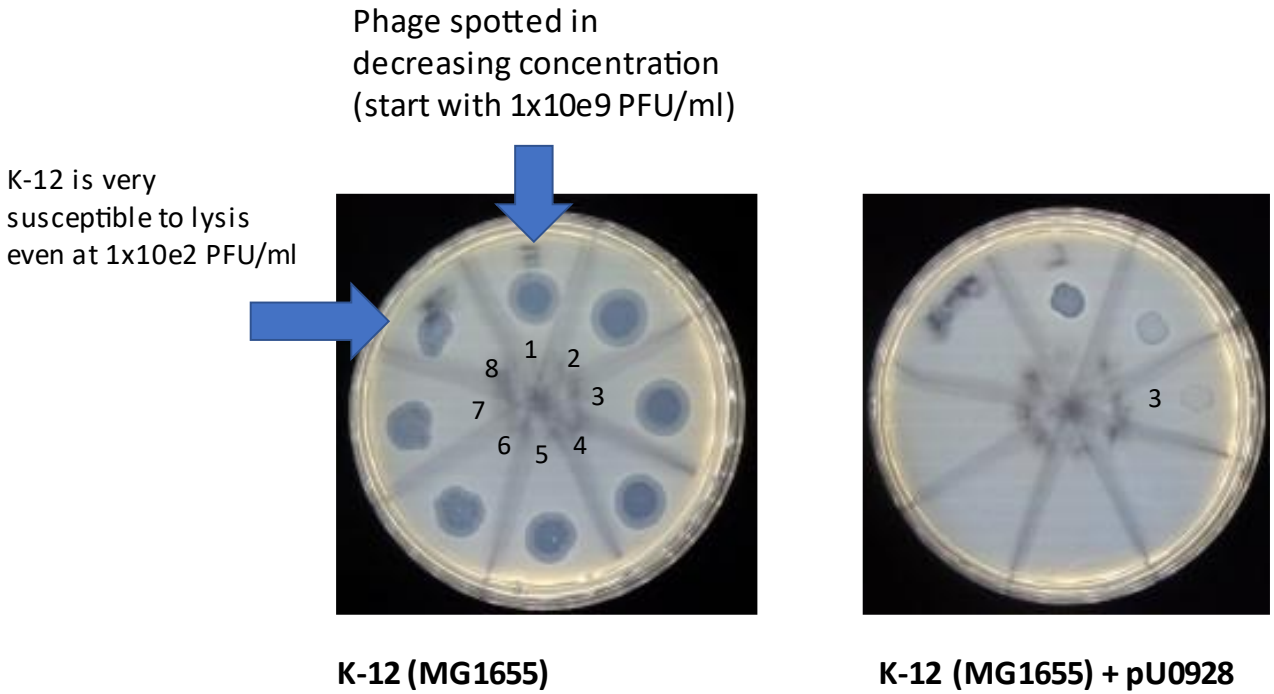


Figure 12. Spot Titration Phenotype in *E. coli* K-12 Transconjugants.

Phage (P1vir, Greed, Lust respectively) were spotted in decreasing titrations on *E. coli* K-12 lawns. The lowest titration where a phage spot was visualized was given as a score for that plate (e.g., *E. coli* K-12 MG1655 was given a score of 8, *E. coli* K-12 MG1655 pU0928 had a score of 3).

Table 24. Spot Titration Phenotype of Conjugation Parents for K-12 Transconjugants.

E. coli strain	Construct	Mutation	ASKA plasmid	Bladder plasmid	Marker	P1vir	Lust	Greed
Urinary	UMB0928	N/A	None	pU0928	Tc	No spot	No spot	No spot
Urinary	UMB1091	N/A	None	pU1091	Tc	No spot	No spot	No spot
Urinary	UMB1223	N/A	None	pU1223	Tc	No spot	No spot	No spot
Urinary	UMB1284	N/A	None	pU1284	Tc	No spot	No spot	No spot
Urinary	UMB6721	N/A	None	pU6721	Tc	No spot	No spot	No spot
MG1655	AJW1776	WT	None	None	None	8	8	8
MG1655	AJW4793	WT	pCA24n-Empty	None	Cm	8	8	8
MG1655	AJW5116	*yfiQ::Cm, cobB::FRT*	None	None	Cm	8	8	8

Table 25. Spot Titration Phenotype in K-12 MG1655 pCA24n-Cm Background.

E. coli strain	Construct	Mutation	ASKA plasmid	Bladder plasmid	Marker	P1vir	Lust	Greed
MG1655	AJW4793	WT	pCA24n-Empty	pU0928	Cm, Tc	3	4	3
MG1655	AJW4793	WT	pCA24n-Empty	pU1091	Cm, Tc	8	8	8
MG1655	AJW4793	WT	pCA24n-Empty	pU1223	Cm, Tc	8	8	8
MG1655	AJW4793	WT	pCA24n-Empty	pU1284	Cm, Tc	3	4	3
MG1655	AJW4793	WT	pCA24n-Empty	pU6721	Cm, Tc	8	8	8

Table 26. Spot Titration Phenotype in K-12 MG1655 *yfiQ::Cm, cobB::FRT* Background.

E. coli strain	Construct	Mutation	ASKA plasmid	Bladder plasmid	Marker	P1vir	Lust	Greed
MG1655	AJW5035	*yfiQ::Kn*	None	None	Kn	8	8	8
MG1655	AJW1776	WT	pCA24n-*yfiQ*	None	Cm	8	8	8
MG1655	AJW5035	*yfiQ::Kn*	pCA24n-*yfiQ*	None	Kn, Cm	7	8	8
MG1655	AJW5035	*yfiQ::Kn*	pCA24n-Empty	None	Kn, Cm	7	8	8
MG1655	AJW5184	*yfiQ::Cm*	None	None	Cm	8	8	8
MG1655	AJW5035	*yfiQ::Kn*	pCA24n-*yfiQ*	pU0928	Cm, Kn, Tc	3	4	3
MG1655	AJW5035	*yfiQ::Kn*	pCA24n-Empty	pU0928	Cm, Kn, Tc	4	5	4
MG1655	AJW5184	*yfiQ::Cm*	None	pU0928	Cm, Tc	3	3	3
BW25113	AJW4688	*yfiQ::Kn*	None	None	Kn, Tc	8	8	8
BW25113	AJW4688	*yfiQ::Kn*	None	pU0928	Kn, Tc	3	3	3
MG1655	AJW5037	*cobB::Cm*	None	None	Cm	8	8	8
MG1655	AJW5037	*cobB::Cm*	None	pU0928	Cm, Tc	2	3	3

Table 27. Spot Titration Phenotype in K-12 *yfiQ* Mutant Background.

E. coli strain	Construct	Mutation	ASKA plasmid	Bladder plasmid	Marker	P1vir	Lust	Greed
MG1655	AJW5116	*yfiQ::Cm, cobB::FRT*	None	None	Cm	8	8	8
MG1655	AJW5116	*yfiQ::Cm, cobB::FRT*	None	pU0928	Cm, Tc	3	3	3
MG1655	AJW5116	*yfiQ::Cm, cobB::FRT*	None	pU1091	Cm, Tc	8	8	8
MG1655	AJW5116	*yfiQ::Cm, cobB::FRT*	None	pU1223	Cm, Tc	8	8	8
MG1655	AJW5116	*yfiQ::Cm, cobB::FRT*	None	pU1284	Cm, Tc	3	3	3
MG1655	AJW5116	*yfiQ::Cm, cobB::FRT*	None	pU6721	Cm, Tc	8	8	8

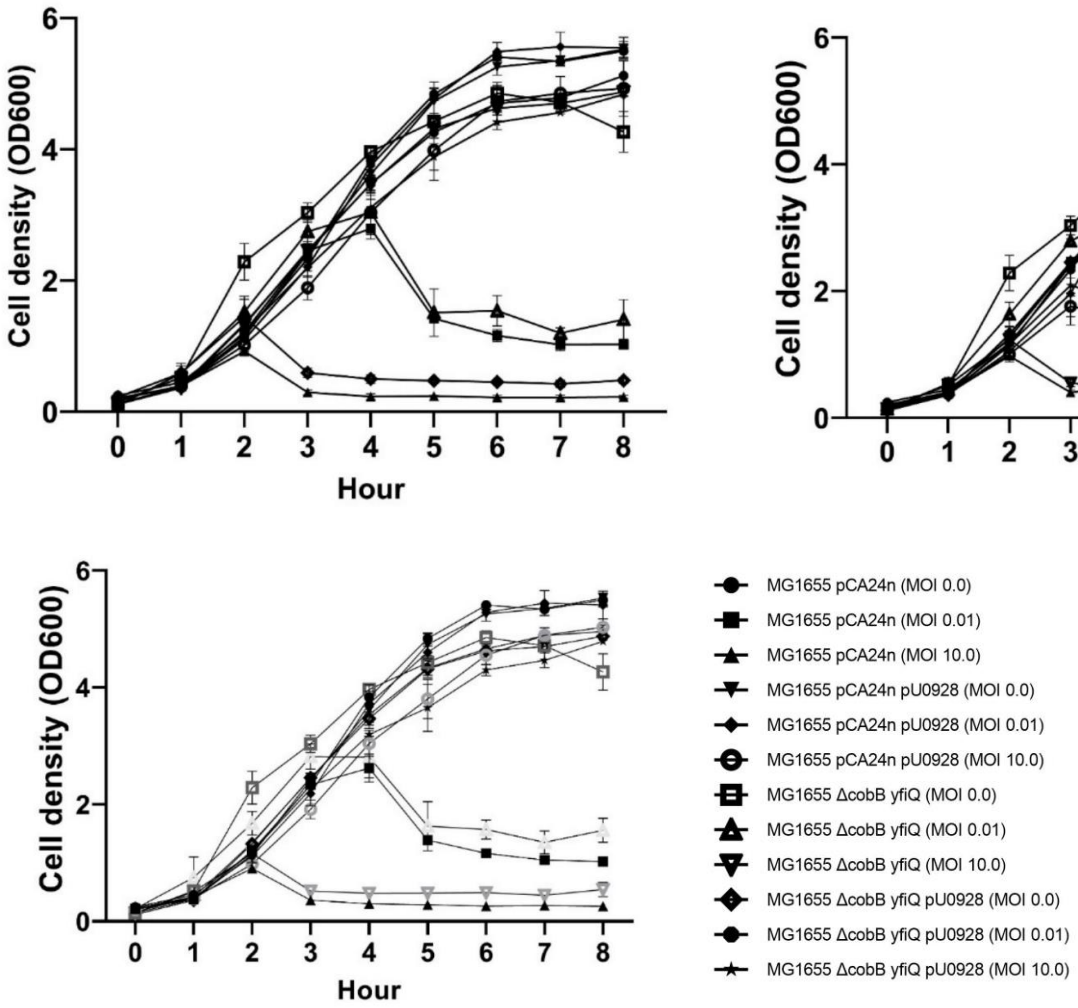


Figure 13. Growth Curve of K-12 Transconjugants with pU0928 Infected with Phage.

Constructs of *E. coli* K-12 were infected with (a) P1vir, (b) Greed, and (c) Lust at OD 0.04 (~1 hour) at three MOIs (0.0, 0.01, 10.0). When infected, constructs with no pU0928 had a severely decreased OD, while those carrying pU0928 exhibited a less pronounced decrease.

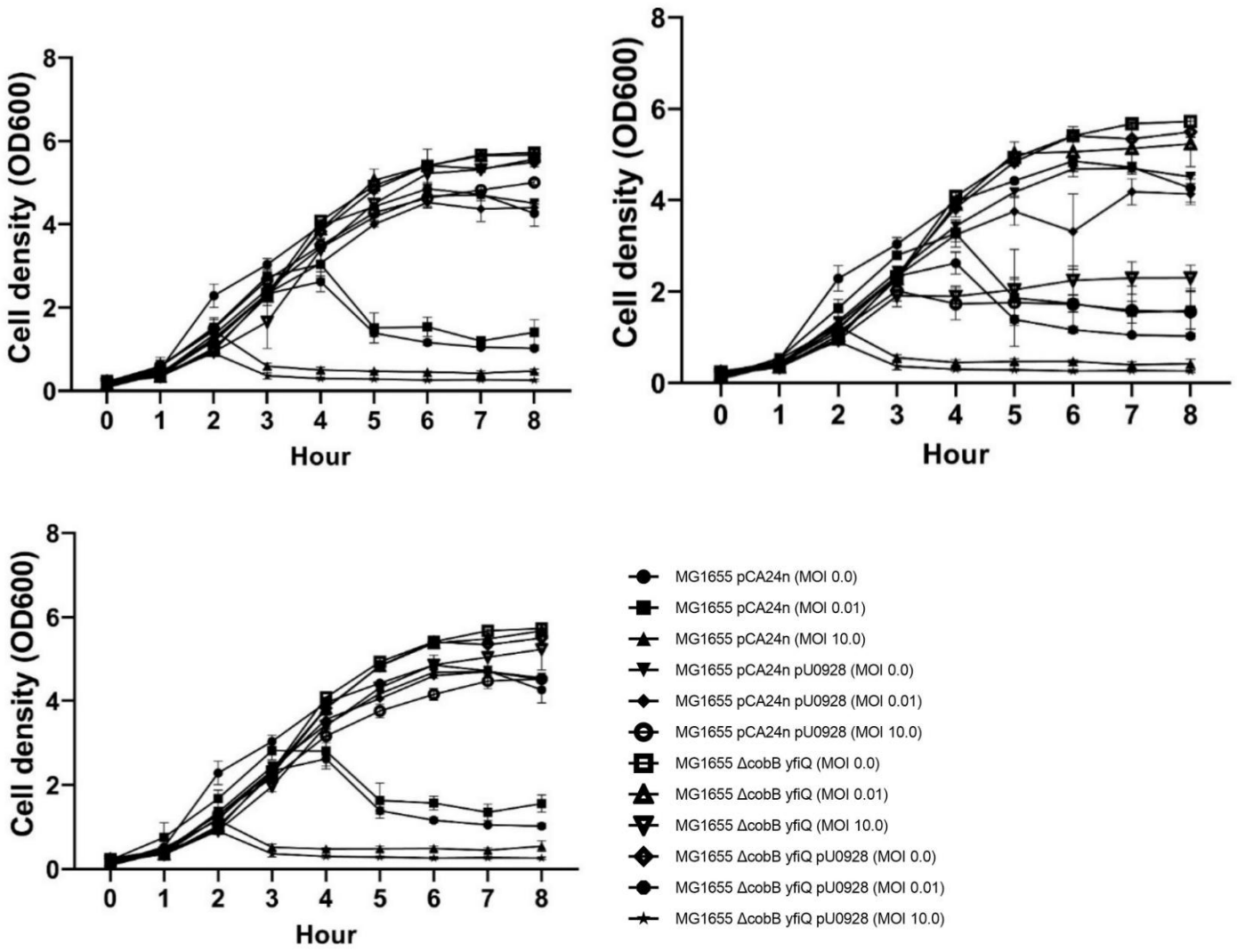


Figure 14. Growth Curve of K-12 Transconjugants with pU1284 Infected with Phage.
Constructs of *E. coli* K-12 were infected with (a) P1vir, (b) Greed, and (c) Lust at OD 0.04 (~1 hour) at three MOIs (0.0, 0.01, 10.0). When infected, constructs with no pU1284 had a severely decreased OD, while those carrying pU1284 exhibited a less pronounced decrease.

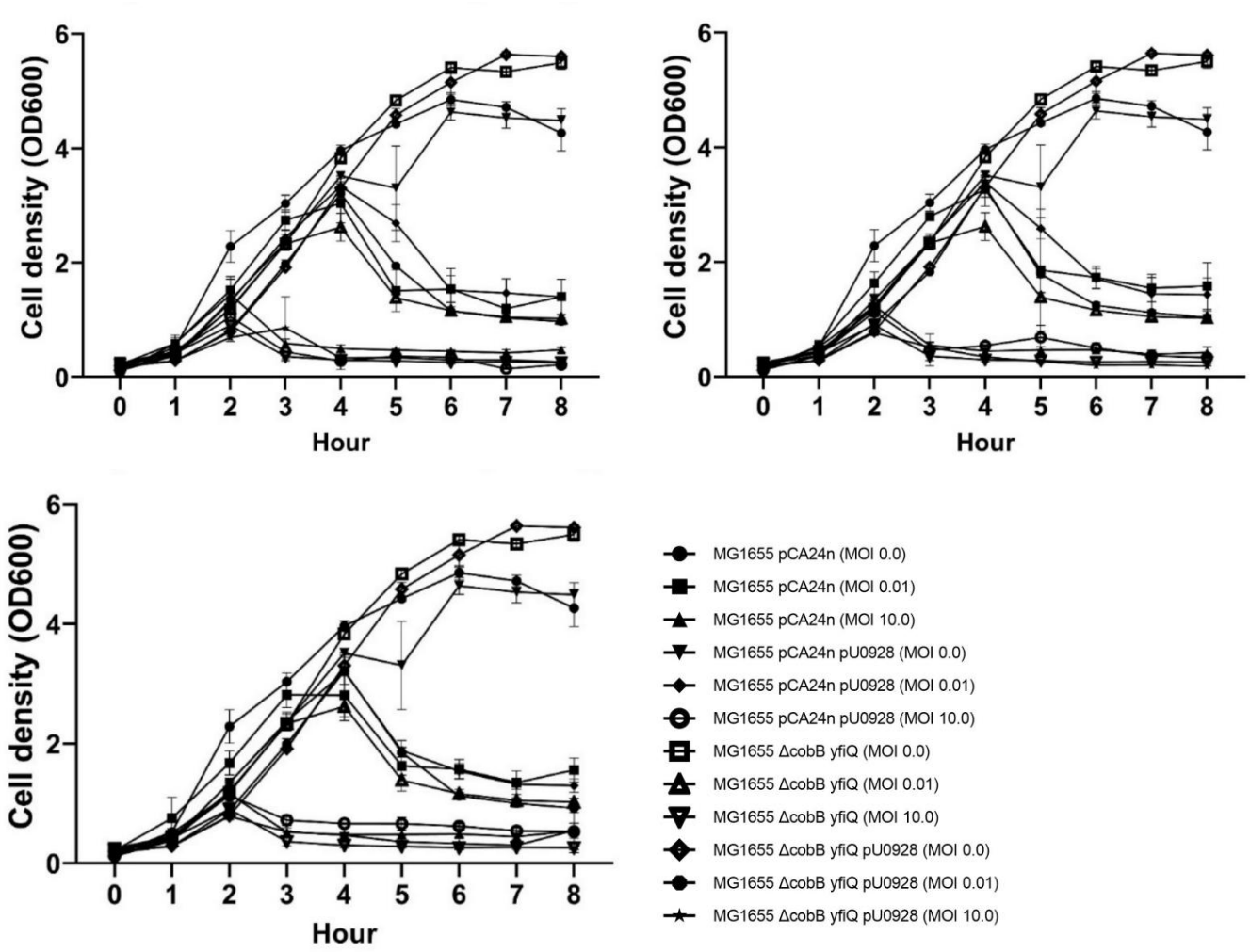


Figure 15. Growth Curve of K-12 Transconjugants with pU1223 Infected with Phage.
Constructs of *E. coli* K-12 were infected with (a) P1vir, (b) Greed, and (c) Lust at OD 0.04 (~1 hour) at three MOIs (0.0, 0.01, 10.0). Constructs exhibited a several decreased OD when infected even in the presence of pU1223.

Table 28. Growth of K-12 with pU0928 and pU1284 on Antibiotic Plates.

	K-12 AJW5116	UMB0928	K-12 AJW5116 pU0928	UMB1284	K-12 AJW5116 pU1284
Strain description	Plasmid recipient	Plasmid donor	Transconjugant	Plasmid donor	Transconjugant
LB	+	+	+	+	+
Tetracycline		+	+	+	+
Kanamycin				+	+
Ampicillin		+	+	+	+
Spectinomycin				+	+
Chloramphenicol	+		+		+

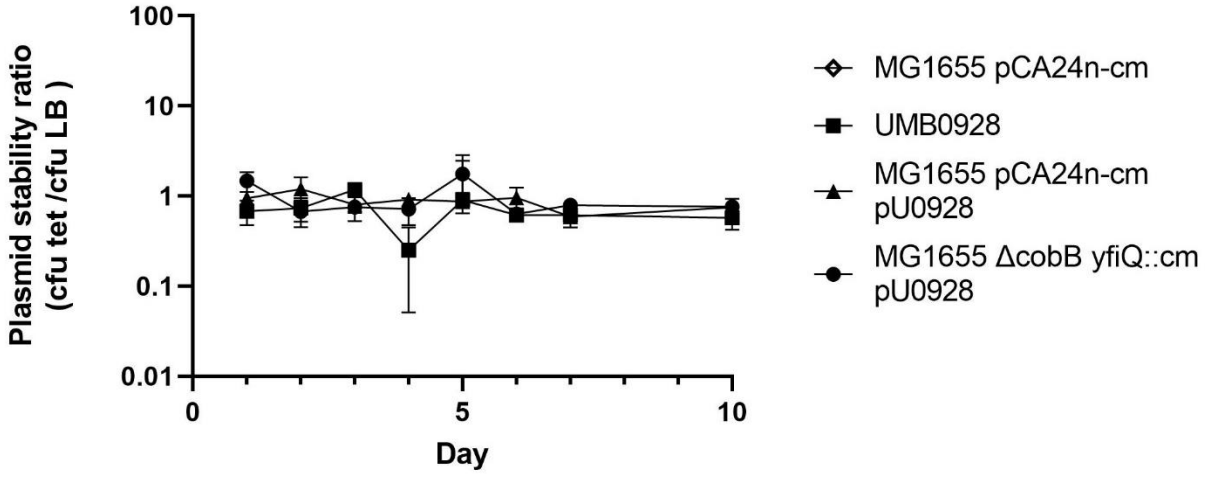


Figure 16. Stability of pU0928 in the Absence of Selection.

Urinary isolate UMB0928 and *E. coli* K-12 variants were grown in the absence of antibiotic selection for plasmid pU0928 for 10 days. Cultures were plated onto tetracycline (pU0928 selection marker) and LB plates daily. A plasmid stability ratio of 1 indicates plasmid retention, while a ratio close to 0 indicates loss of plasmid. The negative control MG1655 pCA24n-Cm without pU0928 did not grow on tetracycline plates.

Sequence Analysis of Anti-phage Plasmids

The plasmidic assemblies for the plasmid donors UMB0928, UMB1284, and UMB1223 were profiled for homology and gene content. The curated plasmid sequence from these assemblies was estimated to be ~100,000 bp and homologous to F plasmids in the NCBI database (Table 28). UMB1284 had plasmidic sequences (UMB1284_1, UMB1284_2) with homology to two distinct plasmids in the NCBI database. Only 127 ORFs (30.82%) in the plasmidic assemblies from UMB0928 and UMB1284 were annotated with a function, while 285 (69.17%) were not assigned a predicted function (i.e., hypothetical proteins).

Table 29. Sequence Overview of Plasmid Donors for K-12 Conjugation.

Predicted plasmid	Predicted size (bp)	BLAST hit	Plasmid replicon type	Size (bp)	GenBank	Query coverage	% sequence identity	E-value
UMB0928_1	107042	pCFS3313-2	Col156, incFIA, incFIB, incFII	111822	CP026941.2	98	99.17	0.0
UMB1223_1	144896	pDA33137-178	Col156, incFIA, incFIB, incFII	178078	CP029580.1	90	99.95	0.0
UMB1284_1	48172	p51008369SK1	incX	33826	CP029976.1	70	99.95	0.0
UMB1284_2	95109	p179-1	incFIA, incFII	122483	CP041560.1	75	100	0.0

We performed both plasmid and whole genome extractions of the *E. coli* K-12 transconjugants carrying either pU928, pU1223, and pU1284 and sequenced via Illumina short read sequencing. The plasmid extractions failed to adequately sequence (data not shown). However, the whole genome extractions resulted in sufficient plasmid genetic content for assembly (Table 29). Plasmid contigs were aligned to plasmid entries in the NCBI database via web BLAST to curate plasmid contigs and prune chromosomal contigs. Curated plasmid assemblies for pU0928, pU1223, and pU1284 were predicted to be ~100k bp (approximately the size of a typical F plasmid). The assemblies for pU0928, pU1223, and pU1284, and pU6721 had homology to *E. coli* plasmid sequences deposited in the NCBI database (Table 30). The genetic content sequenced from *E. coli* K-12 pU1091 only had chromosomal hits.

Table 30. Assembly and Annotation Overview of Urinary Plasmids from K-12 Transconjugants.

Urinary plasmid	Phage phenotype	Contigs	Bases	CDS
pU0928	Less permissive	16	100293	107
pU1284	Less permissive	11	130107	151
pU1223	Permissive	20	148520	173
pU1091	Permissive	11	10546	9
pU6721	Permissive	7	39608	54

The genetic content of the plasmids was analyzed for incompatibility, virulence, and antibiotic resistance genes (Table 31). pU0928, pU1223, and pU1284 had incompatibility genes like those predicted in their respective parent plasmid donor, primarily those associated with F plasmids. pU1091 did not have any predicted *rep* or *inc* genes, further supporting the hypothesis that plasmid content was not sequenced. pU6721 had the incompatibility gene *incX4*, which differs from the incompatibility genes predicted in UMB6721 (*incFIB*, *col*, *incB/O/K/*, *incFII*, *incX4*), in addition to having a different plasmid size than UMB6721. The two plasmids that confer an anti-phage phenotype, pU0928, and pU1284, were predicted to carry virulence genes. pU0928, pU1223, and pU1284 had multiple predicted antibiotic resistance genes, including against tetracycline, which was utilized for antibiotic selection for the *E. coli* K-12 transconjugants.

Table 31. BLAST Hits for Urinary plasmids Extracted and Sequenced from K-12 Transconjugants.

Host	Urinary plasmid sequenced	Urinary plasmid size (bases)	Plasmid BLAST hit	BLAST hit size (bases)	Host	Query coverage	E value	Per. Ident	Accession
K-12	pU0928	100293	pU15A_A	109023	*Escherichia coli*	45%	0	99.60%	CP035721.1
			pU15A_B	87354	*Escherichia coli*	39%	0	99.60%	CP035469.1
			p22C124-2	80146	*Escherichia coli*	32%	0	99.62%	LC501510.1
			p16C47-3	87921	*Escherichia coli*	32%	0	99.62%	LC501471.1
			p17C106-1	23836	*Escherichia coli*	21%	0	99.62%	LC501478.1
K-12	pU1284	130107	p2629-1	109813	*Escherichia coli*	52%	0	100.00%	CP041542.1
			p4_0.1	138672	*Escherichia coli*	72%	0	100.00%	CP023850.1
			pTO148	133139	*Escherichia coli*	73%	0	100.00%	LS992191.1
			RCS59_p	130559	*Escherichia coli*	64%	0	100.00%	LT985271.1
			pECAZ162_KPC	142829	*Escherichia coli*	61%	0	100.00%	CP019014.1
K-12	pU1223	148520	pNMBU-W13E19_01	122112	*Escherichia coli*	76%	0	99.96%	CP043407.1
			pRHBSTW-00322_2	124230	*Escherichia coli*	77%	0	99.77%	CP056599.1
			pB28	112178	*Escherichia coli*	51%	0	99.92%	MK295820.1
			pMSB1_3B-sc-2280406_2	160467	*Escherichia coli*	86%	0	99.80%	LR890300.1
			pRSB225	164550	uncultured bacterium	74%	0	99.95%	JX127248.1
K-12	pU1091	10546	Chromosome	4621656	*Escherichia coli*	63%	0	99.88%	CP011343.2
			Chromosome	4615343	*Escherichia coli*	63%	0	99.85%	CP043205.1
			Chromosome	4615322	*Escherichia coli*	63%	0	99.85%	CP043207.1
			Chromosome	4686434	*Escherichia coli*	65%	0	99.85%	CP060708.1
			Chromosome	4656756	*Escherichia coli*	63%	0	99.83%	CP043193.1
K-12	pU6721	39608	p010_C	32102	*Escherichia coli*	72%	0	99.67%	CP048333.1
			pSAM7	35341	*Escherichia coli*	72%	0	99.66%	JX981514.1
			pCFS3313-3	31764	*Escherichia coli*	72%	0	99.59%	CP026942.1
			pEcIncX4	31653	*Escherichia coli*	71%	0	99.59%	MT349420.1
			pVPS18EC0801-4	31764	*Escherichia coli*	72%	0	99.58%	CP063721.1

Table 32. Gene Analysis of Urinary Plasmids in K-12 Transconjugants.

Host	Urinary plasmid	Phage phenotype	Plasmid replicon	Virulence genes	Antibiotic resistance predicted
K-12	pU0928	Less permissive	IncFIB, Col156, IncQ1, IncI1-I	*cia, senB*	Streptomycin, sulfamethoxazole, trimethoprim, tetracycline
K-12	pU1284	Less permissive	IncFIA, IncFII, IncX4	*traT*	Ciprofloxacin, spectinomycin, trimethoprim, sulfamethoxazole, macrolide, tetracycline,
K-12	pU1223	Permissive	IncFIA, IncFIB, IncFII, Col156		Streptomycin, tetracycline, sulfamethoxazole, macrolide, ampicillin/amoxicillin/ceftriaxone/piperacillin
K-12	pU1091	Permissive	None		Aminoglycoside
K-12	pU6721	Permissive	IncX4		

pU0928 and pU1284 conferred an anti-phage phenotype to *E. coli* K-12, while pU1223 conferred as permissive a phenotype as the no plasmid control. Genes in pU0928 and pU1284 may explain the anti-phage mechanism. Thus, we mapped sequence reads to the respective curated plasmid sequence (Figure 17). No large gaps were observed in the plasmid sequence after mapping reads to the predicted plasmid sequence. The curated plasmid assemblies were annotated via Prokka and ORFs clustered via Usearch with a 0.8 amino acid sequence identity threshold. A known gene function was assigned to 39.05% of ORFs in pU0928, 40.82% of ORFs in pU1284, and 42.18% of ORFs in pU1223 (Table 32). The names of distinct ORFs were processed as a list for each plasmid (Table 33-35). The plasmids were noted to have plasmid replication and maintenance machinery, virulence genes, and the tetracycline resistance that was used for antibiotic selection. pU0928 and pU1284 were queried for the anti-phage genes listed in Table 10 (cover query and sequence identity >90%). From these, the only anti-phage genes present were *imm* in pU0928 and t*traT* in pU1284.

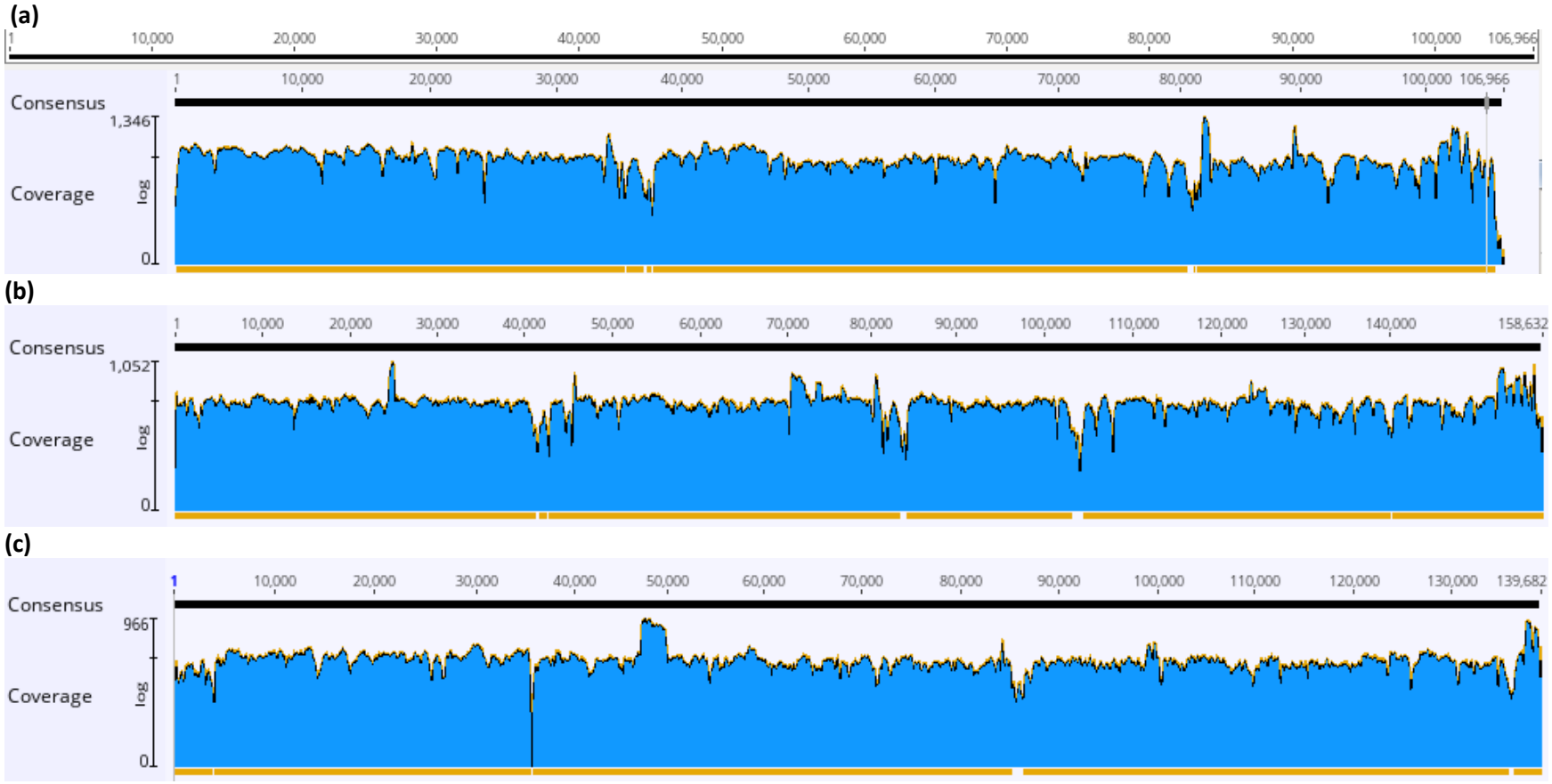


Figure 17. Read Coverage of Urinary F Plasmids in K-12 Transconjugants.

Sequence reads were mapped to curated plasmid assembly from (a) pU0928, (b) pU1284, and (c) pU1223. No large gaps are observed in the plasmid sequence predicted after sequence read assembly.

Table 33. Overview of ORFs Annotated in Urinary Plasmids from K-12 Transconjugants.

	Phage phenotype	Bases	ORFs with function	ORFs hypothetical	% with function
pU0928	Less permissive	100293	41	64	39.05
pU1284	Less permissive	130107	69	100	40.82
pU1223	Permissive	148520	62	85	42.18

Table 34. Unique ORFs in pU0928 Annotated with a Function.

34 kDa membrane antigen	Major pilus subunit operon regulatory protein PapB	RepFIB replication protein A
Aminoglycoside 3'-phosphotransferase	Na(+)-translocating NADH-quinone reductase subunit C	Replication initiation protein
Colicin-Ia	Nucleoid occlusion protein	Tetracycline repressor protein class B from transposon Tn10
Dihydropteroate synthase	Plasmid-derived single-stranded DNA-binding protein	Tetracycline resistance protein, class B
DNA-invertase hin	Plasmid segregation protein ParM	Thiol-disulfide oxidoreductase ResA
Glucose-1-phosphatase	Protein PsiB	Thiol-disulfide oxidoreductase ResA
Hemin receptor	Protein UmuC	Tn3 family transposase TnAs3
IS1 family transposase IS1R	putative protein	Transcription antitermination protein RfaH
IS21 family transposase IS100kyp	putative protein YncE	Transposon Tn10 TetC protein
IS4 family transposase ISVsa5	putative TonB-dependent receptor	Transposon Tn10 TetD protein
IS66 family transposase ISEc22	Putative transposase InsK for insertion sequence element IS150	Tyrosine recombinase XerC
IS66 family transposase ISEc22	Regulatory protein RepA	Tyrosine recombinase XerC
IS66 family transposase ISEc8		Tyrosine recombinase XerC
IS6 family transposase IS26		
ISNCY family transposase ISSen7		
Lipoprotein-releasing system ATP-binding protein LolD		

Note: Unique ORF is based off a 0.8 amino acid sequence identity threshold. Entries with the same name on the list fall below this threshold and are thus distinct ORFs despite being annotated with the same name.

Table 35. Unique ORFs in pU1284 Annotated with a Function.

34 kDa membrane antigen	IS6 family transposase IS26	Protein TraJ
Amino-acid permease RocC	IS6 family transposase IS6100	Protein TraQ
Aminoglycoside 3'-phosphotransferase	IS91 family transposase ISSbo1	Protein TraV
Antirestriction protein KlcA	ISL3 family transposase ISStma11	putative protein
Antitoxin CcdA	Lipoprotein-releasing system ATP-binding protein LolD	putative protein YncE
Antitoxin PemI	Lipoprotein YlpA	putative signaling protein
Beta-lactamase TEM	Mating pair stabilization protein TraN	putative TonB-dependent receptor
Coupling protein TraD	Mercuric reductase	Putative transposase InsK for insertion sequence element IS150
Dihydropteroate synthase	Mercuric transport protein MerC	Relaxosome protein TraM
DNA adenine methyltransferase YhdJ	Mercuric transport protein MerT	Relaxosome protein TraY
DNA-cytosine methyltransferase	Mercuric transport protein periplasmic component	RepFIB replication protein A
Endonuclease YhcR	Methyl-accepting chemotaxis protein McpQ	Replication initiation protein
Endoribonuclease PemK	Multidrug efflux pump Tap	Tetracycline repressor protein class B from transposon Tn10
Fertility inhibition protein	Multifunctional conjugation protein TraI	Tetracycline resistance protein, class B
Glucose-1-phosphatase	Na(+)-translocating NADH-quinone reductase subunit C	Thiol-disulfide oxidoreductase ResA
Hemin receptor	Nucleoid occlusion protein	Thiol-disulfide oxidoreductase ResA
Homocysteine S-methyltransferase	Pilin	Tn3 family transposase ISPa38
IS110 family transposase IS5075	Plasmid-derived single-stranded DNA-binding protein	Toxin CcdB
IS21 family transposase IS100kyp	Protein PsiB	Transposon Tn10 TetC protein
IS3 family transposase ISLad1	Protein SopB	Transposon Tn10 TetD protein
IS4 family transposase ISVsa5	Protein TraC	Type-2 restriction enzyme EcoRII
IS5 family transposase IS5		Tyrosine recombinase XerC
IS66 family transposase ISEc22		Tyrosine recombinase XerC
IS66 family transposase ISEc22		
IS66 family transposase ISEc8		

Table 36. Unique ORFs in pU1223 Annotated with a Function.

2-keto-3-deoxygluconate permease	Hemolysin expression-modulating protein Hha	sn-glycerol-3-phosphate transport system permease protein UgpA
34 kDa membrane antigen	IS1 family transposase IS1R	Streptomycin 3''-adenylyltransferase
3',5'-cyclic adenosine monophosphate phosphodiesterase CpdA	IS3 family transposase IS629	Tetracycline repressor protein class A from transposon 1721
Aldehyde reductase Ahr	IS3 family transposase IS629	Tetracycline resistance protein, class C
Aminoglycoside N(6')-acetyltransferase type 1	IS66 family transposase ISEc23	Thiol-disulfide oxidoreductase ResA
Antirestriction protein KlcA	IS6 family transposase IS26	Tn3 family transposase
Antitoxin CcdA	Lipoprotein-releasing system ATP-binding protein LolD	Tn3 family transposase Tn2
Antitoxin PemI	Lipoprotein YlpA	Tn3 family transposase TnAs1
Beta-lactamase CTX-M-1	Modulating protein YmoA	Toxin CcdB
Beta-lactamase OXA-1	Multidrug efflux pump Tap	Transcriptional regulator
Chaperone protein DnaJ	Multidrug transporter EmrE	Transposase for transposon Tn5
Chloramphenicol acetyltransferase	Na(+)-translocating NADH-quinone reductase subunit C	tRNA(fMet)-specific endonuclease VapC
Coupling protein TraD	Proteasome-associated ATPase	Type IV secretion system protein VirB11
Diacetylchitobiose uptake system permease protein NgcG	Protein SopB	Type IV secretion system protein virB4
Dihydropteroate synthase	putative 2-dehydro-3-deoxy-D-pentonate aldolase YjhH	Type IV secretion system protein virB8
DNA adenine methyltransferase YhdJ	putative chromate transport protein	Type IV secretion system protein virB9
DNA-binding protein H-NS	putative HTH-type transcriptional regulator RhmR	Tyrosine recombinase XerC
DNA topoisomerase 3	Replication initiation protein	Tyrosine recombinase XerC
Endonuclease YhcR	Serine recombinase PinR	Tyrosine recombinase XerC
Endonuclease YhcR	sn-glycerol-3-phosphate-binding periplasmic protein UgpB	
Endoribonuclease PemK	sn-glycerol-3-phosphate import ATP-binding protein UgpC	
Glucose-1-phosphatase		

All ORFs from pU0928, pU1223, and pU1284 were reviewed to identify ORFs exclusive to or shared by these three plasmids (Table 36). Three ORFs are found only in pU0928 and pU1284 but not in pU1223. The predicted functions of these three ORFs were respectively that of a phage integrase (227 amino acids), a dihydrofolate reductase (199 amino acids), and an ORF with an EAL cyclic di-GMP phosphodiesterase domain (116 amino acids) (Table 37) (Figures 18-20). Phage integrases are utilized by

phage to integrate into the host genome. Dihydrofolate reductase is used by some phages for phage tail stability and phage particle assembly. For the third ORF, EAL cyclic di-GMP phosphodiesterase domains are involved in signaling.

Table 37. ORFs Shared in the Anti-phage Plasmids pU0928 and pU1284.

ORF name	pU0928	pU1223	pU1284
	Less phage permissive	Phage permissive	Less phage Permissive
Tyrosine recombinase XerC	+		+
pU1284_1.fasta_OJDLIIBG_00125 hypothetical protein	+		+
pU1284_1.fasta_OJDLIIBG_00127 hypothetical protein	+		+
Putative transposase InsK for insertion sequence element IS150	+	+	+
Tyrosine recombinase XerC	+	+	+
IS3 family transposase IS629	+	+	+
Na(+)-translocating NADH-quinone reductase subunit C	+	+	+
34 kDa membrane antigen	+	+	+
Lipoprotein-releasing system ATP-binding protein LolD	+	+	+
Thiol-disulfide oxidoreductase ResA	+	+	+
Glucose-1-phosphatase	+	+	+
IS6 family transposase IS26	+	+	+
IS1 family transposase IS1R	+	+	+
pU1223.fasta_GJKJIFEI_00100 hypothetical protein	+	+	+
pU1223.fasta_GJKJIFEI_00110 hypothetical protein	+	+	+
pU1223.fasta_GJKJIFEI_00111 hypothetical protein	+	+	+
pU1284_1.fasta_OJDLIIBG_00097 hypothetical protein	+	+	+
pU1284_1.fasta_OJDLIIBG_00131 hypothetical protein	+	+	+
pU1284_1.fasta_OJDLIIBG_00132 hypothetical protein	+	+	+
pU1284_1.fasta_OJDLIIBG_00133 hypothetical protein	+	+	+
pU1284_1.fasta_OJDLIIBG_00136 hypothetical protein	+	+	+

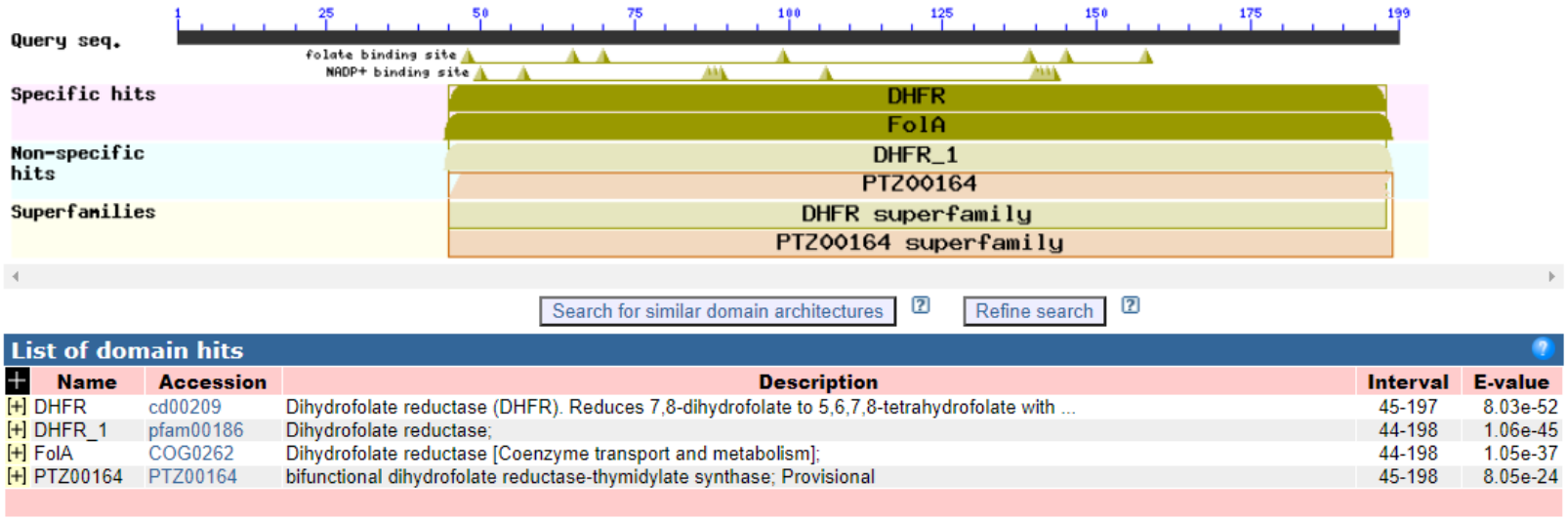
Note: Hypothetical ORFs are given a temporary name based off that run

Table 38. BLAST Result for the Three ORFs Shared by the Anti-phage Plasmids pU0928 and pU1284.

ORF name[1]	Predicted function[2]	Description	Length (aa)	Qcover	Eval	pident	Accession
pU1284_1.fasta_OJDLIIBG_00125 hypothetical protein	Dihydrofolate reductase	Dihydrofolate reductase type 1 [*Escherichia coli*]	199	100	6.00E-144	100	AKN35619.1
Size: 199 amino acids		dihydrofolate reductase type VII [*Escherichia coli*]	224	100	7.00E-144	100	ACQ42056.1
		dihydrofolate reductase [*Escherichia coli*]	210	100	7.00E-144	100	AIG72712.1
		dihydrofolate reductase [*Escherichia coli*]	232	100	9.00E-144	100	AQS26682.1
		dihydrofolate reductase [*Klebsiella pneumoniae*]	233	100	1.00E-143	100	ADD63540.1
Tyrosine recombinase XerC	Phage-integrase	integrase/recombinase [*Klebsiella* pneumoniae]	339	100	0	100	KMG57351.1
Size: 337 amino acids		MULTISPECIES: class 1 integron integrase IntI1 [Bacteria]	337	100	0	100	WP_000845048.1
		class 1 integron integrase IntI1 [*Escherichia coli*]	337	100	0	100	WP_097473620.1
		integron integrase [*Comamonas thiooxydans*]	337	100	0	100	WP_197577570.1
		integron integrase IntI1 [*Klebsiella pneumoniae*]	390	100	0	100	SAY12656.1
pU1284_1.fasta_OJDLIIBG_00127 hypothetical protein	EAL cyclic di-GMP phosphodiesterase domain	hypothetical protein [*Escherichia coli*]	150	100	6.00E-81	100	ABG49198.1
Size: 116 amino acids		Select seq gb|KJO53739.1|	134	100	6.00E-81	100	KJO53739.1
		Select seq gb|KTG85147.1|	141	100	7.00E-81	100	KTG85147.1
		Select seq ref|WP_077776315.1|	198	100	8.00E-81	100	WP_077776315.1
		Select seq ref|WP_077816978.1|	198	100	8.00E-81	100	WP_077816978.1

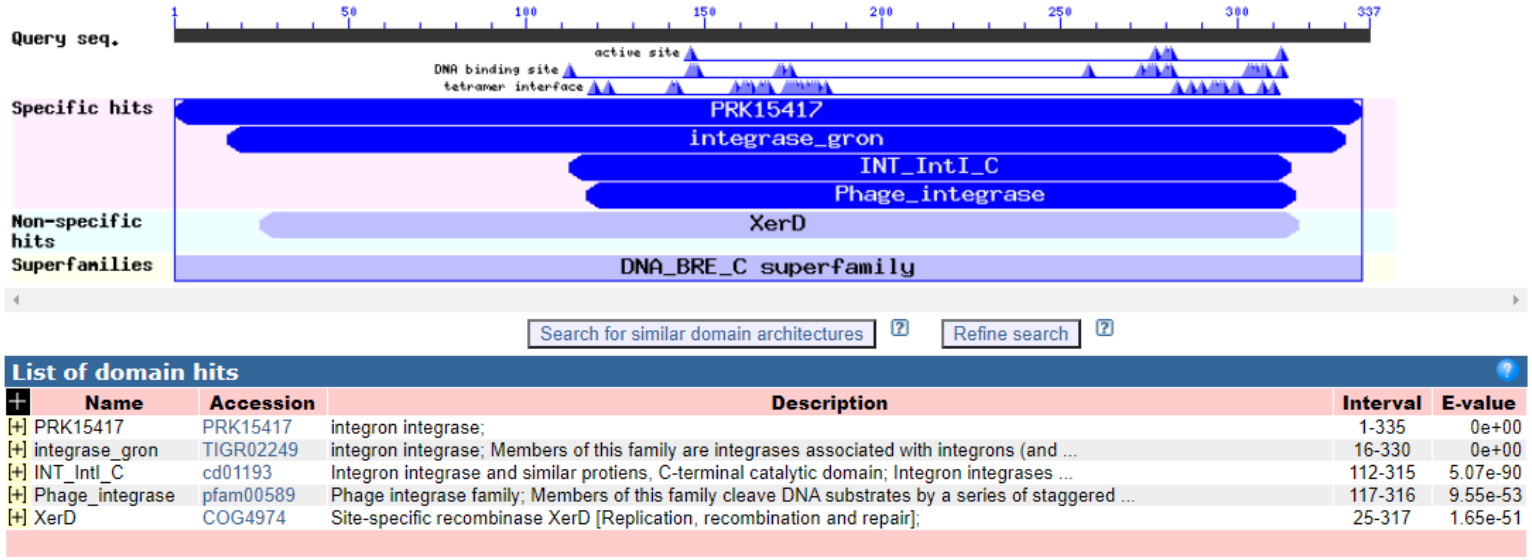
[1]Temporary name given by Prokka, in case of a hypothetical ORF it is based off the contig name

[2]Name given off predicted function and used hereafter



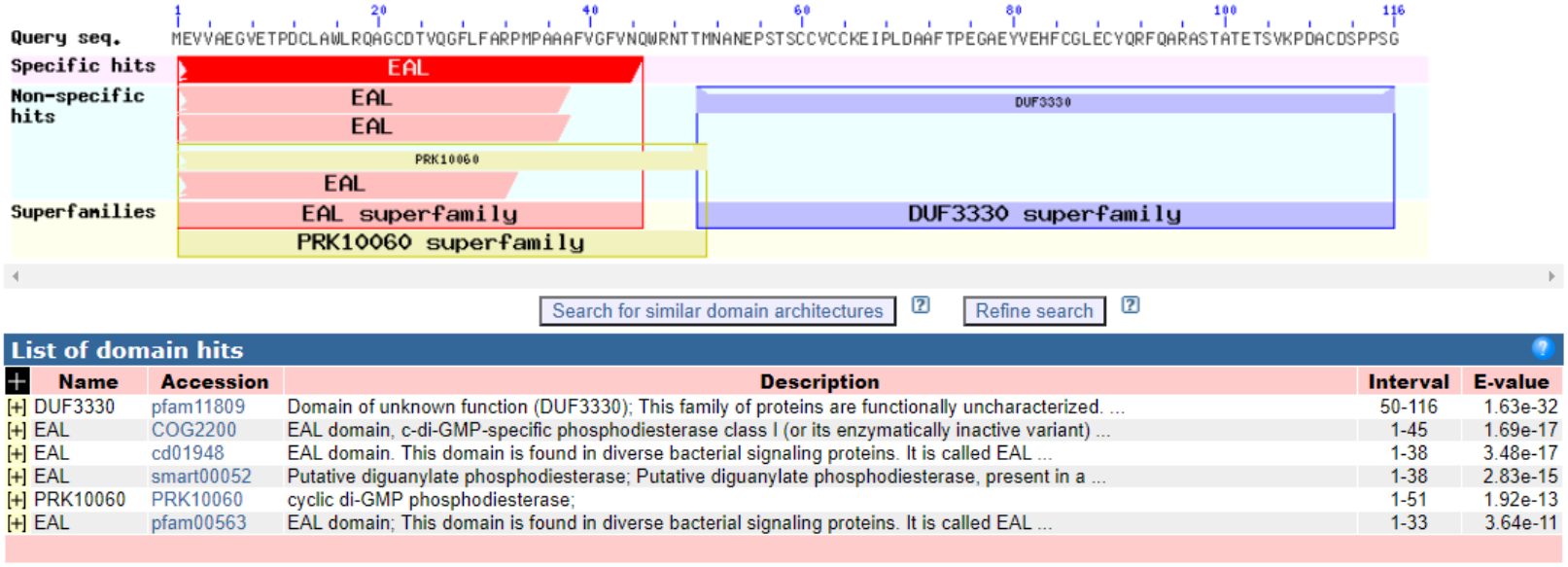
	Name	Accession	Description	Interval	E-value
[+]	DHFR	cd00209	Dihydrofolate reductase (DHFR). Reduces 7,8-dihydrofolate to 5,6,7,8-tetrahydrofolate with ...	45-197	8.03e-52
[+]	DHFR_1	pfam00186	Dihydrofolate reductase;	44-198	1.06e-45
[+]	FolA	COG0262	Dihydrofolate reductase [Coenzyme transport and metabolism];	44-198	1.05e-37
[+]	PTZ00164	PTZ00164	bifunctional dihydrofolate reductase-thymidylate synthase; Provisional	45-198	8.05e-24

Figure 18. pU1284_1.fasta_OJDLIIBG_00125 Has Protein Domains of Dihydrofolate Reductase.



Name	Accession	Description	Interval	E-value
[+] PRK15417	PRK15417	integron integrase;	1-335	0e+00
[+] integrase_gron	TIGR02249	integron integrase; Members of this family are integrases associated with integrons (and ...	16-330	0e+00
[+] INT_IntI_C	cd01193	Integron integrase and similar protiens, C-terminal catalytic domain; Integron integrases ...	112-315	5.07e-90
[+] Phage_integrase	pfam00589	Phage integrase family; Members of this family cleave DNA substrates by a series of staggered ...	117-316	9.55e-53
[+] XerD	COG4974	Site-specific recombinase XerD [Replication, recombination and repair];	25-317	1.65e-51

Figure 19. Tyrosine Recombinase XerC Has Protein Domains of a Phage Integrase.



List of domain hits

Name	Accession	Description	Interval	E-value
[+] DUF3330	pfam11809	Domain of unknown function (DUF3330); This family of proteins are functionally uncharacterized. ...	50-116	1.63e-32
[+] EAL	COG2200	EAL domain, c-di-GMP-specific phosphodiesterase class I (or its enzymatically inactive variant) ...	1-45	1.69e-17
[+] EAL	cd01948	EAL domain. This domain is found in diverse bacterial signaling proteins. It is called EAL ...	1-38	3.48e-17
[+] EAL	smart00052	Putative diguanylate phosphodiesterase; Putative diguanylate phosphodiesterase, present in a ...	1-38	2.83e-15
[+] PRK10060	PRK10060	cyclic di-GMP phosphodiesterase;	1-51	1.92e-13
[+] EAL	pfam00563	EAL domain; This domain is found in diverse bacterial signaling proteins. It is called EAL ...	1-33	3.64e-11

Figure 20. pU1284_1.fasta_OJDLIIBG_00127 Has Protein Domains of a Signaling Protein

Although pU0928 and pU1284 only share 14% of their plasmid sequences, the amino acid sequences of these three ORFs were conserved in both plasmids (Table 38). The phage integrase amino acid sequence from pU0928 and pU1284 had a 99% shared sequence identity, with a difference of two amino acids (Proline to Leucine at position 228, Valine to Phenylalanine at position 232). The amino acid sequence in the dihydrofolate reductase and EAL cyclic di-GMP phosphodiesterase domain were identical in pU0928 and pU1284. All urinary *E. coli* plasmid assemblies were scanned for the presence of these three ORFs, with the threshold for presence set at cover query >90% and sequence identity >90%. The phage-integrase ORF was present in 16 urinary *E. coli* plasmidic assemblies, the dihydrofolate reductase ORF was present in 13, and the EAL cyclic di-GMP phosphodiesterase domain ORF was in 10 (Table 39). The genes for the dihydrofolate reductase and the EAL cyclic di-GMP phosphodiesterase domain ORF were not always present together, but presence of either of these two ORFs always included presence of the phage-integrase ORF. All the urinary *E. coli* isolates with hits to these ORFs were predicted to have an F plasmid and all but three had no evidence of infection when spotted with phage. None of these three ORFs were identified in the plasmidic assembly from UMB1284 but were present in its WGS, potentially due to sequence loss during plasmidic assembly or chromosomal contig pruning in the former. The three ORFs were identified in UMB0928.

To assess if all the plasmids that have these three ORFs are similar, the nucleotide sequence of all urinary *E. coli* plasmidic assemblies containing these three ORFs were compared for sequence homology (Figure 21). Plasmid sequences from pU0928, pU1284, and pU1223 clustered with their respective plasmid donor isolate. Some plasmids clustered together, indicating a similar plasmid type in that subgroup, but the plasmid collection in its entirety did not have high homology. Despite low overall plasmid homology, the amino acid sequence for the three ORFs for the phage-integrase, dihydrofolate reductase, and EAL cyclic di-GMP phosphodiesterase domain ORF were conserved in these plasmids.

Table 39. Comparison of Three Shared ORFs in Anti-phage Plasmids pU0928 and pU1284.

ORF name	Predicted function	Identities	Positives	Gaps
Tyrosine recombinase XerC	Phage-integrase	335/337(99%)	335/337(99%)	0/337(0%)
pU1284_1.fasta_OJDLIIBG_00125 hypothetical protein	Dihydrofolate reductase	199/199(100%)	199/199(100%)	0/199(0%)
pU1284_1.fasta_OJDLIIBG_00127 hypothetical protein	EAL cyclic di-GMP phosphodiesterase domain	116/116(100%)	116/116(100%)	0/116(0%)

Table 40. Presence of ORFs Shared by pU0928 and pU1284 in Other Urinary *E. coli* Plasmids.

Isolate	Phage spot	Inc group	Phage-integrase	Dihydrofolate reductase	EAL cyclic di-GMP phosphodiesterase domain
UMB0906	No	IncFII	+	+	
UMB0928	No	IncFI	+	+	+
UMB0931	Greed	IncFII	+	+	+
UMB0933	No	IncFII	+	+	+
UMB0934	No	IncFII	+	+	
UMB0949	No	IncFII	+	+	+
UMB1091	No	IncFII	+	+	
UMB1160	P1vir	IncFII	+	+	+
UMB1193	No	IncFII	+		+
UMB1223	No	IncFII	+	+	
UMB1229	No	IncFII	+		+
UMB1284	No	IncFII			
UMB1285	No	IncFII	+	+	+
UMB3641	No	IncFII	+		
UMB5924	P1vir	IncFI	+	+	
UMB5978	No	IncFI	+	+	+
UMB6653	No	IncFII	+	+	+
UMB6721	No	IncFII	+	+	+

Note: Presence indicates a cover query and sequence identity of over 90%.

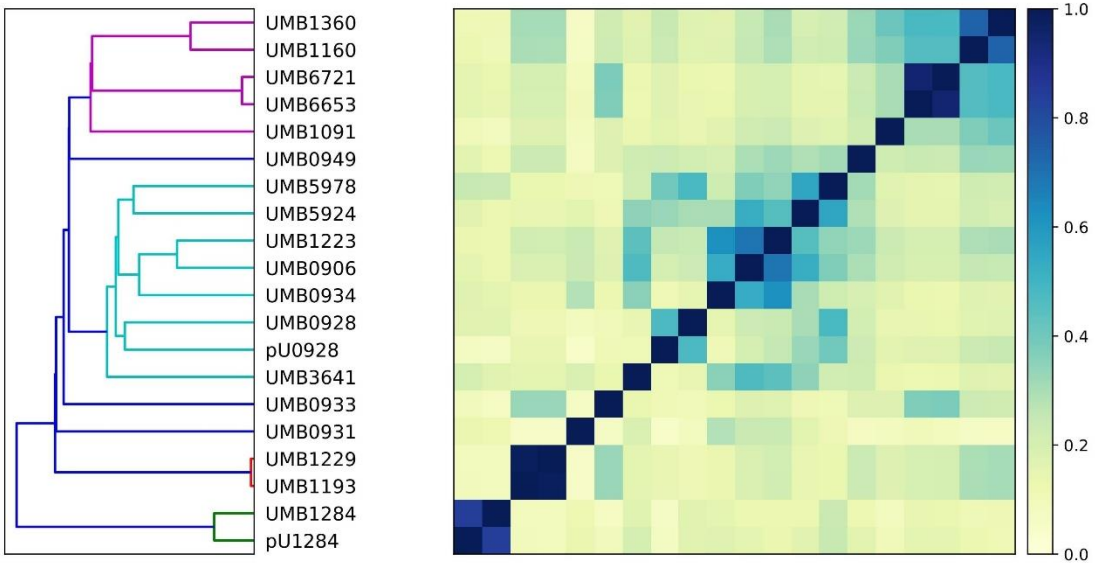


Figure 21. Homology of Urinary *E. coli* Plasmids with ORFs Shared by pU0928 and pU1284.

The plasmid sequence was compared for all urinary *E. coli* plasmids that contain the three ORFs shared by pU0928 and pU1284 (phage integrase, dihydrofolate reductase, EAL domain gene) present at >90% sequence identity and >90% query coverage. pU0928, pU1223, and pU1283 cluster with their respective plasmid donor parent. Multiple clusters were identified but, overall, the sequence homology for all plasmids as a group was low despite the three ORFs of interest being highly conserved. Black denotes 100% sequence homology and white is 0% sequence homology.

Phage-like Genetic Content in pU0928 and pU1284

A phylogenetic tree of phage integrase sequences was made to assess the relationship of the urinary *E. coli* phage integrase to phage integrases in the NCBI database. Phage integrase amino acid sequences were obtained via web BLAST hits from the pU1284 phage integrase, random searches for "phage-integrase" in the NCBI gene database (either plasmid, phage, or whole-genome), and homologs of the Lambda phage integrase (Figure 22). The phage-integrase from pU0928 clustered with phage-integrases of phage origin. The phage-integrase from pU1284 clustered with phage integrases from whole-genome sequences of bacteria (which could be of either chromosomal, plasmid, or phage origin). Because the phage integrase could originate from a prophage in the plasmid, the sequence of pU0928, pU1223, and pU1284 was scanned for phage content via PHAST and PHASTER. At least one phage-like gene was predicted in each plasmid, with varying degrees of phage sequence completeness (Table 40).

Phage sequences obtained from PHAST and PHASTER were compared to prune similar phage hits in the same plasmid (data not shown). Sequence homology comparison indicated four distinct phage-like sequences present in pU0928, three in pU1223, and one in pU1284 (Figure 23). pU0928 and pU1223 shared a predicted phage sequence (pU0928_phaster_1 and pU1223_phaster_1). Aside from some ORFs in common, most of the predicted phage sequences had low sequence homology as indicated by the light-yellow color of the heatmap. PHASTER generated phage maps with the phage-like genes annotated in these sequences (Figure 24-26). Predicted phage genes in pU0928 phage-like sequences were transposases, phage tail shafts, plate proteins, and coat proteins (Figure 24). Predicted phage genes in pU1223 in phage-like sequences were a transposase and other phage-like genes with unknown function (Figure 25). Predicted phage genes in the pU1284 phage-like sequence were tail shaft, transposase, and phage-like genes with unknown function (Figure 26).

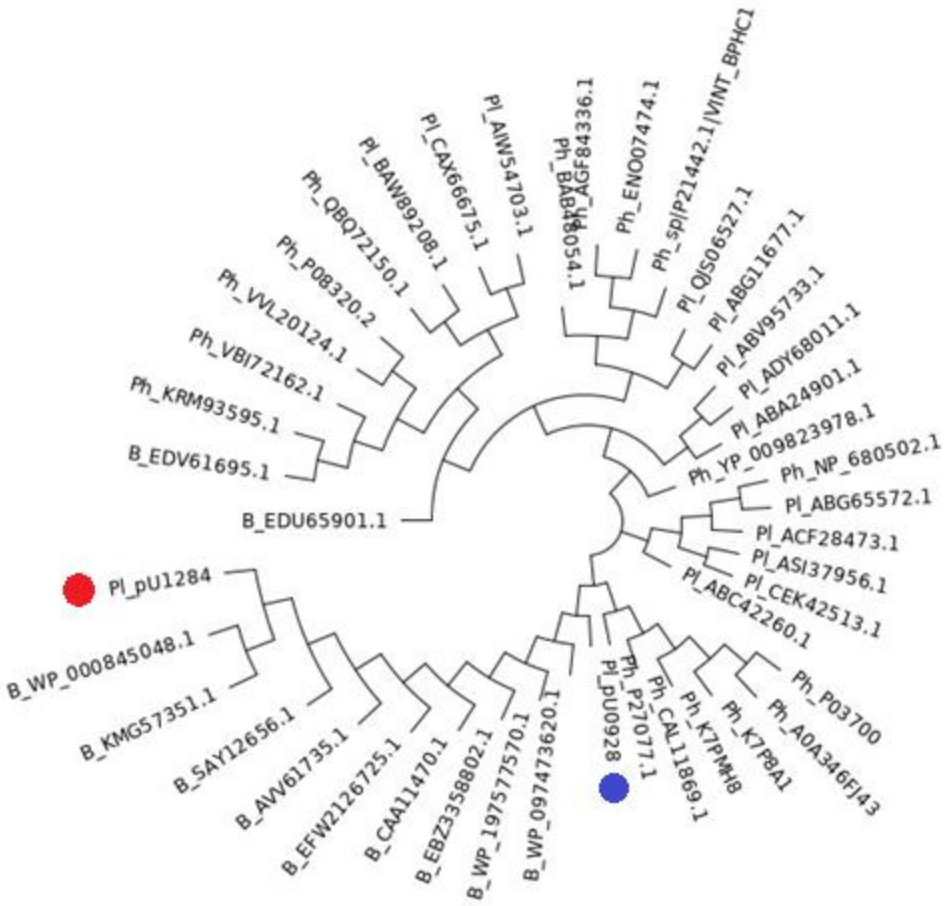


Figure 22. Phage Integrase Genes in Bacteria WGS, Plasmid, and Phages. The sequence for pU1284 has a red dot and the sequence for pU0928 has a blue dot. Phage integrase sequences of bacteria whole genome origin (B_) were EDU65901.1 (*E. coli*), EDV61695.1 (*E. coli*), KMG57351.1 (*K. pneumoniae*), WP_000845048.1 (Multispecies), WP_097473620.1 (*E. coli*), WP_197577570.1 (*Comamonas thiooxydans*), SAY12656.1 (*K. pneumoniae*), EBZ3358802.1 (*Salmonella*), CAA11470.1 (*Pseudomonas aeruginosa*), EFW2126725.1 (*Shigella oydii*), AVV61735.1 (*Serratia proteamaculans*). Phage integrase sequences of phage origin (Ph_) were P03700 (Lambda), A0A346FJ43 (*Enterobacteria phage*), KRM93595.1 (*Lactobacillus senioris* phage), K7P8A1 (*Escherichia* phage), K7PMH8 (*Escherichia* phage), BAB48054.1 (*Mesorhizobium japonicum* phage), AGF84336.1 (*Salmonella enterica* phage), ENO07474.1 (*E. coli* phage), VBI72162.1 (*Burkholderia pseudomallei*), YP_009823978.1 (*Erwinia* phage), P21442.1 (VINT_BPHC1 phage), NP_680502.1 (*Lactobacillus* phage A2), VVL20124.1 (*Vibrio* phage), P08320.2 (VINT_BPP4), P27077.1 (VINT_BPP21), QBQ72150.1 (*Serratia* phage Parlo), CAL11869.1 (*Yersinia enterocolitica* phage). Phage integrase sequences of plasmid origin (Pl_) were pU0928 (*E. coli*), pU1284 (*E. coli*), BAW89208.1 (*E. coli*), ABC42260.1 (*E. coli*), ASI37956.1 (*E. coli*), CAX66675.1 (*Lactobacillus johnsonii*), QJS06527.1 (*Arthrobacter* sp.), ABG11677.1 (*Mycobacterium* sp. MCS), ABV95733.1 (*Dinoroseobacter shibae*), CEK42513.1 (*Pseudomonas* fluorescen), ADY68011.1 (*Agrobacterium* sp), AIW54703.1 (*Clostridium botulinum*), ACF28473.1 (*Azospirillum baldaniorum*), ABG65572.1 *(Chelativorans* sp. BNC1), ABA24901.1 (*Trichormus* variabilis). Phage integrases from pU0928 and pU1284 have a two amino acid difference (P to L at position 228, V to F at position 232).

Table 41. Phage Hits in pU0928, pU1223, and pU12284 via PHAST and PHASTER.

Isolate	Identified	Region	Bases	Score	# ORFs	Most Common Phage
pU0928	PHASTER	1	6Kb	90	10	PHAGE_Escher_SH2026Stx1_NC_049919(3)
pU0928	PHASTER	2	6.4Kb	10	10	PHAGE_Escher_503458_NC_049341(3)
pU0928	PHASTER	3	6.5Kb	40	15	PHAGE_Escher_RCS47_NC_042128(2)
pU0928	PHAST	1	8.1Kb	110	11	PHAGE_Stx2_c_1717_NC_011357
pU0928	PHAST	2	7.4Kb	50	12	PHAGE_Mycoba_Xeno_NC_031243
pU0928	PHAST	3	12.2Kb	30	24	PHAGE_Gordon_Bowser_NC_030930
pU0928	PHAST	4	29Kb	40	9	PHAGE_Salmon_SJ46_NC_031129
pU1223	PHASTER	1	6.1Kb	90	10	PHAGE_Stx2_c_1717_NC_011357(3)
pU1223	PHAST	1	10.1Kb	100	16	PHAGE_Stx2_c_1717_NC_011357
pU1223	PHAST	2	7.7Kb	50	10	PHAGE_Burkho_KL3_NC_015266
pU1223	PHAST	3	6.5Kb	100	12	PHAGE_Entero_BP_4795_NC_004813
pU1284	PHASTER	1	4.4Kb	20	9	PHAGE_Stx2_c_Stx2a_WGPS9_NC_049923(2)
pU1284	PHAST	1	36.7Kb	100	20	PHAGE_Erwini_vB_EamM_Caitlin_NC_031120

Note: Score of >90 is predicted to be an intact phage

The ORFs of these predicted phages were annotated with Prokka to assess if phage-like ORFs were shared between pU0928 and pU1284. Four ORFs were shared in pU0928_PHAST_1 and pU1284_PHAST_1, including two transposases, dihydrofolate reductase, and a phage-integrase (Table 41). The dihydrofolate reductase and phage-integrase ORFs were the same ones identified as shared by the plasmids pU0928 and pU1284 (Table 36). The urinary *E. coli* plasmids previously identified to contain this phage integrase and dihydrofolate reductase ORF (Table 39) were scanned for phage content via PHASTER. Except for UMB3641, all urinary plasmids were predicted to contain at least one phage-like sequence, with varying degrees of completeness (Table 42).

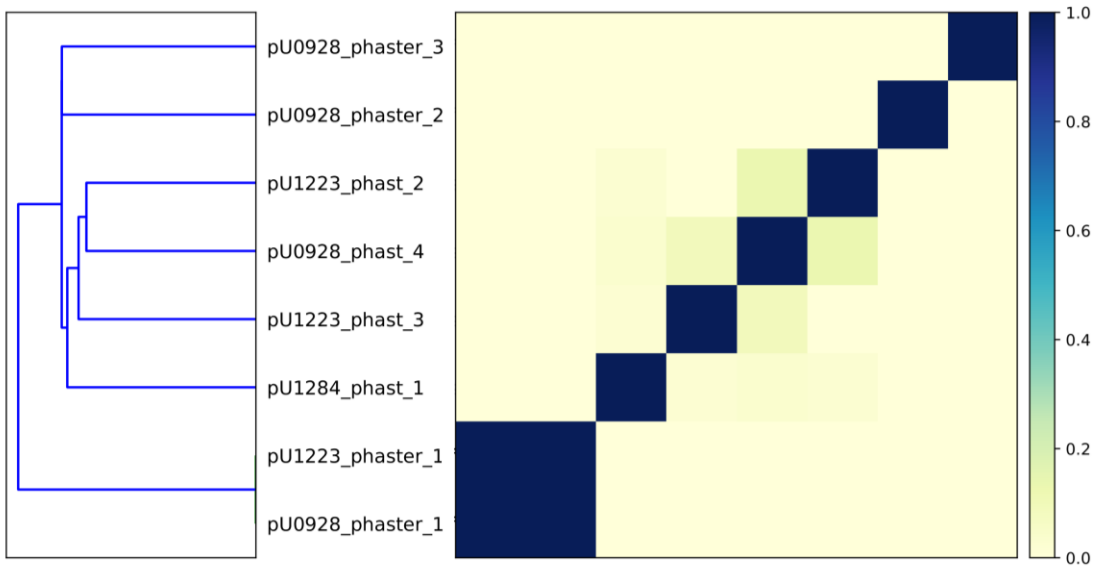


Figure 23. Comparison of Phage Sequences Predicted in pU0928, pU1223, pU1284

pU0928 is predicted to have four phage-like sequences, pU1223 has three, and pU1284 has one. One phage-like sequence from pU0928 (pU0928_phaster_1) is very similar to a phage-like sequence in pU1223 (pU1223_phaster_1). The rest of the phage-like sequences are distinct in all three plasmids. Black denotes 100% sequence homology and white is 0% sequence homology.

Figure 24. Predicted Phage Maps for pU0928_phaster_1, pU0928_phaster_2, pU0928_phaster_3.

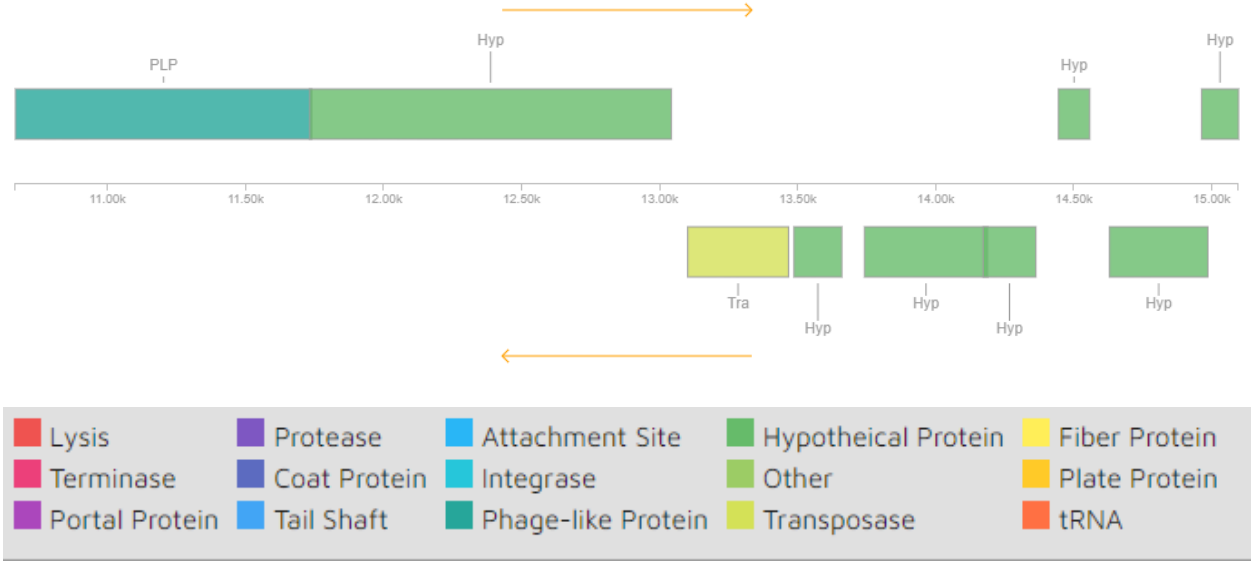


Figure 25. Predicted Phage Maps for pU1223_phaster_1.

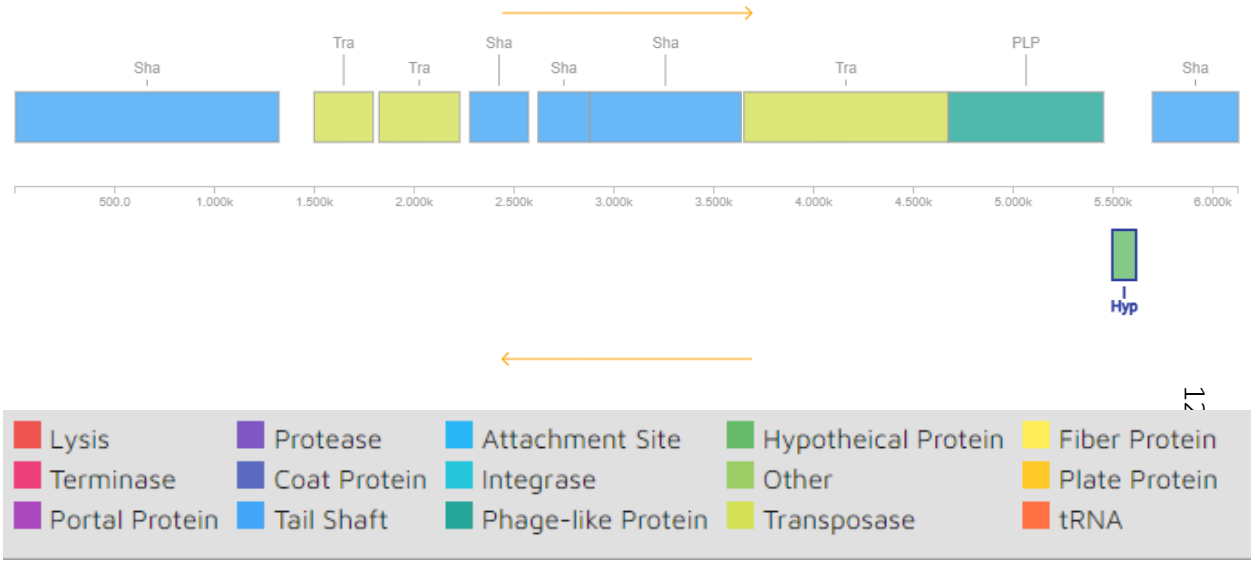


Figure 26. Predicted Phage Maps for pU1284_phaster_1.

Table 42. ORFs Shared in pU0928, pU1223, and pU1284.

		pU0928	pU0928	pU0928	pU0928	pU1223	pU1223	pU1223	pU1284
Annotated ORF	Predicted function	phaster_1	phaster_2	phaster_3	phast_4	phaster_1	phast_2	phast_3	phast_1
Putative transposase InsK for insertion sequence element IS150	Transposase				+			+	+
Tyrosine recombinase XerC	Phage integrase				+				+
>pU1284_phast.prokka_CIILILIG_00032 hypothetical protein	Dihydrofolate reductase				+				+
>pU1284_phast.prokka_CIILILIG_00034 hypothetical protein	Transposase				+				+

Note: The Phage integrase and dihydrofolate ORFs are the same ones identified to be shared by pU0928 and pU1284

Table 43. Predicted Phage Sequences in Urinary *E. coli* Plasmids with Phage-Integrase.

Isolate	Phage spot	Inc group	Phage #	Bases	Score (%)	# ORFs	Phage hit
UMB0906	No	IncFII	1	4.1Kb	10	6	PHAGE_Entero_P1_NC_005856(4)
UMB0928	No	IncFI	1	9.1Kb	40	14	PHAGE_Gordon_Hedwig_NC_031099(1)
			2	6.1Kb	90	10	PHAGE_Stx2_c_Stx2a_F451_NC_049924(3)
UMB0931	Greed	IncFII	1	93.8Kb	70	121	PHAGE_Escher_RCS47_NC_042128(54)
			2	6.6Kb	30	10	PHAGE_Stx2_c_Stx2a_WGPS9_NC_049923(2)
UMB0933	No	IncFII	1	21.7Kb	40	23	PHAGE_Cronob_ENT39118_NC_019934(2)
			2	6.4Kb	30	13	PHAGE_Cronob_vB_CsaM_GAP32_NC_019401(1)
			3	4.3Kb	50	8	PHAGE_Escher_RCS47_NC_042128(2)
UMB0934	No	IncFII	1	4.1Kb	20	7	PHAGE_Escher_RCS47_NC_042128(4)
UMB0949	No	IncFII	1	9.9Kb	40	7	PHAGE_Stx2_c_1717_NC_011357(3)
			2	4Kb	60	7	PHAGE_Stx2_c_1717_NC_011357(2)
			3	4.9Kb	60	10	PHAGE_Stx2_c_1717_NC_011357(3)
UMB1091	No	IncFII	1	6.1Kb	80	9	PHAGE_Stx2_c_Stx2a_F451_NC_049924(3)
			2	5.2Kb	50	7	PHAGE_Salmon_SJ46_NC_031129(3)
			3	3.9Kb	70	7	PHAGE_Stx2_c_1717_NC_011357(3)
UMB1160	P1vir	IncFII	1	28Kb	30	9	PHAGE_Salmon_SJ46_NC_031129(2)
UMB1193	No	IncFII	1	16.9Kb	30	10	PHAGE_Entero_P1_NC_005856(2)
			2	4.3Kb	50	8	PHAGE_Escher_RCS47_NC_042128(2)
UMB1223	No	IncFII	1	4.1Kb	20	7	PHAGE_Escher_RCS47_NC_042128(4)
UMB1229	No	IncFII	1	16.1Kb	30	10	PHAGE_Escher_RCS47_NC_042128(1)
UMB1284	No	IncFII	1	7Kb	10	12	PHAGE_Brucel_BiPBO1_NC_031264(2)
UMB1285	No	IncFII	1	7.6Kb	30	15	PHAGE_Escher_RCS47_NC_042128(3)
UMB3641	No	IncFII	0				
UMB5924	P1vir	IncFI	1	6.1Kb	90	10	PHAGE_Escher_SH2026Stx1_NC_049919(3)

Table 44. Predicted Phage Sequences in Urinary *E. coli* Plasmids that Have the Phage-Integrase.

UMB5978	No	IncFI	1	6.1Kb	90	10	PHAGE_Stx2_c_Stx2a_F451_NC_049924(3)
			2	9.7Kb	10	8	PHAGE_Escher_RCS47_NC_042128(3)
UMB6653	No	IncFII	1	6.4Kb	30	12	PHAGE_Pseudo_phiPSA1_NC_024365(1)
			2	6.1Kb	80	9	PHAGE_Stx2_c_Stx2a_F451_NC_049924(3)
UMB6721	No	IncFII	1	6.4Kb	30	12	PHAGE_Escher_Lys12581Vzw_NC_049917(1)

Note: Score of >90 is an intact phage

E. coli Plasmids Share Similarities to *K. pneumoniae* Plasmids

To assess if urinary *E. coli* can share plasmid content with other urinary species, isolates from clinically relevant urinary species had their whole genome sequence scanned for *inc* and *rep* genes. Evidence of plasmid content was observed in *E. faecalis, K. pneumoniae, S. epidermidis, and S. aeruginosa* (Table 44). Urinary *K. pneumoniae* plasmids had *incF* and *col* genes similar to those present in urinary *E. coli* plasmids. The whole genome raw sequence reads of three urinary *K. pneumoniae* isolates were assembled into plasmidic contigs via plasmidSPAdes and contigs were curated by checking for plasmid homology via web BLAST. The plasmidic contigs from *K. pneumoniae* were scanned with Prokka and web BLAST. *K. pneumoniae* plasmids were approximately 100k bp and had 30-50% of their ORFs annotated with a function (Table 45). These plasmids had web BLAST hits to similarly large plasmids in *K. pneumoniae* (Table 46). The plasmids were predicted to code for conjugation genes (i.e., *tra* genes) in addition to plasmid replication and virulence genes (Table 47). Plasmids from *K. pneumoniae* UMB7783 and UMB8492 were predicted to have a phage-like sequences that could be prophages (Table 48).

Table 45. Inc and Rep Proteins in Urinary Species Whole Genome Sequence.

Species	Clinical relevance	Isolates analyzed	Plasmid signature genes identified by PlasmidFinder
Aerococcus urinae	Pathogen	7	None
Corynebacterium amycolatum	Commensal	5	None
Enterococcus faecalis	Pathogen	3	repUS43, rep9b
Gardnerella vaginalis	Commensal	5	None
Klebsiella pneumoniae	Pathogen	3	Col, IncFIB, IncFII
Lactobacillus gasseri	Commensal	6	None
Lactobacillus jensenii	Commensal	8	None
Staphylococcus epidermidis	Commensal	9	rep7a, rep10, rep15, rep20, rep22, repUS23, rep39, repUS46, repUS76
Streptococcus anginosus	Commensal	8	repUS43
Streptococcus mitis	Commensal	8	None

Table 46. Annotation Summary for Urinary *K. pneumoniae* Plasmidic Assemblies.

Isolate	Species	Size (bp)	ORF annotated	ORF hypothetical	% annotated
UMB7779	*Klebsiella pneumoniae*	96352	53	55	49.07
UMB7783	*Klebsiella pneumoniae*	99217	39	71	35.45
UMB8492	*Klebsiella pneumoniae*	102300	40	81	33.06

Table 47. Top Web BLAST Result for Urinary *K. pneumoniae* Plasmidic Assemblies.

Strain	Taxonomy	Size (bp)	Plasmid hit	Reference size (bp)	Query coverage	E value	Per. Ident	Accession
UMB7783	*Klebsiella pneumoniae*	96352	pAR_0096	100759	100%	0	100.00%	CP027614.1
UMB8492	*Klebsiella pneumoniae*	99217	pAR_0096	100759	100%	0	99.94%	CP027614.1
UMB7779	*Klebsiella pneumoniae*	102300	pKpn3-L132	150325	80%	0	99.87%	CP040025.1

Table 48. Summary of Relevant Genes in Urinary *Klebsiella* Plasmidic Assemblies.

Strain	Taxonomy	Inc/Rep hits	Plasmid transfer	Plasmid replication	Virulence
UMB7783	*Klebsiella pneumoniae*	IncFII(K)	*traA, traM, tra_I_1, traD_1, finO_1, traN, finO_2, traI_2, traS. traD_2*	*ssb* (plasmid)	*traT, corA, merA, merC, merP, merT, clsA, imm*
UMB8492	*Klebsiella pneumoniae*		*traD, traQ, traN, traC, traV, traA, traM, finO*	*ssb* (plasmid)	
UMB7779	*Klebsiella pneumoniae*	IncFII(K)	*fhO, traI, traD, traQ, traN, traC, traV, traA, traY, traJ, traM*		*traT, imm*

Table 49. Phages Predicted in Urinary *Klebsiella* Plasmids.

Isolate	Inc group	Phage #	Region Length	Score	# ORF	Most Common Phage
UMB7779	IncFII(K)	0				
UMB7783	IncFII(K)	1	27.2Kb	50	33	PHAGE_Sodali_phiSG1_NC_007902(2)
UMB8492		1	21.5Kb	80	40	PHAGE_Escher_RCS47_NC_042128(2)

Note: Score of >90 is an intact phage

The sequences of urinary *E. coli* plasmids were compared to plasmids predicted to be in urinary *K. pneumoniae*. Plasmids from urinary *K. pneumoniae* clustered with plasmids in urinary *E. coli* UMB0939 (col), UMB1180 (col), UMB1228 (incX1), UMB4746 (IncFII) (Figure 27). As observed in a previous analysis, urinary *E. coli* UMB0939 had a top web BLAST hit to a *K. pneumoniae* plasmid in the NCBI database (Table 14). The other *E. coli* UMB isolates with *K. pneumoniae* web BLAST hits were UMB1193 (IncFII) and UMB1727 (IncFII), which do not cluster with the urinary *K. pneumoniae* included here but rather the “red” cluster composed of urinary *E. coli* F plasmids (Figure 27).

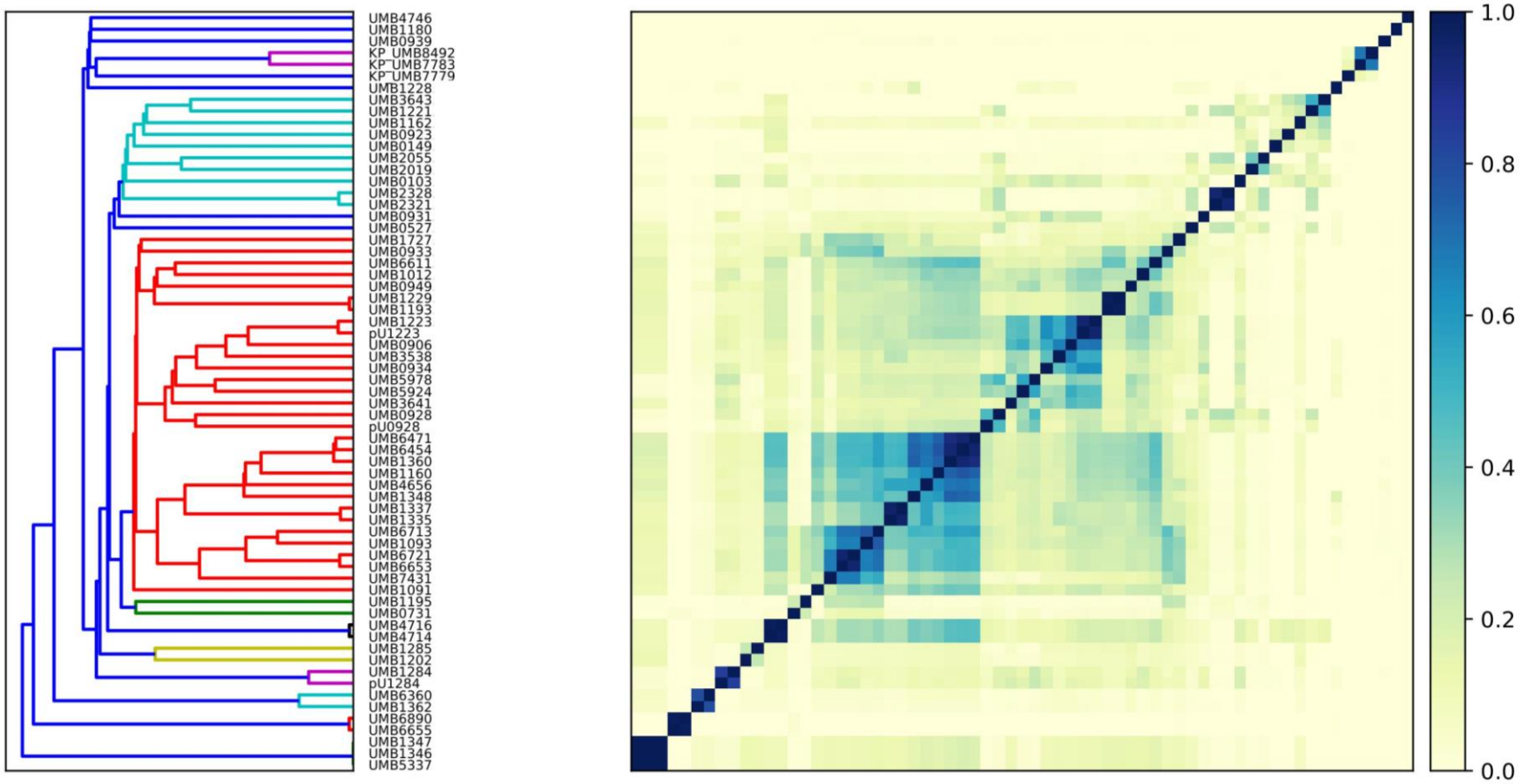


Figure 27. Comparison of Urinary *E. coli* Plasmids to Urinary *K. pneumoniae* Plasmids.

The plasmid sequence from urinary *E. coli* and *K. pneumoniae* was compared. Black denotes 100% sequence homology and white is 0% sequence homology. Plasmids from urinary *K. pneumoniae* (KP_UMB7789, KP_UMB7783, KP_UMB8493) clustered with plasmids in urinary *E. coli* UMB0939 (col), UMB1180 (col), UMB1228 (incX1), UMB4746 (IncFII) (Figure 27). Urinary *E. coli* UMBs with *K. pneumoniae* web BLAST hits were UMB1193 (IncFII) and UMB1727 (IncFII), which do not cluster with the urinary *K. pneumoniae* included here but rather the "red" cluster composed of urinary *E. coli* F plasmids.

CHAPTER FOUR:

DISCUSSION

Phage, Plasmids, and Urinary *E. coli*

Like in other niches in the world, phage may be an important determinant of bacteria population dynamics in the urinary tract[49,75,85,169]. Phage can interact with bacteria in a variety of ways, from facilitating horizontal genetic exchange (e.g., transduction, lysogeny), disrupting populations (e.g., lysis), and protecting against other phage (e.g., superinfection immunity and exclusion)[51,60,99,117,170]. Phage interactions may differ not just between different host species, but even at the level of strains within a species[85]. Understanding the interactions of phage with its host depends on understanding the genetic determinants of infection in both the phage and the host[50,85,171]. My scanning for the presence of 24 anti-phage and phage receptor genes revealed variance of nine genes (*rexB, imm, hok/shok, mazEF, pemIK*, and *traT*) in urinary *E. coli* compared to phage-susceptible laboratory *E. coli* (Figure 1, Table 9, Figure 3). The presence of *rexB* and *imm* genes in urinary *E. coli* could indicate the presence of prophage in the urinary *E. coli* genome, potentially integrated into the host chromosome or a plasmid, or circularized as a plasmid-like entity[116,172]. The genes *hok/shok, mazEF, pemIK*, and *traT* are all plasmid-linked[113,116,125,173]. *hok/shok, mazEF*, and *pemIK* are TA modules involved in plasmid retention, while *traT* blocks invasion by foreign plasmids. The absence of these plasmid-linked genes in phage-permissive *E. coli* coupled with their presence in multiple urinary *E. coli* less susceptible to infection led to my hypothesis that plasmids in urinary *E. coli* could reduce permissivity to phage.

Plasmids are mobile genetic elements that transmit and retain traits in bacterial populations[142,174]. F plasmids are an especially important plasmid type in *E. coli*, characterized by high

host persistence, a heterogenous genetic composition, and prevalence of antibiotic resistance and virulence genes[113,136,142]. There is evidence *in vitro* that plasmids can maintain and transmit genes that antagonize phage infection[106,113,125]. Phage predation can drive the acquisition of plasmids, including those with antibiotic resistance, in bacteria populations[105,119]. Plasmids were predicted to be present in approximately 85% of urinary *E. coli* isolates in this study (n=57/67). Twenty-eight distinct plasmid *inc* genes were present in these isolates; the plasmids were grouped as either IncF, Col, or Inc-various (Table 12). The most common Inc type was the IncF gene of the F plasmid group (68.65% of isolates) (Figure 5a). The overall sequence of these plasmidic assemblies was often homologous to plasmids in the NCBI database, primarily those from *E. coli*, but also plasmids from *Salmonella, Klebsiella,* and *Enterobacter* (Table 14). The widespread presence of plasmids in this set of urinary *E. coli* isolates may indicate that plasmids are relevant to *E. coli* in the urinary tract[142,175]. In this context, F plasmids are an especially important type of plasmid due their ability to maintain and transmit antibiotic, virulence, and fitness genes[113,142,175]. Plasmid replicons were assessed in the terms of phage permissivity, and urinary isolates in the IncFII group were statistically more likely to not be permissive when compared to isolates with plasmids from the other *inc* groups and to laboratory *E. coli* strains (Figure 5b). A key component of this study was to experimentally assess urinary plasmids in the context of phage infection, the first step consisting of profiling the genetic content in these urinary *E. coli* plasmids.

Genes in Urinary *E. coli* Plasmids

The function of ORFs from the urinary *E. coli* plasmid assemblies was predicted via sequence homology. In total, only 24% of plasmid ORFs had a function predicted, while 76% had hypothetical functions (Table 15). The IncF group had the most unique ORFs predicted (2060), but only 26.8% had a function assigned, compared to 18.8% in the col group and 17.0% in the inc-various group (Table 15). These results indicate that urinary *E. coli* plasmid gene content has much room for discovery, even in the clinically relevant F plasmid.

The F plasmids are of particular interest in terms of their potential for conjugation and retention, given how widespread they were in the isolates in this and previous studies. The annotated ORFs were reviewed for transfer genes, primarily *tra* genes, and for type IV secretion genes, which are typically required for F plasmid conjugation (Figure 6, Figure 7, Table 16). The transfer genes *traC, traD, traI, traJ, traM, traN, traQ, traR, traV,* and *traY* were in more than half of F plasmid assemblies. In contrast, the transfer genes *traB, traG, traL*, and *traS* were only in a minority of F plasmid assemblies. In addition, Type IV secretion system genes were identified in 16 isolates. The gene *fim*, which blocks the fertility/transfer of invading plasmids, was present in more than 80% of F plasmid assemblies. The plasmid retention/addiction modules *pemIK* and *ccdAB* were found in approximately half of F plasmid assemblies, with minor hits for other TA systems as well. Taken together, these data indicate that urinary *E. coli* have prototypical mechanisms for conjugation and retention relative to what has been reported in the literature[113,141].

I reviewed the antibiotic resistance content for all urinary *E. coli* plasmid assemblies. The IncF group had hits to all eight antibiotic resistances scanned (aminoglycoside, beta-lactam, macrolide, sulfamethoxazole, tetracycline, trimethoprim, fluoroquinolone, and phenicol) (Table 17, Figure 8a). The most common antibiotic resistance predicted in F plasmids was for beta-lactams, followed by sulfamethoxazole and tetracycline. Col plasmids had hits only to beta-lactam resistance genes, while plasmids from the Inc-various groups had no predicted hits. On average, IncF plasmids had more hits per assembly than did col and inv-various plasmids (Figure 8b). With that said, the variance for hits in the IncF plasmids was so high that statistical significance was not reached (Figure 8c). This can be explained by the observation that some plasmids in the IncF group had zero antibiotic resistance hits, while other plasmids carried multiple resistances, up to a maximum of all eight distinct antibiotic resistances (Figure 8c). However, when comparing multiple distinct antibiotic resistance hits, IncF plasmids skewed towards multiple resistances, whereas Col and Inc-various plasmids are in the zero to one range. These data

agree with reports of F plasmids in the literature, which indicate that multiple antibiotic resistances are common in these plasmids[142,175].

Next, I reviewed virulence gene content for all urinary *E. coli* plasmid assemblies. Following the antibiotic resistance pattern, F plasmids carried a wider breadth of distinct virulence genes and a larger proportion of hits in their assemblies compared to the other plasmid groups (Figure 9). Thirty distinct virulence genes were identified in all combined plasmid assemblies, with 29 distinct genes present in IncF assemblies. The most common virulence genes were *traT* and *senB*, present in 66.7% and 45.6% of plasmid assemblies, respectively. *traT* has been mentioned throughout this dissertation, involved both in blocking invasion by plasmids and antagonizing phage binding to membrane receptors[107,125]. *senB* is an entero-toxin present in F plasmids, linked to severe diarrhea after infection by both *Shigella* and entero-invasive *E. coli* (EIEC)[176,177]. The col plasmids are characterized by the presence of colicin-related virulence genes[140]. However, one urinary col plasmid was predicted to carry *traT* and *senB*, which are common to F plasmids; this could indicate genetic exchange between plasmids of different incompatibility groups[121]. On average, IncF plasmids were predicted to carry more virulence gene hits than col plasmids and inc-various plasmids. However, the large variance of hits in the plasmid groups resulted in the lack of statistical difference. When assessing multiple distinct virulence genes in the plasmids, IncF and col plasmids skew towards multiple hits. The distinction in multiple virulence hits from IncF and col plasmids is that the former had a greater variety of distinct virulence genes, whereas col plasmids had mostly colicin-related genes.

The Prokka ORF annotations were reviewed manually to identify miscellaneous gene functions. Genes of interest were those predicted to be involved in plasmid replication, metal transport and resistance, multidrug resistance genes, phage genes, and virulence regulators. The F plasmid group had the highest count of distinct ORFs with known functions, likely due to including the largest number of plasmids (Table 18). Urinary F plasmids having a high ORF count is in accordance with the literature's

view that F plasmid code for heterogenous and diverse content[175]. Overall, the gene profiling of the urinary *E. coli* plasmids trends for the types of genes expected to be present in F and col plasmids as per other studies. Beyond the genetic content annotated with a known function, a looming question is what do all the hypothetical genes do, especially as genes of unknown function constitute most of the genetic content. Even so, we could utilize the profiled genetic content to facilitate the transfer of these plasmids into a control *E. coli* background to test their effect on phage infection.

Anti-phage Plasmids in Urinary *E. coli*

Experimentation of clinical bacteria isolates at the bench can be challenging due to the unpredictable traits expressed by these untamed genomes. To facilitate testing of specific genetic content, one can transfer this content to a control background, such as *E. coli* K-12[103,104]. Genetic data from previous sections was used to identify plasmids that could be transferred to *E. coli* K-12, primarily those that had potential for conjugation, as well as antibiotic resistance genes that could be used as selection markers. I grew the urinary *E. coli* isolates on antibiotics commonly used at the bench as selection markers and found growth on ampicillin and tetracycline to be the most common (Table 21, Figure 11). Given the genetic tools available in our lab, I chose tetracycline as a selection marker for bench experiments, and reviewed the plasmidic assemblies for urinary *E. coli* for both conjugation and antibiotic resistance genes for tetracycline. Five urinary *E. coli* isolates (UMB0928, UMB1223, UMB1284, UMB1091, UMB6721) met these criteria and generated *E. coli* K-12 transconjugants that could grow on the appropriate selection markers (Table 22).

The most pertinent question to answer was whether *E. coli K*-12 carrying urinary plasmids exhibit an altered phage permissivity phenotype. Two of the *E. coli* K-12 transconjugants (carrying pU0928 and pU1284) exhibited a phage spot titration phenotype that indicated reduced permissivity to infection (Figure 12). Tested on multiple constructs, I observed reduced permissivity to phage infection only in pU0928 or pU1284 transconjugants, but not in pU1223, pU1091, or pU6721 transconjugants.

Variation on the location of the *E. coli* K-12 selection marker (chromosome or vector), *E. coli* K-12 strain, or variations on the *E. coli* K-12 gene deletions did not affect these results (Table 21-26). *E. coi* K-12 transconjugants of pU0928, pU1284, and pU1223 were infected with a range of phage concentrations (MOI 0.0, 0.01, 10.0) and had their OD measured across time. The protective effect was more apparent when *E. coli* K-12 pU0928 and pU1284 transconjugants were infected with P1vir or Lust, with infections of MOI 0.01 and 10.0 displaying similar growth curve results (Figure 13-15). Greed infection of the *E. coli* K-12 pU0928 or pU1284 transconjugants at an MOI of 10.0 was comparable to infection of the *E. coli* K-12 control at an MOI of 0.01 despite the difference in phage concentration on infection differing by four orders of magnitude.

While mechanisms underlying the anti-phage phenotype mediated by these plasmids has not been elucidated, my results allow me to make multiple useful deductions. The protective effect of pU0928 and pU1284 could be overcome at high concentration of phage during initial infection. This indicates that the anti-phage mechanism is not that of complete immunity, as would be expected by modification of the adsorption site of the phage[79,110]. The concentration-dependent effect may indicate a stoichiometric relationship between the infecting phage particles and protective traits expressed by the host. P1vir, Greed, and Lust are distinct lytic phages yet exhibited similar spot titration phenotypes and growth characteristics when infecting K-12 carrying the two anti-phage plasmids, pU0928 and pU1284. One thing to consider is the fact that there may be many distinct phage in the urinary tract; therefore, the odds of pU0928 and pU1284 having defense mechanism specific to these three phages are slim[20,75,178]. The decreased permissivity to infection could be explained by an anti-phage mechanism that targets a conserved aspect of the lytic phage cycle[60,69,117].

A limitation of the spot titration or growth curve assays is that they are relatively short term; therefore, an unanswered question is how these plasmids maintain themselves in urinary isolates in the long term (i.e., days to weeks). Even if anti-phage plasmids are protective, they may not be biologically

relevant if they are easily lost[113]. *E. coli* strains with pU0928, including the parent UMB0928, were passaged for 10 days in the absence of selection for plasmid (i.e., no tetracycline in the growth media to retain the plasmid). Even on the last day tested, all *E. coli* hosts carrying pU0928 were able to grow on tetracycline plates to a similar extent as LB plates, indicating retention of the plasmid (Figure 16). Review of the UMB0928 plasmid assemblies indicated presence the TA systems *ccdAB* and *pemIK*, which could explain the plasmid's ability to persist[113,124].

We can consider the implications of anti-phage plasmid acquisition by *E. coli* in terms of phage interactions and broad urobiota dynamics. Unlike naïve *E. coli* K-12, the pU0928 and pU1284 transconjugants can maintain a higher OD when exposed to phage (Figure 13-15). The plasmids can be readily conjugated due to their innate transfer systems; given how common F plasmids were in all these urinary isolates, we can infer that propagation of anti-phage gene content via plasmids is possible in the urinary tract. Propagation of this content may occur at an expedited rate in the presence of phage selective pressure, like the way in which antibiotic use can drive the dissemination and retention of R plasmids in bacteria populations[123,179]. pU0928 and pU1284 *E. coli* K-12 transconjugants could grow on the same antibiotics as the urinary plasmid donor (compared to the parent *E. coli* K-12, which can only grow on chloramphenicol), a phenotype that persisted even in the absence of selection (Table 27, Figure 16). The high stability of these plasmids and their anti-phage mechanism could be a factor underlying the propagation and maintenance of F plasmids in urinary *E. coli* populations, including plasmid-linked antibiotic and resistance genes[113,139].

Sequence Analysis of Anti-phage Plasmids from Urinary *E. coli*

To profile the genetic content in the urinary plasmids conjugated in previous experiments (pU0928, pU1223, pU1284, pU1091, pU6721), their genomes were extracted and sequenced. This procedure was performed as both a plasmid extraction and a whole genome extraction. The plasmid extraction was inconclusive for all isolates attempted, as there was no evidence of plasmid content after

sequencing (likely due to technical issues). When the sequence raw reads from the whole genome extraction were assembled into plasmidic assemblies, however, there was evidence of plasmid content (Table 29-31). The plasmidic assemblies for pU0928, pU1223, and pU1284 had nucleotide sizes and *inc* genes like those predicted in the urinary parent (Table 29-31). pU6721 was predicted to be 39,608 bp and have an *incX* gene, which differed from the plasmid prediction in UMB6721 (250,312 bp and with *incFIB, col, incB/O/K/, incFII*, and *incX4*). A plausible explanation for these results is that UMB6721 had multiple plasmids and only the one with *incX* was transferred to *E. coli* K-12. The final plasmid, pU1091, had a predicted size of 10,546 bp with no *inc* genes identified. However, the web BLAST hits for pU1091 were those of *E. coli* chromosomal sequence, supporting the conclusion that the extraction and sequencing of UMB1091 did not yield plasmid sequences, likely due to technical limitations (Table 30). The size, inc hits, and web BLAST results for pU0928, pU1223, and pU1284, however, support the conclusion that these three are indeed F plasmids (Table 29-31). All three F plasmids have evidence of multiple antibiotic resistances, including tetracycline resistance, which was used as a selection marker in the conjugations. pU0928 is predicted to encode the virulence genes *cia* and s*enB*, whereas pU1284 is predicted to encode *traT* (Table 31). The curated plasmid assemblies for pU0928, pU1223, and pU1283 were used as a reference to map the raw sequence reads from whole genome sequencing. No major gaps were observed in terms of coverage (Figure 17).

The curated assemblies for pU0928, pU1223, and pU1284 were annotated specifically to identify ORFs responsible for the anti-phage phenotype (Table 32). In all three phages, approximately 40% of the ORFs were assigned a function by the annotation software Prokka, indicating that most of the content in these plasmids has not been elucidated. On reviewing the ORFs with a predicted function for pU0928, there were no genes that stood out as anti-phage genes (Table 33). Of note is that anti-phage genes *traT*, *ccdAB* and *pemIk* are predicted to be in the parent UMB0928 plasmid assembly but were absent in pU0928. This could indicate that some of the plasmid content from UMB0928 was not passed on to the

E. coli K-12 recipient or that there were technical shortcomings in the extraction, sequencing, assembly, or curation steps that filtered out these ORFs from pU0928. Potential anti-phage ORFs in pU1284 are *ccdAB, pemIK*, and a type II restriction enzyme (Table 34). Despite not having an anti-phage phenotype, pU1223 also was predicted to code for *ccdAB* and *pemIK* (Table 35). pU0928 and pU1284 were compared to the anti-phage ORFs scanned in Figure 1, with *imm* being present in pU0928 and *traT* in pU1284. *imm* is often associated with T4 phage and the phage immunity protein *Sp*, but neither of these were present in pU0928 or pU1284[180,181].

All ORFs were reviewed to assess if any were present in both pU0928 and pU1284 (phage non-permissive plasmids) and absent in pU1223 (phage permissive plasmid). There were 18 ORFs present in all three plasmids, but only three ORFs were present in both pU0928 and pU1284 (Table 36). After assessing their web BLAST hits and protein domains, the function of these three ORFs was respectively predicted to be that of a phage integrase, dihydrofolate reductase, and a gene with an EAL/cyclic di-GMP phosphodiesterase domain (Table 37, Figure 18-20). The first two predicted gene functions can be linked to phage biology. Phage integrases are utilized by phage to integrate into the host genome[182,183]. Dihydrofolate reductase is a bacterial enzyme that can be replaced by a phage-encoded homolog during infection, as finely regulated activity of this enzyme often is necessary for proper phage particle assembly[184,185]. The ORF with the with EAL/cyclic di-GMP phosphodiesterase domain is more enigmatic and could be involved in signaling[186].

These three ORFs are highly conserved when comparing their amino acid sequences from pU0928 and pU1284, in contrast to the 14% overall homology of these two plasmids (Table 38). Given that these genes could be common in other urinary *E. coli* isolates, all urinary *E. coli* plasmids were scanned for the three ORFs. The phage integrase was present in 16 urinary *E. coli* plasmidic assemblies, dihydrofolate reductase was present in 13, and the EAL domain containing ORF was in 10 (Table 39). These three ORFs were co-present in nine of these urinary plasmids. All 16 isolates with these ORFs are

predicted to be F plasmids, and the majority have no evidence of phage infection after phage spotting. The nucleotide sequence of these plasmids was compared to assess if these plasmids were similar. Overall, these plasmids are not alike, but rather they form multiple clusters (Figure 21). F plasmids are known to be heterogenous in terms of content, yet these three ORFs are highly conserved (>90 query coverage and sequence identity). If these ORFs are indeed involved in phage protection, potentially their presence and conserved sequence indicates importance in terms of phage selection pressure.

Phage-like Genetic Content in Anti-phage Plasmids from Urinary *E. coli*

The phage integrase ORF may indicate the presence of prophage or phage-like sequences in pU0928 or pU1284[182]. A phylogenetic analysis was performed using the phage integrase sequence from pU0928a and pU1284, in addition to phage integrases from bacterial genomes, plasmids, and phage. The phage integrase from pU0928 clusters with phage integrases from plasmids, while the phage integrase from pU1284 clusters with phage integrases from bacterial genomes (Figure 22). Given the possibility of phage-like content in pU0928 and pU1284, their sequences were scanned via PHAST and PHASTER. There were four phage-like sequences in pU0928 and one in pU1284 (Table 40, Figure 23).

The ORFs in these phage-like sequences were reviewed using PHASTER output and Prokka annotation. The PHASTER plasmid maps indicate multiple predicted genes with phage-like but mostly hypothetical functions, with varying degrees of phage homology completeness (i.e., how similar this sequence is to that of a known phage) (Figure 24-26). pU0928 and pU1284 could either have functional prophages or perhaps phage-like sequences that cannot be induced into a productive phage. A highlight of the Prokka annotation is that the two of the three ORFs shared by phage-like sequences in pU0928 and pU1284 are the same phage integrase and dihydrofolate reductase identified in the previous analysis (Table 39). When reassessing the urinary *E. coli* plasmids that have the phage integrase in their F plasmid, all but one were predicted to have phage-like sequences (Table 42).

Given these data, a mechanism for the anti-phage phenotype could be linked to phage-like

genes in pU0928 and pU1284. Prophage are known to provide superinfection immunity and exclusion, which is further supported by the presence of the infection exclusion gene *imm* in pU0928[60,117,181]. Despite the relatively simple explanation that phage protect urinary *E. coli* from other phage, most of the content in these anti-phage plasmids and their phage-like sequences remain unknown. Furthermore, if phage were responsible for the protective effect, this raises additional questions concerning the role of plasmids in mobilizing and retaining phage-like content in urinary *E. coli*. Plasmids are exceptional in their ability to mobilize genetic content horizontally and persist vertically, which could open new avenues for phage "infection" in the urinary tract, even outside the strict host range of these phages[144,174].

Homology of *E. coli* Plasmids to Plasmids from Other Gram-Negative Species

There is evidence in the literature that *E. coli* can exchange plasmids with other species, which could extend to the urinary tract[173,187]. A total of 62 WGS from 10 urinary bacteria species were scanned for *inc* and *rep* genes as to be compared to those in *E. coli* (Table 44). *inc* or *rep* genes were present in *E. faecalis, K. pneumoniae, S. epidermidis*, and *S. anginosus* isolates. Three urinary *K. pneumoniae* isolates had hits for col, incFIB, and incFII, which were similar to those in urinary *E. coli*. The plasmidic assemblies from these *Klebsiella* isolates had homology to *Klebsiella* plasmids in the NCBI database and were predicted to be approximately 100,000 bp (Table 45). Only 30-50% of the ORFs in these plasmids were annotated with a function, which included the anti-phage *traT* and *tra* genes like those in *E. coli* (Table 47).

tra genes indicate the potential for plasmid conjugation within *Klebsiella* species and potentially to other Gram-negative species[188]. *K. pneumoniae* and *E. coli* have been noted in the literature to be able to exchange plasmids, including those with antibiotic resistance[187]. The urinary *K. pneumoniae* plasmids clustered with urinary *E. coli* plasmids when their entire nucleotide sequence was compared (Figure 27). The urinary *E. coli* plasmids that clustered with urinary *K. pneumoniae* were primarily

predicted to not be F plasmids. That stated, two F plasmid *E. coli* assemblies (UMB1193, UMB1727) had web BLAST top hits to *K. pneumoniae* plasmids in the NCBI database (Table 46). The plasmidic assemblies from UMB0928 and UMB3643 had homology to a plasmid from *Salmonella* and the plasmidic assemblies from UMB1180 and UMB1360 had homology to a plasmid from *Enterobacter* (Table 14).

The plasmidic assemblies from urinary *K. pneumonia* were scanned for phage sequences, with two of these predicted plasmids having a hit for a phage-like sequence (Table 48). If plasmid exchange occurs between different urinary species, phage could be shuttled in plasmids and thus access bacteria outside of their host range. Furthermore, non-*E. coli* species could act as reservoirs for anti-phage content, as these species may not be lysed by coliphage.

Model for *E. coli*, Plasmid, and Phage Interactions in the Urinary Tract

We can use inferences from the results of this project and others in the *E. coli* literature to propose a model for urinary *E. coli*, plasmid, and phage interactions. When an *E. coli* cell resident in the urinary tract is exposed to coliphage, it can defend itself against infection by having anti-phage genes in its chromosome, plasmids, or prophage (with the prophage circularized in the cytoplasm or integrated into the host genome)[75]. An advantage of plasmids is that they are pliable, non-essential mobile genetic elements that can be transmitted without fatally disrupting the host[174,188]. Chromosomal genes may have limits on the content that can be mutated, while prophage may require lytic activation for rapid propagation[99,104]. Moreover, plasmids have replication and conjugation machinery, allowing them to be efficiently transmitted vertically and horizontally[174,189]. Extrapolating from the data in this project, we can estimate that plasmids are widespread in urinary *E. coli* (Figure 5a).

The acquisition of plasmids in bacterial populations by phage predation has been observed *in vitro* but not studied in the context of the urobiome[105,106,119]. We have provided evidence that urinary *E. coli* plasmids can be transmitted to naïve strains (Table 22). This project presents evidence to support the hypothesis that urinary *E. coli* carrying an anti-phage plasmid may better survive exposure to lytic

phage infection relative to a permissive isolate (Table 24-26, Figure 12-15). This phenotype was observed even in the very phage-susceptible *E. coli* K-12. Given the mechanism of lytic infection, we can propose a scenario that parallels the phage-plasmid dynamic to the manner in which antibiotic use coupled with antibiotic resistance genes in plasmids may prime bacteria for survival[123,179]. Furthermore, just like antibiotic use, phage predation in the urinary tract may drive the transmission and persistence of anti-phage plasmids[85,97]. F plasmids are the most common plasmid identified in this collection of urinary *E. coli*, with antibiotic resistance and virulence genes frequent in the plasmid sequence[141,175,188]. K-12 that acquired pU0928 and pU1284 had the same multiple antibiotic resistances as the urinary parent (Table 27). Even in the absence of selection during passaging for ten days, *E. coli* K-12 could still grow on the antibiotic selection marker present on the F plasmid (Figure 16). A major issue in the clinical setting is the increasing frequency of antibiotic resistance in bacteria[94,123,132]. Potentially, phage could be a driver for the retention and spread of antibiotic resistance and virulence genes in the clinical setting.

At present there are multiple genes in this study's urinary *E. coli* plasmids that can be considered anti-phage genes. F plasmids were predicted to carry multiple TA systems, the phage-adsorption blocking *traT*, multiple phage-like sequences, genes with homology to phage integrases and restriction enzymes (Table 28). In addition, there were hundreds of ORFs with unknown functions that could be anti-phage genes (Table 15). The anti-phage phenotype could be a polygenic trait, potentially a redundant defense mechanism that is passed as a bolus on the urinary *E. coli* plasmid.

The final element to consider is that plasmid exchange may occur between *E. coli* and other Gram-negative species in the urinary tract[187]. The literature contains reports of *E. coli* exchanging plasmids with species like *Klebsiella* and *Shigella,* propagating traits like antibiotic resistance[145,190,191]. Multiple urinary *E. coli* plasmids have homology to the plasmids of Gram-negative species in the NCBI database (Table 14). Analysis of three urinary *K. pneumoniae* species indicated presence of *inc* genes similar to those in *E. coli* and with homology to urinary *E. coli* plasmids (Figure 39). Phage are often

specific to a species or a strain; therefore, *Klebsiella* could serve as a reservoir for anti-phage content where it is safe from coliphage predation.

Future Directions

Genes Involved in the Anti-phage Phenotype

Given that the anti-phage phenotype of pU0928 and pU1284 has been validated, the immediate unanswered question centers on the anti-phage mechanism. The initial step to study the mechanism is to identify the genes that result in the anti-phage phenotype. My work completed two major steps towards this pursuit: 1) the genes that result in the anti-phage phenotype have been localized to a discrete genetic element (i.e., a plasmid), and 2) the genetic content of interest is now in the genetically amenable *E. coli* K-12[103,104].

This said, there are challenges to keep in mind in designing strategies to identify the anti-phage mechanism. Most importantly, the anti-phage mechanism may be either a monogenic or polygenic trait[192]. If monogenic, then if the correct ORF is expressed in a vector then we would expect the anti-phage phenotype. It is more complicated for a polygenic phenotype, as all the necessary ORFs must be expressed together. Furthermore, the ORFs could provide redundant mechanisms that decrease permissivity to phage on a spectrum. One approach toward identifying the correct ORFs would be to randomly mutagenize the *E. coli* K-12 transconjugants, then identify those that exhibit increased permissivity to phage, and systematically narrow down the sites mutated. Another approach is to delete or clone whole segments of the urinary plasmid, and see if these result in loss or gain of the anti-phage phenotype, respectively.

The analysis of the ORFs in pU0928 and pU1284 indicated three shared ORFs, a phage integrase, dihydrofolate reductase, and a gene with an EAL cyclic di-GMP phosphodiesterase domain. An immediate step would be to express these ORFs in a vector and assess if they are enough to confer the

anti-phage phenotype in *E. coli* K-12. Once the anti-phage genes are validated, the literature can be reviewed for homologues that may have their mechanism of action reported.

Lifestyle of Phage that Infects *E. coli* Carrying Anti-phage Plasmid

A question relevant to the anti-phage phenotype in the urinary plasmids is how these affect the infection of the lytic phage. In the absence of the protective effect, the phage P1vir will undergo the lytic pathway of infection as it has been bioengineered to not lysogenize[98]. Given that lysis occurs if the phage titer is high enough, the anti-phage plasmid does not appear to provide immunity to phage infection (Figure 13-15). The primary question is what effect the anti-phage plasmid has on the lytic kinetics of the phage. To address this question, a one-step phage growth curve could be performed at decreasing MOI of P1vir to see if there is a threshold where evidence of lytic infection ceases[193]. Given that one of the ORFs of interest in pU0928 and pU1284 is a phage integrase, a secondary hypothesis is that the anti-phage plasmids could be forcing the phage into the lysogenic pathway. This can be tested by exposing the transconjugants to P1vir, passaging surviving cells to dilute out phage, then attempting to amplify P1vir DNA from the genome of the *E. coli* K-12 transconjugant.

Phage Predation as a Driver for the Acquisition of Plasmid-linked Traits

In vitro studies indicate that phage predation can drive the acquisition of plasmids in a bacteria population[118,119,133]. The anti-phage plasmids pU0928 and pU1284 are predicted to carry antibiotic resistance and virulence genes. When *E. coli* K-12 acquired these plasmids, it was not only able to maintain a high OD in the presence of phage, but it was able to grow on multiple antibiotics (Figure 13-15, Table 27). A question to address is whether this phenomenon can be replicated in the context of bacterial population dynamics.

An experiment that could test population dynamics between *E. coli*, its plasmids and its phage is to compete naïve *E. coli* K-12 and its transconjugant in the absence and presence of phage. If phage drive plasmid acquisition, we expect the transconjugants to outcompete their plasmid-naïve parent

when exposed to phage and thus increase the frequency of the plasmid. This could be measured by comparing the frequency of the selection marker in the two populations across time. The plasmid could be extracted from the transconjugants to confirm the presence of antibiotic resistance and virulence genes.

Alternatively, the transconjugants could outcompete their naïve parent even in the absence of phage. In contrast, the transconjugants may be at a disadvantage because plasmids are foreign genetic elements and are assumed to carry a fitness cost. Yet plasmids can carry factors that increase fitness via other means, such as additional metabolic or toxic genes. If this scenario occurs, it indicates additional factors of interest in the urinary plasmids as it relates to *E. coli* fitness in the urinary tract.

Transmission of Anti-phage Plasmids to and from Other Gram-Negative Bacteria

The anti-phage plasmid pU0928 is predicted to possess four phage-like sequences, while pU1284 has one predicted phage-like sequence. While these two plasmids only have approximately 14% query coverage when compared to each other and their phage-like regions are very distinct, they share two highly conserved ORFs (phage integrase, dihydrofolate reductase) in their phage-like regions (Figure 23, Table 41). Furthermore, pU0928 carries *imm*, which is a gene present in the prophage T4, even though there was no sequence homology to T4 in pU0928. The anti-phage phenotype conferred by these plasmids could be explained by prophage or phage-like genes. This project presents predicted phage maps for both plasmids; the sequence of these regions could be cloned in their entirety and expressed in a vector to assess if they confer the anti-phage phenotype.

Outside of understanding the mechanism of these genes, a pertinent question is whether these phage-like regions code for functional prophage. This can be tested by exposing the transconjugants to stressful growth conditions (e.g., limited media, high or low pH, heat) to induce phage and then use standard phage propagation techniques to increase the titer. Naïve *E. coli* K-12 could be infected with the phage lysate and the genetic sequence of the phage amplified via PCR. The phage permissivity of

these *E. coli* K-12 would be tested to assess if the prophage is protective.

The presence of phage-like genetic content in plasmids present new ways to think about bacteria, plasmids, and phage interactions in the urinary tract. Plasmid exchange between *E. coli* and other Gram-negative species has been documented in the literature. This project has evidence of plasmid homology between urinary *E. coli* and other Gram-negative species, particularly urinary *K. pneumoniae*. A relevant question is if the anti-phage plasmids pU0928 and pU1284 can be transmitted to and from urinary *Klebsiella* species. To test this, a conjugation assay could be attempted between *E. coli* K-12 transconjugants and urinary *Klebsiella* species, particularly those predicted to not have a plasmid-transfer blocking mechanism (e.g., *fim, traT*). To accomplish this, one could adapt the strategy for *E. coli* conjugation outlined in this project, where bioinformatics is used to computationally screen a multitude of urinary *Klebsiella* species for *tra* and selection marker genes (e.g., antibiotic resistance).

If transfer of pU0928 and pU1284 from *E. coli* to *K. pneumoniae* is successful, the next step would be to attempt the transfer of pU0928 and pU1284 from *Klebsiella* to a naïve *E. coli* K-12 strain, thus closing the conjugation loop. Finally, *Klebsiella* with the urinary *E. coli* plasmid could be passaged extensively to assess if the plasmid can be lost and thus establish *Klebsiella* as an anti-phage gene reservoir.

Conclusion

To our knowledge, this is the first study of the plasmidome in a diverse set *E. coli* strains from catheterized urine. Plasmids are important biological factors, both for urobiota dynamics and the clinical setting[175,179]. Most of the urinary *E. coli* analyzed in this work were predicted to have a plasmid, most of the F plasmid type. In accordance with F plasmids described in the literature, F plasmids in urinary *E. coli* possess antibiotic resistance and virulence genes, which are found at a higher frequency than in other urinary plasmids observed (col, inc-various)[142,175]. F plasmids in urinary *E. coli* encode the *tra* and Type IV secretion systems known to be involved in conjugation, in addition to multiple plasmid addiction TA

modules[113,188]. The gene content and stability of these plasmids could be relevant to *E. coli*'s role in urinary health[75].

Two urinary F plasmids were conjugated into *E. coli* K-12 and shown to subsequently decrease permissivity to three lytic phages. These F plasmids were stable in the absence of selection and provided multiple antibiotic resistances to the *E. coli* K-12 host. These anti-phage plasmids were sequenced and their ORFs determined. Most of the ORFs in these plasmids are predicted to have hypothetical functions, but anti-phage genes were identified: *imm* in pU0928 and *traT, ccdAB, pemIK* in pU1284[107,124,181]. Three ORFs are shared by the anti-phage plasmids pU0928 and pU1284; they are predicted to be a phage integrase, dihydrofolate reductase, and an EAL cyclic di-GMP phosphodiesterase domain protein. The first two ORFs can be linked to phage biology[182,184]. These three ORFs are highly conserved in both plasmids, despite pU0928 and pU1284 having low overall homology (14% cover query). The phage integrase ORF is present in 16 urinary *E. coli* isolates (n=16/57, 28% of isolates); presence of dihydrofolate reductase or the EAL domain-containing protein in these isolates always included the phage integrase. All isolates with these ORFs are predicted to have an F plasmid and most of them have no evidence of lytic infection when spotted with phage. In pU0928 and pU1284, both the phage-integrase and dihydrofolate sequence localizes to phage-like regions in the plasmids. All but one of the 16 urinary *E. coli* that have the phage integrase gene are predicted to have phage-like regions. These predicted phage-like sequences vary in completeness, indicating that they could potentially be intact prophages or uninducible phage-like sequences.

Phage are known to integrate in plasmids and propagate through them, which coupled with pU0928 and pU1284's ability to decrease phage permissivity, could bring a new understanding to *E. coli* dynamics in the urinary tract[194–196]. The final element to consider is that there is homology between urinary *E. coli* plasmids and other Gram-negative species. Anti-phage content in prophage and its inter-species exchange via plasmids may be another layer to consider in urinary dynamics and health[197].

In the immediate future, a direction to pursue is the elucidation of the anti-phage mechanism in these plasmids and their impact on bacteria-phage dynamics. Identification of genes that can reduce permissiveness could better help us understand why some strains of *E. coli* thrive in the urinary tract, which is especially relevant for UTI[28,75]. Identification of genes that antagonize the lytic life cycle could be employed to boost the defenses provided by commensal bacteria and probiotics[198,199]. Furthermore, the findings in this project should be tested in the context of bacteria-phage population dynamics, to assess if the presence of these anti-phage plasmids enriches specific genetic content (e.g., antibiotic resistance). Phage predation could be a factor that selects for the acquisition of plasmids by urinary *E. coli* reminiscent of the manner in which antibiotic use drives the acquisition and persistence of R plasmids in bacteria populations[123]. This bacteria-plasmid-phage model may not be limited to *E. coli*, and phage could play an important role in driving the plasmid-related gene flow in other species of the urobiome.

APPENDIX A

BASH SCRIPTS AND USEFUL COMMANDS

Bash Scripts and Useful Commands

This section lists scripts and commands I commonly used for analyzes. These scripts should only be used as a guide for commands, tools, and syntax to use. They should be updated accordingly as new tools, parameters, and commands are released. If scripts are not working but no errors are found, retype into a new text document (writing scripts in Windows sometimes introduces invisible characters).

Script to process genomes from raw sequencing reads to annotated ORFs

```
# Trim reads (Sickle)
#!/bin/Bash
for r1 in *R1*;
do
r2=${r1/R1/R2}
singles=trimmed${r1/R1/singles}
trimmed1=trimmed$r1
trimmed2=trimmed$r2;
/data/apps/sickle-master/sickle pe -f "$r1" -r "$r2" -t sanger -o "$trimmed1" -p "$trimmed2" -s
"$singles";
done

# Assemble reads (plasmidSPAdes)
#!/bin/Bash
for file1 in *trimmed*R1*fastq;
do
file2=${file1/R1/R2}
file3=${file1/R1/singles}
out=${file1%%.fastq}_output
/data/apps/SPAdes-3.12.0-Linux/bin/spades.py -k 55,77,99,127 -t 20 --only-assembler -1 $file1 -2 $file2 -
s $file3 -o $out;
done

# Copy and rename contig files with directory name, then move to main directory for downstream
manipulation (i.e. all contig files are originally named contig.fasta in the assembly directory)
#!/bin/Bash
for subdir in *;
do cp $subdir/contigs.fasta $subdir.fasta;
done;
```

```
# Get rid of reads under 500 bp (bioawk)
#!/bin/Bash
for file1 in *.fasta;do
file2=prune_${file1%%.fasta}.fasta
/data/apps/bioawk/bioawk -c fastx 'length($seq) > 500 {print ">"$name"\n"$seq"\n+\n"$qual}' $file1 >
$file2;
done

# Annotate contigs (Prokka)
#!/bin/Bash
for file1 in *prune*.fasta;
do
out=${file1%%.fasta}_prokka
/data/apps/prokka/bin/prokka --outdir $out --prefix $file1 --centre XXX -compliant $file1;
done

# Copy annotation summary of all assemblies
#!/bin/Bash
for subdir in *;
do cp $subdir/*.txt $subdir.txt;
done;

# Then rename the top header in the file with >'filename'
#!/bin/Bash
for file1 in *txt;
do
file2=${file1%%.txt}.txt2
awk 'NR==1{$0=$0">"FILENAME}1' $file1 > $file2;
done

# Concatenate all annotation summaries into single file
#!/bin/Bash
for file in *.txt2
do
cat *.txt2 >> annotation_summary;
rm *.txt
rm *.txt2
done;

# Move TSV files to directory
#!/bin/Bash
mkdir annotation

# Copy TSV files for all annotation
#!/bin/Bash
for subdir in *;
do cp $subdir/*.tsv $subdir.tsv;
done;
```

```
mv *.tsv annotation/

# Copy faa files for all annotation
#!/bin/Bash
for subdir in *;
do cp $subdir/*.faa $subdir.faa;
done;

mv *.faa ORF/
```

Script to process raw sequencing reads into annotated plasmidic assembly

```
# Trim reads (Sickle)
#!/bin/Bash
for r1 in *R1*;
do
r2=${r1/R1/R2}
singles=trimmed${r1/R1/singles}
trimmed1=trimmed$r1
trimmed2=trimmed$r2;
/data/apps/sickle-master/sickle pe -f "$r1" -r "$r2" -t sanger -o "$trimmed1" -p "$trimmed2" -s
"$singles";
done

# Assemble reads (plasmidSPAdes)
#!/bin/Bash
for file1 in *trimmed*R1*fastq;
do
file2=${file1/R1/R2}
file3=${file1/R1/singles}
out=${file1%%.fastq}_output
/data/apps/SPAdes-3.12.0-Linux/bin/plasmidspades.py -k 55,77,99,127 -t 20 --only-assembler -1 $file1 -
2 $file2 -s $file3 -o $out;
done

# Copy and rename contig files with directory name, then move for downstream manipulation (i.e. all
contig files are originally named contig.fasta)
#!/bin/Bash
for subdir in *;
do cp $subdir/contigs.fasta $subdir.fasta;
done;

# Get rid of reads under 500 bp
#!/bin/Bash
for file1 in *.fasta;
do
file2=prune_${file1%%.fasta}.fasta
```

```
/data/apps/bioawk/bioawk -c fastx 'length($seq) > 500 {print ">"$name"\n"$seq"\n+\n"$qual}' $file1 >
$file2;
done

# Annotate contigs (Prokka)
#!/bin/Bash
for file1 in *prune*.fasta;
do
out=${file1%%.fasta}_prokka
/data/apps/prokka/bin/prokka --outdir $out --prefix $file1 --centre XXX -compliant $file1;
done

# Copy annotation summary of all assemblies
#!/bin/Bash
for subdir in *;
do cp $subdir/*.txt $subdir.txt;
done;

# Then rename the top header in the file with >'filename'
#!/bin/Bash
for file1 in *txt;
do
file2=${file1%%.txt}.txt2
awk 'NR==1{$0=$0">"FILENAME}1' $file1 > $file2;
done

# Concatenate all annotation summaries into single file
#!/bin/Bash
for file in *.txt2
do
cat *.txt2 >> annotation_summary;
rm *.txt
rm *.txt2
done;

# Move TSV files to directory
#!/bin/Bash
mkdir annotation

# Copy TSV files for all annotation
#!/bin/Bash
for subdir in *;
do cp $subdir/*.tsv $subdir.tsv;
done;

mv *.tsv annotation/

# Copy faa files for all annotation
```

```
#!/bin/Bash
for subdir in *;
do cp $subdir/*.faa $subdir.faa;
done;

mv *.faa ORF/

# We will move to a new directory (cluster) to begin clustering our assemblies
#!/bin/Bash
mkdir cluster
#!/bin/Bash
mv *prune*.fasta cluster
#!/bin/Bash
cd cluster

# We want to generate single concatenated file with all assemblies for later clustering
# First we delete all the ">" within a file
#!/bin/Bash
for file1 in *fasta;
do
sed -i '/^>/ s/>.*//' $file1;
done

# Then rename the top header in the file with >'filename'
#!/bin/Bash
for file1 in *fasta;
do
file2=${file1%%.fasta}.txt
awk 'NR==1{$0=$0">"FILENAME}1' $file1 > $file2;
done

# Finally we create a single text document (asemblies) that has all assemblies as a single read each
#!/bin/Bash
for file1 in *txt;
do
cat *.txt > assemblies;
done

#We remove extranous symbols from the assemblies file and remove the modified fasta files
#!/bin/Bash
sed -e '/^[^>]/s/[^ATGCatgc]/N/g' assemblies > assemblies_final
rm *fasta
cd cluster
```

Script to use Local BLAST on multiple databases

```
#!/bin/Bash
#We append the file name to each read and concatenate all reads
```

```
for file1 in *.fastq;
do
file2=${file1%%.fastq}.fasta
awk '/>/{sub(">","&"FILENAME"_");sub(/\.faa/,x)}1' $file1 > $file2;
done

#We convert each assembly into a BLAST database
for file1 in *.fasta;
do
file2=${file1%%.fasta}.db
makeblastdb -in $file1 -parse_seqids -out $file2 -title "$file2" -dbtype nucl;
done

#We can now blast our queries into each database and output in a tabular format
for file1 in *.fasta;
do
database=${file1%%.fasta}.db
tblastn -db "$database" -query query.fa -out "$database".csv -outfmt "6 qseqid sseqid pident qcovs
evalue"
done
```

Script to Local BLAST multiple databases for only the top hit per query

```
#!/bin/Bash
#We append the file name to each read and concatenate all reads
for file1 in *.fastq;
do
file2=${file1%%.fastq}.fasta
awk '/>/{sub(">","&"FILENAME"_");sub(/\.faa/,x)}1' $file1 > $file2;
done

#We convert each assembly into a BLAST database
for file1 in *.fasta;
do
file2=${file1%%.fasta}.db
makeblastdb -in $file1 -parse_seqids -out $file2 -title "$file2" -dbtype nucl;
done

#We can now blast our queries into each database and output in a tabular format
for file1 in *.fasta;
do
database=${file1%%.fasta}.db
tblastn -db "$database" -query query.fa -out "$database".csv -outfmt "6 qseqid sseqid pident qcovs
evalue" -max_target_seqs 1
done
```

Scripts to cluster amino acid sequences with UCLUST

```
#!/bin/Bash
```

```
mkdir cluster

#Then we copy the .faa files from the Prokka directories to the cluster subdirectory
#!/bin/Bash
for subdir in *;
do cp $subdir/*.faa $subdir.faa;
done;
mv *.faa cluster
cd cluster;

#!/bin/Bash
#We append the file name to each read and concatenate all reads
for file1 in *.faa;
do
file2=${file1%%.faa}.fasta
awk '/>/{sub(">","&"FILENAME"_");sub(/\.faa/,x)}1' $file1 > $file2;

cat *.fasta > reads.fasta;
done

#We sort reads by lenght
#!/bin/Bash
/home/cesar/apps/bbmap/sortbyname.sh in=reads.fasta out=sorted.fasta length descending
rm sorted.fasta
/home/cesar/apps/bbmap/sortbyname.sh in=reads.fasta out=sorted.fasta length descending -
ignorejunk

#Then we copy the .faa files from the Prokka directories to the cluster subdirectory

#We cluster the reads
mkdir cluster_dir

#!/bin/Bash
/home/cesar/apps/usearch11.0.667_i86linux32 -cluster_fast sorted.fasta -id 0.8 -clusters cluster_dir/c_;

#We process the cluster files for ease of use
#!/bin/Bash
for file1 in cluster_dir/*;
do
mv "$file1" "$file1.txt";
done;

for file1 in cluster_dir/*.txt
do
mv "$file1" cluster_dir/"$(head -1 "$file1").fasta";
done;

#!/bin/Bash
```

```
#Clean up extra files and file names
cd cluster_dir
rm *.txt;
done;
```

Script to compare multiple nucleotide assemblies via sourmash

```
#!/bin/Bash
#activate sourmash. If this doesn't work, run this step manually
conda activate smash

#make a directory to place the sig files
mkdir sigs

#convert each assembly into a sig file
for file1 in *.fa; do
file2=${file1%%.fa}.sig
sourmash compute --scaled 1000 -k 31 $file1 -o sigs/$file2;
done

#index all sig files into one file
sourmash index -k 31 assemblydb sigs/*.sig

#search your reference in your index
sourmash search reference.sig assemblydb.sbt.json -n 20

#compare all sigs to one another
sourmash compare sigs/* -o assembly_cmp

#make a tree and heatmap of your comparison
sourmash plot --pdf --labels assembly_cmp

done
```

This script changes file extensions

```
#!/bin/Bash
#First we want to convert each assembly into a BLAST database
#Convert *.fasta file into a BLAST database
#!/bin/Bash
for f in *.fasta; do
mv -- "$f" "${f%.fasta}.fa";
done
```

This script appends the file name to each read in a FASTA file

```
#!/bin/Bash
#We append the file name to each read and concatenate all reads
for file1 in *.faa;
do
file2=${file1%%.faa}.fasta
```

```
awk '/>/{sub(">","&"FILENAME"_");sub(/\.faa/,x)}1' $file1 > $file2;
done
```

This script changes a multi-read FASTA file into a single-read file and concatenates multiple files into a one

```
# We want to generate single concatenated file with all assemblies for later clustering
#Each assembly will be represented as a single read
# First we delete all the ">" within a file
#!/bin/Bash
for file1 in *fasta;
do
sed -i '/^>/ s/>.*//' $file1;
done

# Then rename the top header in the file with >'filename'
#!/bin/Bash
for file1 in *fasta;
do
file2=${file1%%.fasta}.txt
awk 'NR==1{$0=$0">"FILENAME}1' $file1 > $file2;
done

# Finally we create a single text document (asemblies) that has all assemblies as a single read each
#!/bin/Bash
for file1 in *txt;
do
cat *.txt > assemblies;
done

#the 'assemblies' file has all assemblies as a single read each
```

This script moves files with an extension

```
#Use this to bin output from a pipeline (e.g., .faa or .tsv files)
# Move TSV files to directory
#!/bin/Bash
mkdir faa

# Copy TSV files for all annotation
#!/bin/Bash
for subdir in *;
do cp $subdir/*.faa $subdir.faa;
done;

mv *.faa faa/
```

This script removes strings from file names

```
#this removes words from the file name
for f in *.fasta; do
```

```
   mv -- "$f" "${f/_R1_output/}"
done
```

This script renames and removes portions of file names

```
#we want to get rid of “single_” in the file names
#!/bin/Bash
makedir renamed
for f in *.fastq; do
   mv -- "$f" renamed/"${f/single_/}"
done
```

This script removes characters from file names

```
#Replace 000 with characters to remove
#!/bin/Bash
for i in *000.tga
do
   mv "$i" "`echo $i | sed 's/000//'`"
done
```

These commands deletes strings within a file

```
#!/bin/Bash
#This command deletes a string inside a file
#single word
sed -i -e 's/goodbye//g' filename

#multiple words
sed -i -e 's/\(goodbye\|hello\|test\|download\)//g' filename

#everything before space
sed -i -e 's/[^ ]* //' *.txt

#single word in multiple files
sed -i 's/.fasta//g' *.txt
```

This command lists the headers of all files with a given extension

```
#!/bin/Bash
grep -e ">" *.fasta;
sort -u *.fasta
done
```

This command splits reads in a single FASTA file into individual files

```
awk -F "|" '/^>/ {close(F) ; F = $1".fasta"} {print >> F}' yourfile.fa
```

This command removes repeated strings in a file

```
#use this to remove repeated names on a gene list
#!/bin/Bash
awk '!seen[$0]++' plasmid_ref.txt > plasmid_ref_list.txt;
done
```

REFERENCE LIST

1. Thomas-White K, Brady M, Wolfe AJ, Mueller ER. The bladder is not sterile: History and current discoveries on the urinary microbiome. *Curr Bladder Dysfunct Rep*. 2016;11(1):18-24. doi:10.1007/s11884-016-0345-8

2. Roberts W. On the occurrence of micro-organisms in fresh urine. *Br Med J*. 1881;2(1085):623-625. doi:10.1136/bmj.2.1085.623

3. Wolfe AJ, Toh E, Shibata N, et al. Evidence of uncultivated bacteria in the adult female bladder. *J Clin Microbiol*. 2012;50(4):1376-1383. doi:10.1128/JCM.05852-11

4. KASS EH. Asymptomatic infections of the urinary tract. *Trans Assoc Am Physicians*. 1956;69:56-64.

5. Kass EH. Bacteriuria and the Diagnosis of Infections of the Urinary Tract: With Observations on the Use of Methionine as a Urinary Antiseptic. *AMA Arch Intern Med*. 1957;100(5):709-714. doi:10.1001/archinte.1957.00260110025004

6. Stamm WE, Counts GW, Running KR, Fihn S, Turck M, Holmes KK. Diagnosis of coliform infection in acutely dysuric women. *N Engl J Med*. 1982;307(8):463-468. doi:10.1056/NEJM198208193070802

7. Stark RP, Maki DG. Bacteriuria in the Catheterized Patient. *N Engl J Med*. 1984;311(9):560-564. doi:10.1056/nejm198408303110903

8. Lipsky BA, Ireton RC, Fihn SD, Hackett R, Berger RE. Diagnosis of bacteriuria in men: specimen collection and culture interpretation. *J Infect Dis*. 1987;155(5):847-854. doi:10.1093/infdis/155.5.847

9. Price TK, Dune T, Hilt EE, et al. The Clinical Urine Culture: Enhanced Techniques Improve Detection of Clinically Relevant Microorganisms. *J Clin Microbiol*. 2016;54(5):1216-1222. doi:10.1128/JCM.00044-16

10. Wolfe AJ, Brubaker L. "Sterile Urine" and the Presence of Bacteria. *Eur Urol*. 2015;68(2):173-174. doi:10.1016/j.eururo.2015.02.041

11. Schmiemann G, Kniehl E, Gebhardt K, Matejczyk MM, Hummers-Pradier E. The diagnosis of urinary tract infection: a systematic review. *Dtsch Arztebl Int*. 2010;107(21):361-367. doi:10.3238/arztebl.2010.0361

12. Maskell RM. The natural history of urinary tract infection in women. *Med Hypotheses*. 2010;74(5):802-806. doi:10.1016/j.mehy.2009.12.011

13. Seal D V, Cuthbert EH. Doubtful significance of fastidious bacteriuria in the urethral syndrome. *Lancet (London, England)*. 1982;1(8263):115. doi:10.1016/s0140-6736(82)90260-4

14. Brumfitt W, Hamilton-Miller JM, Ludlam H, Gooding A. Lactobacilli do not cause frequency and dysuria syndrome. *Lancet (London, England)*. 1981;2(8243):393-396. doi:10.1016/s0140-6736(81)90833-3

15. Drabu YJ, Sanderson PJ. URINE CULTURE IN URETHRAL SYNDROME. *Lancet*. 1980;315(8158):37-38. doi:10.1016/S0140-6736(80)90569-

16. Hooton TM, Roberts PL, Cox ME, Stapleton AE. Voided Midstream Urine Culture and Acute Cystitis in Premenopausal Women. *N Engl J Med*. 2013;369(20):1883-1891. doi:10.1056/nejmoa1302186

17. Hilt EE, McKinley K, Pearce MM, et al. Urine is not sterile: use of enhanced urine culture techniques to detect resident bacterial flora in the adult female bladder. *J Clin Microbiol*. 2014;52(3):871-876. doi:10.1128/JCM.02876-13

18. Thomas-White K, Forster SC, Kumar N, et al. Culturing of female bladder bacteria reveals an interconnected urogenital microbiota. *Nat Commun*. 2018;9(1):1557. doi:10.1038/s41467-018-03968-5

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

20. Garretto A, Miller-Ensminger T, Wolfe AJ, Putonti C. Bacteriophages of the lower urinary tract. *Nat Rev Urol*. Published online May 9, 2019:1. doi:10.1038/s41585-019-0192-4

21. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. *Nat Biotechnol*. 2017;35(9):833-844. doi:10.1038/nbt.3935

22. Sender R, Fuchs S, Milo R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol*. 2016;14(8). doi:10.1371/journal.pbio.1002533

23. Lloyd-Price J, Abu-Ali G, Huttenhower C. The healthy human microbiome. *Genome Med*. 2016;8(1):51. doi:10.1186/s13073-016-0307-y

24. Huttenhower C, Gevers D, Knight R, et al. Structure, function and diversity of the healthy human microbiome. *Nature*. 2012;486(7402):207-214. doi:10.1038/nature11234

25. Nelson DE, Dong Q, Van der Pol B, et al. Bacterial communities of the coronal sulcus and distal urethra of adolescent males. *PLoS One*. 2012;7(5):e36298. doi:10.1371/journal.pone.0036298

26. Pearce MM, Hilt EE, Rosenfeld AB, et al. The female urinary microbiome: A comparison of women with and without urgency urinary incontinence. *MBio*. 2014;5(4):1283-1297. doi:10.1128/mBio.01283-14

27. Cai T, Mazzoli S, Mondaini N, et al. The Role of Asymptomatic Bacteriuria in Young Women With Recurrent Urinary Tract Infections: To Treat or Not to Treat? *Clin Infect Dis*. 2012;55(6):771-777. doi:10.1093/cid/cis534

28. Andrea Garretto, Taylor Miller-Ensminger, Adriana Ene, Zubia Merchant, Aashaka Shah, Athina Gerodias, Anthony Biancofiori, Stacey Canchola, Stephanie Canchola, Emanuel Castillo, Tasnim Chowdhury, Nikita Gandhi, Sarah Hamilton, Kyla Hatton, Syed Hyder, Kot AJW and CP. Genomic Survey of *E. coli* from the Bladders of Women with and without Lower Urinary Tract Symptoms. *Front Microbiol Sect Infect Dis*. Published online 2020.

29. Malki K, Shapiro JW, Price TK, et al. Genomes of Gardnerella Strains Reveal an Abundance of Prophages within the Bladder Microbiome. *PLoS One*. 2016;11(11):e0166757. doi:10.1371/journal.pone.0166757

30. Miller-Ensminger T, Mormando R, Maskeri L, Shapiro JW, Wolfe AJ, Putonti C. Introducing Lu-1, a novel lactobacillus jensenii phage abundant in the urogenital tract. *PLoS One*. 2020;15(6). doi:10.1371/journal.pone.0234159

31. Hartmann KE, McPheeters ML, Biller DH, et al. Treatment of overactive bladder in women. *Evid Rep Technol Assess (Full Rep)*. 2009;(187):1-120, v.

32. Brubaker L, Nager CW, Richter HE, et al. Urinary bacteria in adult women with urgency urinary incontinence. *Int Urogynecol J*. 2014;25(9):1179-1184. doi:10.1007/s00192-013-2325-2
33. Brubaker L, Wolfe AJ. The female urinary microbiota, urinary health and common urinary disorders. *Ann Transl Med*. 2017;5(2):34. doi:10.21037/atm.2016.11.62

34. Brubaker L, Wolfe AJ. The new world of the urinary microbiota in women. *Am J Obstet Gynecol*. 2015;213(5):644-649. doi:10.1016/j.ajog.2015.05.032

35. Brubaker L, Wolfe A. The urinary microbiota: a paradigm shift for bladder disorders? *Curr Opin Obstet Gynecol*. 2016;28(5):407-412. doi:10.1097/GCO.0000000000000298

36. Upadhyayula S, Kambalapalli M, Asmar BI. Staphylococcus epidermidis Urinary Tract Infection in an Infant . *Case Rep Infect Dis*. 2012;2012:1-2. doi:10.1155/2012/983153

37. Leclercq SY, Sullivan MJ, Ipe DS, Smith JP, Cripps AW, Ulett GC. Pathogenesis of Streptococcus urinary tract infection depends on bacterial strain and β-hemolysin/cytolysin that mediates cytotoxicity, cytokine synthesis, inflammation and virulence. *Sci Rep*. 2016;6. doi:10.1038/srep29000

38. Kalyoussef S, Nieves E, Dinerman E, et al. Lactobacillus Proteins Are Associated with the Bactericidal Activity against *E. coli* of Female Genital Tract Secretions. *PLoS One*. 2012;7(11). doi:10.1371/journal.pone.0049506

39. Price TK. Assessing the termporal dynamics of the lower urinary tract microbiota and the effects of lifestyle. Published online 2019.

40. Price TK, Wolff B, Halverson T, et al. Temporal Dynamics of the Adult Female Lower Urinary Tract Microbiota Downloaded from. Published online 2020. doi:10.1128/mBio

41. Brubaker L, Wolfe AJ. Microbiota in 2016: Associating infection and incontinence with the female urinary microbiota. *Nat Rev Urol*. 2017;14(2):72-74. doi:10.1038/nrurol.2016.262

42. Falagas ME, Betsi GI, Tokas T, Athanasiou S. Probiotics for Prevention of Recurrent Urinary Tract Infections in Women. *Drugs*. 2006;66(9):1253-1261. doi:10.2165/00003495-200666090-00007

43. Thomas-White KJ, Hilt EE, Fok C, et al. Incontinence medication response relates to the female urinary microbiota. *Int Urogynecol J*. 2016;27(5):723-733. doi:10.1007/s00192-015-2847-x

44. Schwaderer AL, Wolfe AJ. The association between bacteria and urinary stones. *Ann Transl Med*. 2017;5(2):32. doi:10.21037/atm.2016.11.73

45. Price TK, Dune T, Hilt EE, et al. The Clinical Urine Culture: Enhanced Techniques Improve Detection of Clinically Relevant Microorganisms. Forbes BA, ed. *J Clin Microbiol*. 2016;54(5):1216-1222. doi:10.1128/JCM.00044-16

46. Larsen P, Dynesen P, Nielsen KL, Frimodt-Møller N. Faecal Escherichia coli from patients with *E. coli* urinary tract infection and healthy controls who have never had a urinary tract infection. *J Med Microbiol*. 2014;63(4):582-589. doi:10.1099/jmm.0.068783-0

47. Chibani-Chennoufi S, Bruttin A, Dillmann M-L, Brussow H. Phage-Host Interaction: an Ecological Perspective. *J Bacteriol*. 2004;186(12):3677-3686. doi:10.1128/JB.186.12.3677-3686.2004

48. Salmond GPC, Fineran PC. A century of the phage: past, present and future. *Nat Rev Microbiol*. 2015;13(12):777-786. doi:10.1038/nrmicro3564

49. Koskella B, Meaden S. Understanding Bacteriophage Specificity in Natural Microbial Communities. *Viruses*. 2013;5(3):806-823. doi:10.3390/v5030806

50. Koskella B, Brockhurst MA. Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol Rev*. 2014;38(5):916-931. doi:10.1111/1574-6976.12072

51. Casjens SR, Hendrix RW. Bacteriophage lambda: Early pioneer and still relevant. *Virology*. 2015;479-480:310-330. doi:10.1016/j.virol.2015.02.010

52. Paul JH. Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas? *ISME J*. 2008;2(6):579-589. doi:10.1038/ismej.2008.35

53. Baugher JL, Durmaz E, Klaenhammer TR. Spontaneously induced prophages in Lactobacillus gasseri contribute to horizontal gene transfer. *Appl Environ Microbiol*. 2014;80(11):3508-3517. doi:10.1128/AEM.04092-13

54. Lood R, Ertürk G, Mattiasson B. Revisiting Antibiotic Resistance Spreading in Wastewater Treatment Plants – Bacteriophages as a Much Neglected Potential Transmission Vehicle. *Front Microbiol*. 2017;8. doi:10.3389/fmicb.2017.02298

55. Harrison E, Brockhurst MA. Ecological and Evolutionary Benefits of Temperate Phage: What Does or Doesn't Kill You Makes You Stronger. *BioEssays*. 2017;39(12):1700112. doi:10.1002/bies.201700112

56. Waldor MK, Mekalanos JJ. Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science*. 1996;272(5270):1910-1914. doi:10.1126/SCIENCE.272.5270.1910

57. Payet JP, Suttle CA. To kill or not to kill: The balance between lytic and lysogenic viral infection is driven by trophic status. *Limnol Oceanogr*. 2013;58(2):465-474. doi:10.4319/lo.2013.58.2.0465

58. Fang Y, Mercer RG, McMullen LM, Gänzle MG. Induction of Shiga Toxin-Encoding Prophage by Abiotic Environmental Stress in Food. Schaffner DW, ed. *Appl Environ Microbiol*. 2017;83(19). doi:10.1128/AEM.01378-17

59. Drexler H. TRANSDUCTION BY BACTERIOPHAGE T1. *Proc Natl Acad Sci*. 1970;66(4):1083-1088. doi:10.1073/pnas.66.4.1083

60. Scott JR. Superinfection immunity and prophage repression in phage P1. *Virology*. 1975;65(1):173-178. Accessed May 24, 2019. http://www.ncbi.nlm.nih.gov/pubmed/1096454

61. Norman JM, Handley SA, Baldridge MT, et al. Disease-Specific Alterations in the Enteric Virome in Inflammatory Bowel Disease. *Cell*. 2015;160(3):447-460. doi:10.1016/j.cell.2015.01.002

62. Ly M, Abeles SR, Boehm TK, et al. Altered oral viral ecology in association with periodontal disease. *MBio*. 2014;5(3):e01133-14. doi:10.1128/mBio.01133-14

63. Navarro F, Muniesa M. Phages in the Human Body. *Front Microbiol*. 2017;8:566. doi:10.3389/fmicb.2017.00566

64. Manrique P, Bolduc B, Walk ST, van der Oost J, de Vos WM, Young MJ. Healthy human gut phageome. *Proc Natl Acad Sci*. 2016;113(37):10400-10405. doi:10.1073/pnas.1601060113

65. Ma Y, You X, Mai G, Tokuyasu T, Liu C. A human gut phage catalog correlates the gut phageome with type 2 diabetes. *Microbiome*. 2018;6(1):24. doi:10.1186/s40168-018-0410-y

66. Adriaenssens EM, Cowan DA. Using Signature Genes as Tools To Assess Environmental Viral Ecology and Diversity. Wood TK, ed. *Appl Environ Microbiol*. 2014;80(15):4470-4480. doi:10.1128/AEM.00878-14

67. Rani A, Ranjan R, McGee HS, et al. A diverse virome in kidney transplant patients contains multiple viral subtypes with distinct polymorphisms. *Sci Rep*. 2016;6(1):33327. doi:10.1038/srep33327

68. Malki K, Sible E, Cooper A, et al. Seven Bacteriophages Isolated from the Female Urinary Microbiota. *Genome Announc*. 2016;4(6). doi:10.1128/genomeA.01003-16

69. Garretto A. Exploring associations between lysogeny and host abundance. Published online 2019. Thesis

70. M.F. d'Herelle presented by MR. On_an_invisible_microbe_antagonistic_to_dysentery_bacilli. *Comptes Rendus Acad des Sci 1917*. 2011;165:373–5. doi:10.4161/bact.1.1.14941

71. Brown-Jaque M, Muniesa M, Navarro F. Bacteriophages in clinical samples can interfere with microbiological diagnostic tools. *Sci Rep*. 2016;6(1):33000. doi:10.1038/srep33000

72. Moustafa A, Li W, Singh H, et al. Microbial metagenome of urinary tract infection. *Sci Rep*. 2018;8(1):4333. doi:10.1038/s41598-018-22660-8

73. Garretto A, Thomas-White K, Wolfe AJ, Putonti C. Detecting viral genomes in the female urinary microbiome. *J Gen Virol*. 2018;99(8):1141-1146. doi:10.1099/jgv.0.001097

74. Kim M-S, Bae J-W. Lysogeny is prevalent and widely distributed in the murine gut microbiota. *ISME J*. 2018;12(4):1127-1141. doi:10.1038/s41396-018-0061-9

75. Miller-Ensminger T, Garretto A, Brenner J, et al. Bacteriophages of the Urinary Microbiome. *J Bacteriol*. 2018;200(7). doi:10.1128/JB.00738-17

76. Thannesberger J, Hellinger H-J, Klymiuk I, et al. Viruses comprise an extensive pool of mobile genetic elements in eukaryote cell cultures and human clinical samples. *FASEB J*. 2017;31(5):1987-2000. doi:10.1096/fj.201601168R

77. Yilmaz S, Allgaier M, Hugenholtz P. Multiple displacement amplification compromises quantitative analysis of metagenomes. *Nat Methods*. 2010;7(12):943-944. doi:10.1038/nmeth1210-943

78. Price TK, Mehrtash A, Kalesinskas L, et al. Genome sequences and annotation of two urinary isolates of *E. coli*. *Stand Genomic Sci*. 2016;11(1):79. doi:10.1186/s40793-016-0202-6

79. Barr JJ, Auro R, Furlan M, et al. Bacteriophage adhering to mucus provide a non-host-derived immunity. *Proc Natl Acad Sci*. 2013;110(26):10771-10776. doi:10.1073/pnas.1305923110

80. Shan J, Ramachandran A, Thanki AM, Vukusic FBI, Barylski J, Clokie MRJ. Bacteriophages are more virulent to bacteria with human cells than they are in bacterial culture; insights from HT-29 cells. *Sci Rep*. 2018;8(1):5091. doi:10.1038/s41598-018-23418-y

81. Duerkop BA, Hooper L V. Resident viruses and their interactions with the immune system. *Nat Immunol*. 2013;14(7):654-659. doi:10.1038/ni.2614

82. Dabrowska K, Opolski A, Wietrzyk J, et al. Antitumor activity of bacteriophages in murine experimental cancer models caused possibly by inhibition of beta3 integrin signaling pathway. *Acta Virol*. 2004;48(4):241-248. Accessed May 24, 2019. http://www.ncbi.nlm.nih.gov/pubmed/15745047

83. Huddleston JR. Horizontal gene transfer in the human gastrointestinal tract: potential spread of antibiotic resistance genes. *Infect Drug Resist*. 2014;7:167-176. doi:10.2147/IDR.S48820

84. Roberts AP, Kreth J. The impact of horizontal gene transfer on the adaptive ability of the human oral microbiome. *Front Cell Infect Microbiol*. 2014;4:124. doi:10.3389/fcimb.2014.00124

85. Howard-Varona C, Hargreaves KR, Solonenko NE, et al. Multiple mechanisms drive phage infection efficiency in nearly identical hosts. *ISME J*. 2018;12(6):1605-1618. doi:10.1038/s41396-018-0099-8

86. Lee DS, Lee SJ, Choe HS, Giacobbe DR. Community-Acquired Urinary Tract Infection by Escherichia coli in the Era of Antibiotic Resistance. *Biomed Res Int*. 2018;2018. doi:10.1155/2018/7656752

87. Gupta K, Hooton TM, Stamm WE. Increasing antimicrobial resistance and the management of uncomplicated community-acquired urinary tract infections. *Ann Intern Med*. 2001;135(1):41-50. doi:10.7326/0003-4819-135-1-200107030-00012

88. Foxman B. Urinary Tract Infection Syndromes. *Infect Dis Clin North Am*. 2014;28(1):1-13. doi:10.1016/j.idc.2013.09.003

89. Johnson TJ, Elnekave E, Miller EA, et al. Phylogenomic Analysis of Extraintestinal Pathogenic *Escherichia coli* Sequence Type 1193, an Emerging Multidrug-Resistant Clonal Group. *Antimicrob Agents Chemother*. 2018;63(1). doi:10.1128/AAC.01913-18

90. Johnson JR, Porter S, Johnston B, et al. Host Characteristics and Bacterial Traits Predict Experimental Virulence for Escherichia coli Bloodstream Isolates From Patients With Urosepsis. *Open Forum Infect Dis*. 2015;2(3):ofv083. doi:10.1093/ofid/ofv083

91. Luo Y, Ma Y, Zhao Q, et al. Similarity and Divergence of Phylogenies, Antimicrobial Susceptibilities, and Virulence Factor Profiles of Escherichia coli Isolates Causing Recurrent Urinary Tract Infections That Persist or Result from Reinfection. *J Clin Microbiol*. 2012;50(12):4002-4007. doi:10.1128/JCM.02086-12

92. Moriel DG, Tan L, Goh KGK, et al. A Novel Protective Vaccine Antigen from the Core Escherichia coli Genome. *mSphere*. 2016;1(6). doi:10.1128/msphere.00326-16

93. SuBashchandrabose S, Hazen TH, Brumbaugh AR, et al. Host-specific induction of Escherichia coli fitness genes during human urinary tract infection. *Proc Natl Acad Sci U S A*. 2014;111(51):18327-18332. doi:10.1073/pnas.1415959112

94. SuBashchandrabose S, Mobley HLT. Virulence and Fitness Determinants of Uropathogenic Escherichia coli. *Microbiol Spectr*. 2015;3(4). doi:10.1128/microbiolspec.uti-0015-2012

95. Yamamoto S, Tsukamoto T, Terai A, Kurazono H, Takeda Y, Yoshida O. Genetic evidence supporting the fecal-perineal-urethral hypothesis in cystitis caused by Escherichia coli. *J Urol*. 1997;157(3):1127-1129. Accessed August 10, 2019. http://www.ncbi.nlm.nih.gov/pubmed/9072556

96. Young R. Bacteriophage lysis: Mechanism and regulation. *Microbiol Rev*. 1992;56(3):430-481.

97. Hendrix RW, Smith MC, Burns RN, Ford ME, Hatfull GF. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc Natl Acad Sci U S A*. 1999;96(5):2192-2197. doi:10.1073/PNAS.96.5.2192

98. Thomason LC, Costantino N, Court DL. *E. coli* Genome Manipulation by P1 Transduction. In: *Current Protocols in Molecular Biology*. Vol Chapter 1. John Wiley & Sons, Inc.; 2007:1.17.1-1.17.8. doi:10.1002/0471142727.mb0117s79

99. Ikeda H, Tomizawa J ichi. Transducing fragments in generalized transduction by phage P1: I. Molecular origin of the fragments. *J Mol Biol*. 1965;14(1):85-109. doi:10.1016/S0022-2836(65)80232-7

100. Yarmolinsky MB. Bacteriophage P1 in retrospect and in prospect. *J Bacteriol*. 2004;186(21):7025-7028. doi:10.1128/JB.186.21.7025-7028.2004

101. Blount ZD. The unexhausted potential of *E. coli*. *Elife*. 2015;4. doi:10.7554/eLife.05826

102. Lee PS, Lee KH. Escherichia coli - A Model System That Benefits from and Contributes to the Evolution of Proteomics. *Biotechnol Bioeng*. 2003;84(7):801-814. doi:10.1002/bit.10848

103. Kitagawa M, Ara T, Arifuzzaman M, et al. Complete set of ORF clones of Escherichia coli ASKA library (A complete set of *E. coli* K-12 ORF archive): unique resources for biological research. *DNA Res*. 2005;12(5):291-299. doi:10.1093/dnares/dsi012

104. Baba T, Ara T, Hasegawa M, et al. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: The Keio collection. *Mol Syst Biol*. 2006;2. doi:10.1038/msb4100050

105. Fineran PC, Blower TR, Foulds IJ, Humphreys DP, Lilley KS, Salmond GPC. The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. *Proc Natl Acad Sci U S A*. 2009;106(3):894-899. doi:10.1073/pnas.0808832106

106. Klaenhammer TR. Genetic Characterization of Multiple Mechanisms of Phage Defense from a Prototype Phage-Insensitive Strain, Lactococcus lactis ME2. *J Dairy Sci*. 1989;72(12):3429-3443. doi:10.3168/jds.S0022-0302(89)79505-9

107. Riede I, Eschbach M-L. *Evidence That TraT Interacts with OmpA of Escherichia Coli TraTprotein OmpA Protein Protein Interaction*. Vol 205.; 1986. Accessed June 17, 2019. https://febs.onlinelibrary.wiley.com/doi/pdf/10.1016/0014-5793%2886%2980905-X

108. Van Houte S, Buckling A, Westra ER. Evolutionary Ecology of Prokaryotic Immune Mechanisms. Published online 2016. doi:10.1128/MMBR.00011-16

109. Ornellas EP, Stocker BAD. Relation of lipopolysaccharide character to P1 sensitivity in Salmonella typhimurium. *Virology*. 1974;60(2):491-502. doi:10.1016/0042-6822(74)90343-2

110. Grahn AM, Haase J, Lanka E, Bamford DH. Assembly of a functional phage PRD1 receptor depends on 11 genes of the IncP plasmid mating pair formation complex. *J Bacteriol*. 1997;179(15):4733-4740. doi:10.1128/jb.179.15.4733-4740.1997

111. Vasu K, Nagaraja V. Diverse Functions of Restriction-Modification Systems in Addition to Cellular Defense. *Microbiol Mol Biol Rev*. 2013;77(1):53-72. doi:10.1128/mmbr.00044-12

112. Díez-Villaseñor C, Almendros C, García-Martínez J, Mojica FJM. Diversity of CRISPR loci in Escherichia coli. *Microbiology*. 2010;156(5):1351-1361. doi:10.1099/mic.0.036046-0

113. Mnif B, Vimont S, Boyd A, et al. Molecular characterization of addiction systems of plasmids encoding extended-spectrum beta-lactamases in Escherichia coli. *J Antimicrob Chemother*. 2010;65(8):1599-1603. doi:10.1093/jac/dkq181

114. Schaufler K, Wieler LH, Semmler T, Ewers C, Guenther S. ESBL-plasmids carrying toxin-antitoxin systems can be "cured" of wild-type Escherichia coli using a heat technique. *Gut Pathog*. 2013;5(1):34. doi:10.1186/1757-4749-5-34

115. Song S, Wood TK. Post-segregational Killing and Phage Inhibition Are Not Mediated by Cell Death Through Toxin/Antitoxin Systems. *Front Microbiol*. 2018;9:814. doi:10.3389/fmicb.2018.00814

116. Engelberg-Kulka H, Reches M, Narasimhan S, et al. rexB of bacteriophage lambda is an anti-cell death gene. *Proc Natl Acad Sci U S A*. 1998;95(26):15481-15486. doi:10.1073/pnas.95.26.15481

117. Kliem M, Dreiseikelmann B. The superimmunity gene sim of bacteriophage P1 causes superinfection exclusion. *Virology*. 1989;171(2):350-355. doi:10.1016/0042-6822(89)90602-8

118. Jalasvuori M, Friman VP, Nieminen A, Bamford JKH, Buckling A. Bacteriophage selection against a plasmid-encoded sex apparatus leads to the loss of antibioticresistance plasmids. *Biol Lett*. 2011;7(6):902-905. doi:10.1098/rsbl.2011.0384

119. Harrison E, Truman J, Wright R, Spiers AJ, Paterson S, Brockhurst MA. Plasmid carriage can limit bacteria-phage coevolution. *Biol Lett*. 2015;11(8). doi:10.1098/rsbl.2015.0361

120. Shintani M, Sanchez ZK, Kimbara K. Genomics of microbial plasmids: Classification and identification based on replication and transfer systems and host taxonomy. *Front Microbiol*. 2015;6(MAR). doi:10.3389/fmicb.2015.00242

121. Osborn AM, da Silva Tatley FM, Steyn LM, Pickup RW, Saunders JR. Mosaic plasmids and mosaic replicons: Evolutionary lessons from the analysis of genetic diversity in IncFII-related replicons. *Microbiology*. 2000;146(9):2267-2275. doi:10.1099/00221287-146-9-2267

122. Lili LN, Britton NF, Feil EJ. The persistence of parasitic plasmids. *Genetics*. 2007;177(1):399-405. doi:10.1534/genetics.107.077420

123. Hughes C, Bauer E, Roberts AP. Spread of R-plasmids among Escherichia coli causing urinary tract infections. *Antimicrob Agents Chemother*. 1981;20(4):496-502. doi:10.1128/aac.20.4.496

124. Wu AY, Kamruzzaman M, Iredell JR. Specialised functions of two common plasmid mediated toxin-Antitoxin systems, ccdAB and pemIK, in Enterobacteriaceae. *PLoS One*. 2020;15(6 June). doi:10.1371/journal.pone.0230652

125. Sukupolvi1 S, David O'connor2 C. *TraT Lipoprotein, a Plasmid-Specified Mediator of Interactions between Gram-Negative Bacteria and Their Environment*.; 1990. Accessed June 17, 2019. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC372785/pdf/microrev00039-0011.pdf

126. Carattoli A, Bertini A, Villa L, Falbo V, Hopkins KL, Threlfall EJ. Identification of plasmids by PCR-based replicon typing. *J Microbiol Methods*. 2005;63(3):219-228. doi:10.1016/j.mimet.2005.03.018

127. Walker A. Welcome to the plasmidome. *Nat Rev Microbiol*. 2012;10(6):379. doi:10.1038/nrmicro2804

128. Brolund A, Franzén O, Melefors Ö, Tegmark-Wisell K, Sandegren L. Plasmidome-Analysis of ESBL-Producing Escherichia coli Using Conventional Typing and High-Throughput Sequencing. *PLoS One*. 2013;8(6). doi:10.1371/journal.pone.0065793

129. Brown Kav A, Rozov R, Bogumil D, et al. Unravelling plasmidome distribution and interaction with its hosting microbiome. *Environ Microbiol*. 2020;22(1):32-44. doi:10.1111/1462-2920.14813

130. Delaney S, Murphy R, Walsh F. A Comparison of Methods for the Extraction of Plasmids Capable of Conferring Antibiotic Resistance in a Human Pathogen From Complex Broiler Cecal Samples. *Front Microbiol*. 2018;9(AUG):1731. doi:10.3389/fmicb.2018.01731

131. Suzuki Y, Nishijima S, Furuta Y, et al. Long-read metagenomic exploration of extrachromosomal mobile genetic elements in the human gut. *Microbiome*. 2019;7(1):119. doi:10.1186/s40168-019-0737-z

132. Steinig EJ, Duchene S, Robinson DA, et al. Evolution and global transmission of a multidrug-resistant, community-associated methicillin-resistant staphylococcus aureus lineage from the Indian subcontinent. *MBio*. 2019;10(6). doi:10.1128/mBio.01105-19

133. Harrison E, Jamie Wood A, Dytham C, et al. Bacteriophages limit the existence conditions for conjugative plasmids. *MBio*. 2015;6(3). doi:10.1128/mBio.00586-15

134. Cox KEL, Schildbach JF. Sequence of the R1 plasmid and comparison to F and R100. *Plasmid*. 2017;91:53-60. doi:10.1016/J.PLASMID.2017.03.007

135. Froggatt JW, Johnston JL, Galetto DW, Archer GL. Antimicrobial resistance in nosocomial isolates of Staphylococcus haemolyticus. *Antimicrob Agents Chemother*. 1989;33(4):460-466. doi:10.1128/AAC.33.4.460

136. Fernandez-Lopez R, de Toro M, Moncalian G, Garcillan-Barcia MP, de la Cruz F. Comparative genomics of the conjugation region of F-like plasmids: Five shades of F. *Front Mol Biosci*. 2016;3(NOV). doi:10.3389/fmolb.2016.00071

137. Goodson M, Rowbury RJ. *Altered Phage P1 Attachment to Strains of Escherichia Coli Carrying the Piasmid ColV,I-K94*. Vol 68.; 1987.

138. Jalasvuori M, Friman V-P, Nieminen A, Bamford JKH, Buckling A. Bacteriophage selection against a plasmid-encoded sex apparatus leads to the loss of antibiotic-resistance plasmids. *Biol Lett*. 2011;7(6):902-905. doi:10.1098/rsbl.2011.0384

139. Yang QE, Sun J, Li L, et al. IncF plasmid diversity in multi-drug resistant Escherichia coli strains from animals in China. *Front Microbiol*. 2015;6(SEP). doi:10.3389/fmicb.2015.00964

140. Riley MA, Gordon DM. *A Survey of Col Plasmids in Natural Isolates of Escherichia Coli and an Investigation into the Stability of Col-Plasmid Lineages*. Vol 138.; 1992.

141. Lawley TD, Klimke WA, Gubbins MJ, Frost LS. F factor conjugation is a true type IV secretion system. *FEMS Microbiol Lett*. 2003;224(1):1-15. doi:10.1016/S0378-1097(03)00430-0

142. Koraimann G. Spread and Persistence of Virulence and Antibiotic Resistance Genes: A Ride on the F Plasmid Conjugation Module. *EcoSal Plus*. 2018;8(1). doi:10.1128/ecosalplus.esp-0003-2018

143. Bergstrom CT, Lipsitch M, Levin BR. Natural selection, infectious transfer and the existence conditions for bacterial plasmids. *Genetics*. 2000;155(4):1505-1519. doi:10.1093/genetics/155.4.1505

144. Zhou W, Liu L, Feng Y, Zong Z. A P7 Phage-Like Plasmid Carrying mcr-1 in an ST15 *Klebsiella* pneumoniae Clinical Isolate. *Front Microbiol*. 2018;9(JAN):11. doi:10.3389/fmicb.2018.00011

145. Schjørring S, Struve C, Krogfelt KA. Transfer of antimicrobial resistance plasmids from *Klebsiella* pneumoniae to Escherichia coli in the mouse intestine. *J Antimicrob Chemother*. 2008;62(5):1086-1093. doi:10.1093/jac/dkn323

146. Dahlberg C, Bergström M, Andreasen M, Christensen BB, Molin S, Hermansson M. Interspecies Bacterial Conjugation by Plasmids from Marine Environments Visualized by gfp Expression. *Mol Biol Evol*. 1998;15(4):385-390. Accessed December 18, 2020. https://academic.oup.com/mbe/article/15/4/385/1021398

147. Schreiber HL, Conover MS, Chou W-C, et al. Bacterial virulence phenotypes of Escherichia coli and host susceptibility determine risk for urinary tract infections. *Sci Transl Med*. 2017;9(382). doi:10.1126/scitranslmed.aaf1283

148. Thomas-White KJ, Gao X, Lin H, et al. Urinary microbes and postoperative urinary tract infection risk in urogynecologic surgical patients. *Int Urogynecol J*. 2018;29(12):1797-1805. doi:10.1007/s00192-018-3767-3

149. Price TK, Hilt EE, Dune TJ, Mueller ER, Wolfe AJ, Brubaker L. Urine trouble: should we think differently about UTI? *Int Urogynecol J*. 2018;29(2):205-210. doi:10.1007/s00192-017-3528-8

150. Wilhelm SW, Suttle CA. Viruses and Nutrient Cycles in the Sea. *Bioscience*. 1999;49(10):781-788. doi:10.2307/1313569

151. McCarthy AJ, Lindsay JA. The distribution of plasmids that carry virulence and resistance genes in Staphylococcus aureus is lineage associated. *BMC Microbiol*. 2012;12. doi:10.1186/1471-2180-12-104

152. Joshi N, JN F. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33). *Software*. Published online 2011. https://github.com/najoshi/sickle

153. Bankevich A, Nurk S, Antipov D, et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19(5):455-477. doi:10.1089/cmb.2012.0021

154. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30(14):2068-2069. doi:10.1093/bioinformatics/btu153

155. Geneious. Geneious 2021.0.3. Published 2021. https://www.geneious.com

156. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403-410. doi:10.1016/S0022-2836(05)80360-2

157. Antipov D, Hartwick N, Shen M, Raiko M, Lapidus A, Pevzner PA. plasmidSPAdes: assembling plasmids from whole genome sequencing data. *Bioinformatics*. 2016;32(22):btw493. doi:10.1093/bioinformatics/btw493

158. Bethesda (MD): National Library of Medicine (US) NC for BI. National Center for Biotechnology Information (NCBI)[Internet]. Published 1988. Accessed March 9, 2021. https://www.ncbi.nlm.nih.gov/

159. Titus Brown C, Irber L. sourmash: a library for MinHash sketching of DNA. *J Open Source Softw*. 2016;1(5):27. doi:10.21105/joss.00027

160. B B. BBMap 38.90. Accessed March 9, 2021. sourceforge.net/projects/bbmap/

161. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. Volume 26(Issue 19):Pages 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

162. Carattoli A, Zankari E, Garciá-Fernández A, et al. In Silico detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother*. 2014;58(7):3895-3903. doi:10.1128/AAC.02412-14

163. Bortolaia V, Kaas RS, Ruppe E, et al. ResFinder 4.0 for predictions of phenotypes from genotypes. *J Antimicrob Chemother*. 2020;75(12):3491-3500. doi:10.1093/jac/dkaa345

164. Joensen KG, Scheutz F, Lund O, et al. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli. *J Clin Microbiol*. 2014;52(5):1501-1510. doi:10.1128/JCM.03617-13

165. Barrick J. Determining phage titer. Published 202AD. https://barricklab.org/twiki/bin/view/Lab/ProtocolsPhageTiters

166. Barrick J. General conjugation protocol. Published 2020. Accessed March 9, 2021. https://barricklab.org/twiki/bin/view/Lab/ProtocolsConjugation

167. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: A Fast Phage Search Tool. *Nucleic Acids Res*. 2011;39(SUPPL. 2):W347-W352. doi:10.1093/nar/gkr485

168. Arndt D, Grant JR, Marcu A, et al. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res*. 2016;44(W1):W16-W21. doi:10.1093/nar/gkw387

169. Navarro F, Muniesa M. Phages in the Human Body. *Front Microbiol*. 2017;8:566. doi:10.3389/fmicb.2017.00566

170. Abeles SR, Pride DT. Molecular Bases and Role of Viruses in the Human Microbiome. *J Mol Biol*. 2014;426(23):3892-3906. doi:10.1016/j.jmb.2014.07.002

171. Shapiro JW, Putonti C. Gene Co-occurrence Networks Reflect Bacteriophage Ecology and Evolution. *MBio*. 2018;9(2). doi:10.1128/mBio.01870-17

172. Labrie SJ, Samson JE, Moineau S. Bacteriophage resistance mechanisms. *Nat Rev Microbiol*. 2010;8(5):317-327. doi:10.1038/nrmicro2315

173. Zienkiewicz M, Kern-Zdanowicz I, Gołębiewski M, et al. Mosaic structure of p1658/97, a 125-kilobase plasmid harboring an active amplicon with the extended-spectrum β-lactamase gene blaSHV-5. *Antimicrob Agents Chemother*. 2007;51(4):1164-1171. doi:10.1128/AAC.00772-06

174. Smillie C, Garcillan-Barcia MP, Francia M V., Rocha EPC, de la Cruz F. Mobility of Plasmids. *Microbiol Mol Biol Rev*. 2010;74(3):434-452. doi:10.1128/mmbr.00020-10

175. Stephens C, Arismendi T, Wright M, et al. F Plasmids Are the Major Carriers of Antibiotic Resistance Genes in Human-Associated Commensal Escherichia coli . *mSphere*. 2020;5(4). doi:10.1128/msphere.00709-20

176. Šmajs D, Weinstock GM. The iron- and temperature-regulated cjrBC genes of Shigella and enteroinvasive Escherichia coli strains code for colicin Js uptake. *J Bacteriol*. 2001;183(13):3958-3966. doi:10.1128/JB.183.13.3958-3966.2001

177. Mao BH, Chang YF, Scaria J, et al. Identification of Escherichia coli genes associated with urinary tract infections. *J Clin Microbiol*. 2012;50(2):449-456. doi:10.1128/JCM.00640-11

178. Żaczek M, Weber-Dąbrowska B, Międzybrodzki R, Górski A. Phage Prevalence in the Human Urinary Tract—Current Knowledge and Therapeutic Implications. *Microorganisms*. 2020;8(11):1802. doi:10.3390/microorganisms8111802

179. Arredondo-Alonso S, Top J, McNally A, et al. Plasmids shaped the recent emergence of the major nosocomial pathogen Enterococcus faecium. *MBio*. 2020;11(1). doi:10.1128/mBio.03284-19

180. Cornett JB. Spackle and Immunity Functions of Bacteriophage T4. *J Virol*. 1974;13(2):312-321. doi:10.1128/jvi.13.2.312-321.1974

181. Lu MJ, Henning U. The immunity (imm) gene of Escherichia coli bacteriophage T4. *J Virol*. 1989;63(8):3472-3478. doi:10.1128/jvi.63.8.3472-3478.1989

182. Piazolla D, Calì S, Spoldi E, et al. Expression of phage P4 integrase is regulated negatively by both Int and Vis. *J Gen Virol*. 2006;87(8):2423-2431. doi:10.1099/vir.0.81875-0

183. Fogg PCM, Colloms S, Rosser S, Stark M, Smith MCM. New applications for phage integrases. *J Mol Biol*. 2014;426(15):2703-2716. doi:10.1016/j.jmb.2014.05.014

184. Mosher RA, DiRenzo AB, Mathews CK. Bacteriophage T4 Virion Dihydrofolate Reductase: Approaches to Quantitation and Assessment of Function. *J Virol*. 1977;23(3):645-658. doi:10.1128/jvi.23.3.645-658.1977

185. Mosher RA, Mathews CK. Bacteriophage T4-coded dihydrofolate reductase: synthesis, turnover, and location of the virion protein. *J Virol*. 1979;31(1):94-103. doi:10.1128/jvi.31.1.94-103.1979

186. Schmidt AJ, Ryjenkov DA, Gomelsky M. The ubiquitous protein domain EAL is a cyclic diguanylate-specific phosphodiesterase: Enzymatically active and inactive EAL domains. *J Bacteriol*. 2005;187(14):4774-4781. doi:10.1128/JB.187.14.4774-4781.2005

187. Goren MG, Carmeli Y, Schwaber MJ, Chmelnitsky I, Schechner V, Navon-Venezia S. Transfer of Carbapenem-resistant plasmid from *Klebsiella* pneumoniae ST258 to Escherichia coli in patient. *Emerg Infect Dis*. 2010;16(6):1014-1017. doi:10.3201/eid1606.091671

188. Bragagnolo N, Rodriguez C, Samari-Kermani N, et al. Protein dynamics in f-like bacterial conjugation. *Biomedicines*. 2020;8(9). doi:10.3390/BIOMEDICINES8090362

189. Jensen LB, Garcia-Migura L, Valenzuela AJS, Løhr M, Hasman H, Aarestrup FM. A classification system for plasmids from enterococci and other Gram-positive bacteria. *J Microbiol Methods*. 2010;80(1):25-43. doi:10.1016/j.mimet.2009.10.012

190. Zagaglia C, Casalino M, Colonna B, Conti C, Calconi A, Nicoletti M. Virulence plasmids of enteroinvasive Escherichia coli and Shigella flexneri integrate into a specific site on the host chromosome: Integration greatly reduces expression of plasmid-carried virulence genes. *Infect Immun*. 1991;59(3):792-799. doi:10.1128/iai.59.3.792-799.1991

191. Venkatesan MM, Goldberg MB, Rose DJ, Grotbeck EJ, Burland V, Blattner FR. Complete DNA sequence and analysis of the large virulence plasmid of Shigella flexneri. *Infect Immun*. 2001;69(5):3271-3285. doi:10.1128/IAI.69.5.3271-3285.2001

192. Freddolino PL, Goodarzi H, Tavazoiea S. Revealing the genetic basis of natural bacterial phenotypic divergence. *J Bacteriol*. 2014;196(4):825-839. doi:10.1128/JB.01039-13

193. Kropinski AM. Practical advice on the one-step growth curve. In: *Methods in Molecular Biology*. Vol 1681. Humana Press Inc.; 2018:41-47. doi:10.1007/978-1-4939-7343-9_3

194. KIM SY, KO KS. Effects of prophage regions in a plasmid carrying a carbapenemase gene on survival against antibiotic stress. *Int J Antimicrob Agents*. 2019;53(1):89-94. doi:10.1016/j.ijantimicag.2018.09.002

195. Galetti R, Andrade LN, Varani AM, Darini ALC. A Phage-Like Plasmid Carrying blaKPC-2 Gene in Carbapenem-Resistant Pseudomonas aeruginosa. *Front Microbiol*. 2019;10(MAR):572. doi:10.3389/fmicb.2019.00572

196. Pfeifer E, Moura de Sousa JA, Touchon M, Rocha EPC. Bacteria have numerous phage-plasmid families with conserved phage and variable plasmid gene repertoires. *bioRxiv*. Published online November 9, 2020:2020.11.09.375378. doi:10.1101/2020.11.09.375378

197. Ojala V, Laitalainen J, Jalasvuori M. Fight evolution with evolution: Plasmid-dependent phages with a wide host range prevent the spread of antibiotic resistance. *Evol Appl*. 2013;6(6):925-932. doi:10.1111/eva.12076

198. Grin PM, Kowalewska PM, Alhazzani W, Fox-Robichaud AE. Lactobacillus for preventing recurrent urinary tract infections in women: Meta-analysis. *Can J Urol*. 2013;20(1):6607-6614.

199. Akgül T, Karakan T. The role of probiotics in women with recurrent urinary tract infections. *Turkish J Urol*. 2018;44(5):377-383. doi:10.5152/tud.2018.48742

VITA

The author, Cesar E. Montelongo Hernandez, was born in Ciudad Juarez, Chihuahua, Mexico on May 26, 1989 to Cesar Montelongo and Gabriela Hernandez, and grew up in Las Cruces, New Mexico with his siblings (Levi, Tito, Ismael). He attended New Mexico State University in Las Cruces, New Mexico where he earned a Bachelor of Science in Microbiology and Biology, a Bachelor of Arts in Spanish, and a Master's of Science in Biology by 2015. In July 2015, Cesar joined the MD-PhD program at Loyola University Chicago Stritch School of Medicine in Maywood, Illinois.

In 2017, Cesar joined the laboratory of Dr. Alan Wolfe, where he studied the microbial genetics and genomics of urinary bacteria. Cesar plans to return to medical school in 2021 to complete his MD by 2023. He hopes to be involved in a career field that combines scientific research, medicine, and public health.