Supplementary Information for

Urinary bacteriophage cooperation with bacterial pathogens during human urinary tract infections supports lysogenic phage therapy

Mahmood Almosuli, Anna Kirtava, Archil Chkhotua, Lia Tsveniashvili, Nina Chanishvili, Sumaiya Safia Irfan, Emily Ng, Hope McIntyre, Adam J Hockenberry, Robyn P. Araujo, Weidong Zhou, Ngoc Vuong, Barbara Birkaya, Lance Liotta, Alessandra Luchini

Table of Contents:

Supplementary Data 1. Proteomic analysis of T3 phage protein lysate spiked in human urine at a concentration of 2 ng/mL and 0.2 ng/mL. Peptides detected using affinity particle concentration and LC-MS/MS analysis are reported. 26 and 17 phage peptides were identified at 2 and 0.2 ng/mL spike ins, respectively. (Supplementary Data Tab 1)

Supplementary Data 2. Proteomic analysis of urine samples collected from UTI and non-UTI patients. This file contains data pertaining to bacteriophages. Patient identification codes are captured in the first lane of columns AK to BN. Samples F1-F10 = non-UTI, US, Samples F11, F15-F17, F21-F22, F26-F29 = non UTI, Country of Georgia, Samples F12-F14, F18-F20, F20-F25, F30 = UTI. (Supplementary Data Tab 2)

Supplementary Data 3. Proteomic analysis of urine samples collected from UTI and non-UTI patients. This file contains data pertaining to uropathogenic bacteria. Patient identification codes are captured in the first lane of columns BD to CG. Samples F1-F10 = non-UTI, US, Samples F11, F15-F17, F21-F22, F26-F29 = non UTI, Country of Georgia, Samples F12-F14, F18-F20, F20-F25, F30 = UTI. (Supplementary Data Tab 3)

Supplementary Data 4. Proteomic analysis of urine samples collected from UTI and non-UTI patients. This file contains data pertaining to *Homo sapiens*. Patient identification codes are captured in the first lane of columns BY to DB. Samples F1-F10 = non-UTI, US, Samples F11, F15-F17, F21-F22, F26-F29 = non UTI, Country of Georgia, Samples F12-F14, F18-F20, F20-F25, F30 = UTI. (Supplementary Data Tab 4)

Supplementary Data 5. Phage annotation predicted by Bacphlip. Name: Phage Organism ID, Virulent: Bacphlip score capturing the likelihood of the phage to be virulent, 0 being the lowest and 1 being the highest. Temperate: Bacphlip score capturing the likelihood of the phage to be virulent, 0 being the lowest and 1 being the highest lifestyle: assigned lifestyle based on the score; confidence: level of attribution confidence. (Supplementary Data Tab 5)

Supplementary Data 6. Results of the Limma analysis of phage protein abundance. Every lane is a protein. A three class limma test was conducted for each protein, and then post-hoc pairwise limma tests. P values and p values adjusted for multiple testing are reported. (Supplementary Data Tab 6)

Supplementary Data 7. Results of the Limma analysis of bacterial protein abundance. Every lane is a protein. A three class limma test was conducted for each protein, and then post-hoc pairwise limma tests. P values and p values adjusted for multiple testing are reported. (Supplementary Data Tab 7)

Supplementary Data 8. Plaque numbers supporting Figure 6B. (Supplementary Data Tab 8)

Supplementary Data 9. Relative expression values supporting Figure 6J. (Supplementary Data Tab 9)

Supplementary Results:

1 Combined phage and bacterial pathogen protein discriminate between UTI cases and non-UTI controls. A LASSO logistic regression model [https://www.jstor.org/stable/2346178] was employed on the combined phage and bacterial pathogen proteins to discriminate cases and controls. This model identified a set of 14 combined phage and pathogen proteins that accurately discriminate cases and controls in the training set. These proteins hold promise as potential biomarkers for UTI prediction and provide insights into the biological mechanisms underlying the survival and pathogenicity of UTI related organisms in the urinary tract. Those 14 proteins are described below.

a) The *E coli* Colicin I receptor (Cir) was positively associated with UTIs. The high abundance of this protein in UTI patients might be due to several reasons related to bacterial survival and pathogenicity. Cir involvement in iron and nutrient acquisition is important for bacterial survival in iron limited environments like the urinary tract [PMID: 136550]. Cir is also involved biofilm formation, development of antibiotic resistance and virulence of bacteria [PMID: 31737922].

b) A second protein that was found to have greater abundance in UTIs is *E coli* Cysteine desulfurase. This protein mobilizes sulfur from cysteines to produce sulfur containing biomolecules and in particular iron-sulfur clusters, which are essential co-factors for enzymes involved in respiration, DNA repair and metabolism [PMID: 18322036]. During UTI, *E. coli* might experience oxidative stress and nutrient deprivation, and the production of iron-sulfur clusters might be beneficial to maintain essential enzyme functions.

c) The *E. coli* large ribosomal subunit proteins uL24 and uL2 were inversely related to UTI, that is to say, were less abundant in UTIs than in non-UTIs. The reduced abundance of these proteins during infection might be related to ribosomal hibernation and translational regulation in response to an unhospitable and nutrient poor host environment [PMID: 33555244, PMID: 36264977].

d) The *Pseudomonas aeruginosa* YfdX protein was more abundant in UTIs than in non-UTI controls. YfdX is predicted to have peptidyl-prolyl cis-trans isomerase activity, thus catalyzing the cis-trans isomerization of peptide bonds at proline residues. This enzymatic activity is required in protein folding, chaperones, signal transduction, protein trafficking, and cell cycle regulation [PMID: 30692978]. YdfX might support the bacterial cells in stress response and adaptation to environmental challenges.

e) The MAG TPA protein, known as endosialidase chaperone, is involved in folding and assembly of phage endosialidases, which are trimeric tail spike proteins found in Escherichia coli K1-specific bacteriophages. Downregulation in UTI could be indicative of lack of phage particle assembly and thus supporting the concept of temperate phages being in their integrated prophage state [PMID: 9367760].

f) The *E coli* putative pesticin receptor is over-expressed in UTIs. The higher abundance of this receptor can be explained by its role in iron and nutrient uptake, biofilm formation and possible antibiotic resistance, which are survival and virulence mechanisms [PMID: 7984105].

g) *Enterococcus faecalis* cyclic pyranopterin monophosphate synthase MoaC, which is involved with the synthesis of molybdenum co-factors, is positively associated with UTIs. Its importance for molybdoenzyme activity and its overabundance during infection might be related to specific metabolic needs of the bacterium in face of oxidative stress, biofilm formation and virulence factor regulation [PMID: 31341072, PMID: 33362753].

h) The *Klebsiella* Cytochrome b561 homolog 2 was less abundant in UTIs than in non-UTI controls. A diminished amount of this protein, which is involved in the electron transport chain, in infection can signal a metabolic shift of the bacterium in response to urinary tract specific conditions [PMID: 22439737].

i) The *Lactococcus* phage N4-gp56 family major capsid protein is over-expressed in UTI cases. This protein forms the shell (capsid) that encloses the viral genetic information, playing a central role in the assembly of phage particles and the subsequent infection of bacterial host cells [PMID: 18374942]. If confirmed, this information can be explained by an inverse relationship between the *Lactococcus* phage particles and its bacterial host, which is present in low abundance in the UTI infection, where the ecological equilibrium has been disrupted and the pathogenic bacteria have proliferated and prevailed.

j) Four uncharacterized phage proteins CPT_Myduc_012, A2_158, A0A7H0XAJ2, and A0A0K1LKF6 were included in the discriminatory panel between UTI cases and controls. The function of these proteins as pertains to phage – bacteria interactions in UTI warrants more investigations.

2 Phage proteins discriminate between cases and controls. To select phage protein that could discriminate cases from controls, we performed feature selection using a least absolute shrinkage and selection operator (LASSO) logistic regression model applied to the 10 cases and 20 controls. A 10-fold cross validation to fit the model and validate on a holdout sample was performed to identify lambda minimum (Supplementary Figure 2). Using lambda minimum, we identified 11 proteins (Table 2) that could successfully assign the UTI status to the training dataset (Membrane lipoprotein, A0A0K1LKF6, Host factor I for bacteriophage Q beta replication, a growth-related protein A0A1Y3FEC6, Uncharacterized protein KMC53_gp11 A0A2H4YGX6, Bacteriophage Mu tail sheath A0A564WHB1, MAG TPA: endosialidase chaperone DAQ14039.1, Virion structural protein I6S6X7, dUTP diphosphatase Q2NP83, N4-gp56 family major capsid protein Q9AZ61, uncharacterized protein HWC59_gp12 QFG06635.1, uncharacterized protein bas50_0172 QXV77638.1, Tail tape measure domain-containing protein OS S5MLP4.

3 Proteins that mediate all phage life phases were identified in the urine. 143 protein groups attributed to bacteriophages were identified in this study. The urinary proteome analysis yielded proteins that play a fundamental role in all phases of the bacteriophage life cycle, both in the lytic and lysogenic phenotype. Lytic lifecycle phases include 1) **phage adsorption**, 2) **penetration**, 3) **replication-recombination-repair**, 4) **maturation**, and 5) **release** [PMID: 36561998], while lysogenic lifecycle includes the specialized phases 6) **integration**, and 7) **lysogenic conversion**, whereby selected prophage genes are expressed during lysogeny [PMID: 36561998]. Below, proteins related to each lifecycle phase that were identified in the urine of UTI patients and controls are reported.

3.1) Adsorption: tail fiber protein Peptides belonging to the lateral tail fiber protein of Escherichia phage Phieco32 (of the Podoviridae family PMID: 18294652), and to Klebsiella

phage Sweeny gp12 (tail adhesin) (syphoviridae PMID: 31558643) were identified. Fiber proteins are the furthest appendages of the phage tail and initiate the adsorption process. Fiber proteins come into direct contact with the bacterial host cell surface receptors and act as adhesins. Target molecules include lipopolysaccharide (LPS), porin transmembrane proteins, teichoic acids, and even organelles such as pili or flagella. Together with spikes, the tail fiber proteins determine the host range and specificity. [PMID: 36292999, PMID: 27471503]. Binding to the bacterium cell wall through tail 5 fiber proteins is the first step of phage – bacteria interaction. Phage tail conformational changes then lead to phage tail tube penetration of the bacterial cell wall and to phage DNA injection. Although phage families (myoviridae, siphoviridae, and podoviridae) differ in their tail conformation, the fiber proteins maintain the conserved task to interact with bacterial receptors and to initiate the adsorption process [PMID: 30104690]. Dissimilarly from other phage tail proteins described below, fibers do not have any enzymatic activity. Endosialidase chaperone from Bacteriophage sp. Many tail spike proteins, which like fibers are main determinants of bacterial host specificity, have the capability to enzymatically degrade the thick layer of glycans that protect encapsulated bacteria. The spike polysaccharide depolymerase activity (lyase or endoglycosidase) allows the phage to access the bacterial cell membrane. The E coli K1 phage tail spike structure revealed the presence of a small C-terminus domain of ~150 amino acids that acts as chaperone. Such domain is enzymatically cleaved during phage maturation and it is necessary for proper folding and trimerization of the spike protein [PMID: 17158460].

3.2) Penetration: DNA injection: Bacteriophage Mu tail sheath. Mu phages, together with T4 phages, are members of the Myoviridae family, which includes phages that infect enterobacteria and that have long, contractile tails [PMID: 30104690]. Their contractile tails confer great viral infection efficiency to the myoviridae family. A protein sheath envelops the myoviridae tail that is capable to shrinking down to less than half of its initial length. 138 copies of the tail sheath protein, gene product (gp) 18 form the myoviridae tail sheath, which encases a central, non - contractile tail tube. When the phage is securely anchored to the bacterial cell surface, and thanks to sheath contraction, the tail tube penetrates the outer membrane and creates a channel for viral genome delivery [PMID: 19229296]. The tail tape measure protein of various phages was identified. Tape measure proteins, mainly characterized in siphoviridae, are named so because they play a role in defining the tail length. During phage penetration of the bacterial host cell wall, tape measure proteins are hypothesized to form a channel between the outer and inner membrane membranes of the host, thus facilitating the DNA transit to cell cytoplasm. Tape measure proteins include three domains: a sensor, a transmembrane domain and a metallopeptidase domain. The sensor domain triggers conformational changes in the capsid proteins that lead to the release of capsid DNA. The metallopeptidase domain degrades the bacterial peptidoglycan, which is followed by the outer and inner membrane fusion and leads to the formation of a channel for DNA injection [PMID: 27824135]. In some phages (e.g., T4) the tape measure protein fuses the two 6 membranes, in other cases (e.g., coliphage HK97) the pore formation is aided by host proteins. Phage species identified: Streptococcus phage Spn1, Streptococcus phage IPP44, Escherichia phage T5, Escherichia phage IrisVonRoten, Pseudoalteromonas phage TW1, Proteus phage vB_PmiS_Jing313.

3.3) Replication, recombination, repair: we identified phage coded proteins that support phage replication, recombination and repair with different functions: a) nucleotide metabolism, b) nucleic acid processing, c) translation and d) carbohydrate metabolism.

a) Nucleotide metabolism: NrdC thioredoxin (Escherichia phage JS98) is a small ubiquitous redox protein, which is expressed at all stages of the T4-like phage infection cycle. It is a hydrogen donor for the phage ribonucleotide reductase, which is the key enzyme for deoxyribonucleotide biosynthesis [PMID: 8662944]. We identified the ribonucleoside diphosphate reductase enzymes of Escherichia phage UB, Escherichia phage 121Q. This enzyme is the limiting factor in the initiation and rate of deoxyribonucleotide synthesis of T4-like bacteriophages in infected cells [PMID: 2228963]. Another nucleotide metabolism enzyme we found in our clinical study is dUTP diphosphatase (Xanthomonas phage OP2). This enzyme catalyzes the Mg2+-dependent hydrolysis of deoxyuridine triphosphate (dUTP) to deoxyuridine monophosphate (dUMP), providing the substrate for thymidylate synthase (EC 2.1.1.45) that leads to the production of thymidine nucleotides. By reducing the effective ratio of dUTP to TTP, the enzyme also reduces the possibility of dUTP incorporation into DNA, which is abnormal and would result in DNA instability [PMID: 4291202].

b) Nucleic acid processing: DNA replication. The DNA primase / DNA helicase of Enterococcus phage EFDG1 (Myoviridae PMID: 31010053) was identified in our clinical study. In model organism phage T7, the primase-helicase performs several key tasks in the DNA replication, acting both as the molecular motor that advances the replication fork by unwinding the DNA and as the trigger for synthesis of each Okazaki fragment [PMID: 34578319]. Single stranded DNA-binding protein of Vibrio phage VvAW1, podoviridae [PMID: 23408718]: studied in bacteriophage T7, the single-stranded DNA binding protein is essential for replication, recombination, repair, and the maturation of viral genomes. In addition to binding ssDNA, the protein binds to DNA polymerase and DNA helicase, and regulates their activities. The protein has potent homologous DNA annealing activity, and plays an important role in recombination [PMID: 29588157]. ATP-dependent DNA ligase (Enterococcus phage nattily, siphoviridae 7, PMID: 33488539) is the enzymes responsible for joining breaks in the phosphodiester backbone of DNA during replication and repair. This enzyme displays a great heterogeneity in size and structure throughout phage families. In bacteriophage T7, for example, this enzyme repairs damaged DNA strand through esterification of a 5'- phosphoryl to a 3'- hydroxyl group using ATP [PMID: 34548548]. Bacteriophage replication gene A protein (GPA, Raoultella terrigena) is an endonuclease capable of inducing a single-strand cut to initiate DNA replication, and to ligate the 3’ and 5’ ends of the viral DNA to form a circle. Studies of the Bacteriophage phi X174 demonstrated that the gene A protein is required for initiation and termination of successive rounds of circular phi X DNA replication [PMID: 2940511]. MAG TPA: ASCH domain protein [Bacteriophage sp.]: members of the ASCH superfamily are thought to function as RNA-binding domains and sustain RNA-processing and prokaryotic translation regulation [PMID: 16322048, PMID: 16452142]. Anti-sigma70 protein (Escherichia phage PaulHMueller) and sigma factor (Staphylococcus phage vB_SsapH-Golestan-100): transcription initiation is regulated in bacteria by the σ factor, which is provides the RNA polymerase with the ability to engage with specific promoters and to initiate RNA synthesis initiation [PMID: 32232532]. The primary σ factor of E. coli is σ70, which is involved in the transcription of housekeeping genes. Other σ factors are used under certain growth conditions or at times of stress. Anti-σ factors directly interact with σ proteins, and inhibit or modulate RNA polymerase activity [PMID: 15561138]. Genes for sigma factors and anti-sigma proteins, which are the predominant constituents of transcription regulation in bacteria, are carried in the phage genome and are expressed during phage replication to co-opt the bacterial transcription machinery [PMID: 25232540, PMID: 15561138]. Bacteriophage

T4 polynucleotide kinase (PNK, Proteus phage Privateer, podophage PMID: 33614267): PNK has the key function to repair bacterial RNA and DNA. In bacteriophage T4, PNK repairs tRNA anticodon loops that are cleaved as a defense against T4 phage infection [PMID: 12110598, PMID: 14754987]. PNK is a multifunctional enzyme that has 5′-kinase, 3′-phosphatase, and phosphodiester hydrolysis activity [PMID: 22499790].

c) Translation: glutamine--tRNA ligase (Escherichia phage D6, temperate myoviridae): catalyzes the attachment of the amino acid glutamine to its cognate transfer RNA during mRNA translation, PMID: 6382258. Glutamine-tRNA ligase is a relatively rare ligase, found in 8 Escherichia coli and other Gram-negative bacteria, in Deinococcus radiodurans, and in the cytosol of eukaryotic cells [PMID: 10704480].

d) Carbohydrate metabolism: pyruvate formate-lyase (Shigella phage CM8, myoviridae): Pyruvate formate lyase (PFL) is responsible for converting pyruvate and coenzyme A (CoA) into formate and acetyl CoA and is an essential enzyme in bacterial anaerobic metabolism [ISBN: 0080453821, 9780080453828].

3.4) Maturation: a) capsid assembly, b) DNA packaging, c) structural proteins.

a) Capsid assembly: portal protein (Streptococcus phage SpSL1, siphoviridae, PMID: 32582130) The portal protein is part of the portal complex, which plays a crucial role in all aspects of bacteriophage assembly from capsid assembly, DNA packaging, portal ring assembly and incorporation. The assembly pathway of Bacteriophage P22 involves a nucleation complex composed of coat protein monomers, scaffolding proteins, and portal protein monomers. This complex forms the portal ring, serving as a foundation for the assembly of the coat protein shell by scaffolding proteins. Ejection proteins are incorporated into the procapsid prior to DNA packaging by terminase proteins. The pressure exerted by the DNA leads to a conformational change in the capsid and portal protein, which, in turn, facilitates the dissociation of the terminase complex and the addition of the tail machinery [PMID: 31337287]. Head vertex assembly chaperone (Proteus phage PM2 myoviridae) is another protein identified in the urine that mediates portal initiator assembly during maturation [PMID: 36851741]. Additionally, we identified the *Escherichia* phage vB_EcoP_ZX6 scaffolding protein. This protein plays a pivotal role as a catalyst and chaperone, guiding the formation and closure of the capsid shell into an icosahedral structure. The scaffolding protein is not a component of the mature shell; it exits the procapsid interior, either intact or after degradation by a viral protease, around the time of DNA packaging.

b) DNA packaging: the *Streptococcus* phage IPP16 HNH endonucleases [PMID: 32582130] is a key component of the phage DNA packaging machine and plays a variety of roles in the phage lifecycle [PMID: 28211904]. Terminase small subunit (Escherichia phage lambda, siphoviridae): the terminase proteins play a crucial role in phages by recognizing DNA and initiating the DNA packaging process. They exhibit ATPase and nuclease functions and have the capacity to attach to specific DNA sequences when it is in its unassembled state. [PMID: 23011306].

c) Structural proteins: a number of structural proteins were identified in the urine study and are reported below. *Klebsiella* phage AmPh_EK29 Major capsid protein, Streptococcus phage phiNJ2 Gp7envelope protein, *Lactococcus* phage TP901-1 MHP (major head protein) [PMID: 8610457], *Staphylococcus* phage vB_SsapH-Golestan-100 tail morphogenetic

protein, which can stimulate humoral immunity with antibody production in mice [PMID: 31803179]. KMC53_gp11 is a structural component of the short non-contractile tail [PMID: 23884409] N4-gp56 family major capsid protein, gene product 56 (gp56), Internal virion protein B (*Escherichia* phage vB_EcoP_ZX6). Several of these proteins are known to have specific effect on their microenvironment and specific functions. For example, the tail morphogenetic protein of staphylococcal bacteria interacts with the mammalian host to stimulate humoral immunity with antibody production, as demonstrated in mouse model studies [PMID: 31803179]. The capsid protein Gp56 encoded by the lytic bacteriophage SP01, has been shown to inhibit *Bacillus subtilis* cytokinesis thus hindering the bacterium reproductive capabilities. Mechanistically, gp56 localizes at the division machinery and blocks the recruitment of proteins needed for septal cell wall synthesis [PMID: 33077634]. The *Escherichia* coli phage T7 contains three internal core proteins, gp14, gp15, and gp16, which are essential for phage morphogenesis and translocation of the ~40 kbp T7 genome [PMID: 35336080].

3.5) Release: Host lysis protein (*Bacillus* phage): to release the offspring virions, phages have to induce bacterium host cell lysis, and they do so by two known mechanisms, either by degrading the host cell wall or by inhibiting the cell wall synthesis [PMID: 10760296].

Phage lytic proteins include endolysins, anti-/holins, and spanins [PMID: 36762092]. The proteomic analysis yielded Paenibacillus sp. PDC88 toxin secretion/phage lysis holin. Typically, double-stranded DNA phages carry a minimum of two genes responsible for triggering the lysis of the host cell: an endolysin, an enzyme that breaks down the cell wall, and a holin, which, at a genetically predetermined moment, permits the endolysin to traverse the cytoplasmic membrane and reach the cell wall. [PMID: 10760296, PMID: 11444771].

The Gp138_N domain-containing protein of *Pseudomonas* phage JHP OX Gp138 is believed to play a role in the mechanism of opening the host cell membrane during the infection process [PMID: 22325780]. Vibrio phage vB_VpaP_KF2 glycosyl hydrolase: encoded by the phage DNA, this protein is a lytic enzyme that hydrolyzes the peptidoglycan layer in the bacterial cell wall [PMID: 22991936]. A recent study shows that the expression of this peptidoglycan hydrolase in lytic phages was sufficient (Atu-ph02 and Atu-ph03), without the expression of other genes, to block cell division and cell lysis in *Agrobacterium tumefaciens* [PMID: 28970228].

Phage coded proteins that are associated with the phenomenon of lysis inhibition were identified in this clinical study (e.g., *Roseobacter* phage DSS3P8 RIIA lysis inhibitor). Lysis inhibition is a strategy that lytic phages use to temporarily delay the destruction of the host bacterium to allow the production of a larger number of offspring phage virions before their release. When *Escherichia coli* cells are individually infected by bacteriophage T4, they typically undergo lysis within 25–30 minutes at 37 °C. However, in high-density cell cultures with a high multiplicity of infection, lysis is delayed for several hours. During this delay, the infecting phage continues to replicate and mature within the host before eventually releasing approximately 10^3 progeny particles [PMID: 10400598]. This specific behavior provides a distinct advantage to the phage because it significantly increases the number of offspring produced during a single burst, thereby enhancing the likelihood of these progeny finding new hosts, particularly in situations with limited host availability [PMID: 14637004].

Phages with a lysogenic lifestyle undergo adsorption and penetration similarly to the lytic phages, and then they perform integration, and expression of selected prophage genes during lysogeny (lysogenic conversion). Urinary proteins involved in lysogenic-specific stages include the following.

3.6) Integration, recombination: Bacteriophage genomes undergo rapid evolution through mutation and horizontal gene transfer as a response to the evolving defensive mechanisms of their bacterial hosts. Phages in a natural community possess core genome elements, and numerous flexible gene modules that are exchanged by homologous recombination to ensure phage diversity to overcome evolving bacterial host defenses [PMID: 32879312]. The recombination system of lambda phage has been extensively studied and includes two proteins, an exonuclease that binds to double strand DNA ends and digest the 5'-strand, and a single strand annealing protein that binds to the resulting 3' strand and anneals it to a complementary strand from another DNA molecule. Homologous recombination occurs by several different pathways, and phage-encoded proteins participate with host proteins in various ways. Recombination machinery of the bacterial host are used by the phage in its lytic cycle. Lambda also carries two genes in the nin region that affect recombination, orf/ninB and rap/ninG. In this study, we found the Bacteriophage lambda (syphoviridae, temperate) NinG/Rap protein, which is a DNA structure-specific endonuclease that cleaves Holliday junction branch points In [PMID: 30904699]. The phage AB09 YqaJ-like viral 11 recombinase Mu-like cryoconite is an exonuclease involved in red recombination, a mechanism of homologous recombination of lambda and other temperate phages. This is an alkaline exonuclease that participates in the digestion of linear double-stranded DNA in a Mg^{2+}- dependent reaction, and this process is part of the genome excision and integration mechanism [PMID: 33841378]. *Escherichia* phage D6 (temperate myoviridae) exonuclease: recombination-associated exonucleases in temperate phages (e.g., lambda phage) are involved in homologous DNA recombination, which enables both genomic stability and plasticity [PMID: 29304472]. Lambda exonuclease is a 5'-->3' exonuclease that degrades double-stranded (ds)DNA. The single stranded DNA produced by lambda exonuclease is utilized by homologous pairing proteins to carry out homologous recombination [PMID: 12626699]. bacteriophage Mu (Myoviridae PMID: 7813916) Gam like protein MuGam is part of a two-component bacterial DNA repair system with bacterial ligase. [PMID: 30487222]. MuGam binds to DNA ends and repairs DNA double strand breaks, facilitates Mu replication and protects linear Mu phage DNA from nuclease digestion [PMID: 12524520]. Mu phage is a very efficient transposable element, and Gam is also involved in transposition events [PMID: 12524520].

3.7) Lysogenic conversion. Expression during lysogeny. Enterobacteria phage PA-2 Outer membrane porin protein LC. Porin proteins are predominant components located in the outer membranes of Gram-negative bacteria, as well as in mitochondria and chloroplasts. These proteins form channels allowing for the unrestricted passage of small solutes. Their expression is primarily regulated by factors such as medium osmolarity, temperature, and carbon source. A porin gene is also present in the genomes of certain lambdoid bacteriophages, and it is expressed in the lysogenic state. The expression of this gene leads to a substantial reduction in the expression of other porins, including OmpC and OmpF proteins, that are expressed by the bacterium. OmpC is a receptor that phage PA-2 uses for adsorption, and with the expression of a competing porin, the phage is able to modulate phage susceptibility of the bacterial host [PMID: 3017988]. Enterobacteria phage fiAA91-ss Cytolethal distending toxin V, C subunit: the cdt toxin gene encodes for the cytolethal distending toxin (CDT), a protein toxin that exhibits

cytotoxic effects on mammalian cells. CDT is a potent virulence factor that can cause cell cycle arrest, DNA damage, and cellular distension. Bacterial cdts are categorized as AB2-type toxins known for their ability to impede mammalian host G2 and early M phases during mitosis. As a result, mammalian cells are unable to divide, and they continue to grow, causing them to expand up to five times their normal size before eventually disintegrating. Cdt serves as a bacterial virulence factor that contributes to bacterial survival and amplifies the pathogenicity of microorganisms. Current experimental findings support the pivotal role of Cdts in the in vivo pathogenicity of Cdt-producing bacteria. These bacteria can trigger conditions such as hepatocellular carcinoma, as well as inflammation in the stomach, intestine, and liver in susceptible mouse strains during chronic infection [PMID: 34868002]. The genes responsible for Cdt production are typically located within the bacterial chromosome and plasmids, but also in bacteriophages [PMID: 21646456]. The impacts of the lytic or lysogenic lifestyle on bacterial fitness depend on the resistance strategies employed by the bacteria. Bacteria can develop resistance to phages through various mechanisms, such as modifying surface receptors or undergoing lysogenization. However, these resistance strategies may come with fitness costs that can affect bacteria. The specific fitness costs associated with each resistance strategy are not yet fully understood. It is likely that the fitness costs of lysogenization may differ from those of surface receptor modification. Further research is required to comprehensively examine the effects of these lifestyles on bacterial fitness. The phage lifestyle (lytic and lysogenic) depends on many factors (bacterial host physiological state, phage density). Lysogeny can have important consequences on the bacterial host, including engineering the bacterial genome, changing the bacterial physiology, facilitate transduction to transfer bacterial DNA, and induce and release phages that modify the whole bacterial community. The effects of phage lysogeny and its symbiotic relationship with the bacterial host might be unfavorable to the human host.

4 Pathogen proteins discriminate between cases and controls. A LASSO logistic regression model was used to identify a reduced set of pathogen proteins that could discriminate UTI samples from controls. A set of 11 proteins (Supplementary Table 3, Supplementary Figure S4) was identified that could correctly discriminate samples in the training set: *Streptococcus* sp. pneumococcal-type histidine triad protein, *Escherichia coli* beta-D-glucuronidase, *Escherichia coli* L-fucose isomerase, *Enterococcus faecalis* cyclic pyranopterin monophosphate synthase MoaC, *Klebsiella sp*. Cytochrome b561 homolog, *Escherichia coli* Biopolymer transport protein ExbB, *Pseudomonas aeruginosa* Phage major capsid protein, *Klebsiella* sp. Betaine aldehyde dehydrogenase, *Pseudomonas aeruginosa* Serine hydroxymethyltransferase 2, *Escherichia coli* Large ribosomal subunit protein uL2, and *Escherichia coli* Colicin I receptor. Four of these proteins were discussed in Supplementary Results S1. Histidine triad proteins (Pht proteins) in bacteria such as *Streptococcus pneumoniae* are involved in metal ion binding and transport. These proteins play a role in virulence and bacterial adherence to host cells. They help the bacteria acquire essential metal ions like manganese and zinc, which are critical for bacterial growth and pathogenicity [PMID: 28624420]. Beta-D-Glucuronidase is involved in the degradation of complex carbohydrates, allowing the bacteria to utilize glucuronic acid as a carbon source [PMID: 18685721]. It can also play a role in pathogenicity by breaking down host glycosaminoglycans. L-Fucose Isomerase is an enzyme that is part of the metabolic pathway for the utilization of L-fucose, a sugar found in animal glycoproteins and glycolipids. Bacteria use this enzyme to metabolize L-fucose as a carbon and energy source [PMID: 9367760]. Biopolymer Transport Protein ExbB is part of the TonB-ExbB-ExbD complex, which is involved in the energy-dependent transport of nutrients such as iron-siderophore complexes and vitamin

B12 across the outer membrane in Gram-negative bacteria [PMID: 2670903]. Betaine Aldehyde Dehydrogenase catalyzes the oxidation of betaine aldehyde to betaine (trimethylglycine). This enzyme is part of the pathway for the biosynthesis of betaine from choline. This enzyme might have the dual role of assimilating carbon and nitrogen from choline, abundant at the infection site, and producing betaine, which is an important osmoprotectant that helps bacteria to survive in high osmolarity environments by maintaining cell turgor and protecting cellular functions [PMID: 16315011]. Serine Hydroxymethyltransferase 2 catalyzes the reversible conversion of serine and tetrahydrofolate (THF) to glycine and 5,10-methylenetetrahydrofolate. This reaction is a key step in one-carbon metabolism, which is essential for the synthesis of nucleotides, amino acids, and other biomolecules. SHMT plays a critical role in cellular metabolism and growth [PMID: 35739195].

5 Proteins that mediate all known bacterial resistance mechanisms to phages were identified in the urine. We identified 845 protein groups attributed to UTI-causing pathogens (including UPEC, Klebsiella pneumonia, Pseudomonas aeruginosa, Enterococcus faecalis, Proteus mirabilis) with UPEC being the most frequent. Identified proteins are major functional players in all the known bacteria defense mechanisms: a) **surface alterations to avoid phage adsorption and prevention of phage DNA injection**, b) **restriction of incoming DNA**, c) **blocking phage assembly**, d) **acquiring phage-specific immunity through clustered regularly interspaced short palindromic repeats and abortive infection** (Abi). PMID: 26066799.

5.1) Prevention of phage adsorption: Type IV pilus inner membrane component PilN Pseudomonas aeruginosa Pili are filamentous structures present on the bacterial surface that mediate a vast array of functions, including adhesion, motility, microcolony formation, secretion of proteases and colonization factors; however, their primary function is bacterial conjugation. Type IV pili consist of homopolymers constituted by monomeric pilin proteins, each weighing approximately 15–20 kDa. These pili are present on the surfaces of numerous Gram-negative bacteria and are also found in at least one Gram-positive organism. During bacterial conjugation, the pili extend outward from the donor cell and adhere to specific receptors on the recipient cell's wall. Cells are drawn close together when the pilus retract and undergoes depolymerization. Genetic material is transferred in this conjugating connection. Several phages, primarily from the Cystoviridae and Inoviridae families, use the pilus structure to initiate the adsorption to the bacterial host. Adsorption mediated by pilus adhesion enables the phages to have a wide range of hosts, since the pilus structure is conserved among many bacterial species [PMID: 26755501]. To confirm the important role of the pilus structure, recent literature reports show that deletions of or mutations in the genes involved in pilus structure and biogenesis prevented bacteriophages DLP1 and DLP2 binding and lysis in P. aeruginosa strains [PMID: 29925793]. P. aeruginosa PilN is an inner-membrane assembly protein and aids with pilus biogenesis. Other reports show that P. aeruginosa type IV pilus (TFP) can be modulated by lysogenic phages. For example, the phage D3112 carries a protein known as Tip, which attaches to a TFP ATPase, hindering its positioning. This interference leads to a reduction in surface piliation and offers protection against superinfection with other phages relying on TFP for their adsorption [PMID: 26066799]. An Escherichia coli (strain K12) Outer membrane protein TolC, which is needed for T7 family phage infection [PMID: 29259138], was identified in the urine.

5.2) Restriction-Modification Systems (RMSs) have been identified as intrinsic bacterial defense mechanisms, employing methylation of specific motifs to distinguish between native and foreign DNA sources and selectively degrading foreign DNA. RMSs are present in 90% of all documented prokaryotic genomes, with 80% of them featuring multiple systems, indicating a rich and widespread reservoir of potential epigenetic information. RMSs are categorized into four types (I to IV) based on their co-factor requirements, subunit composition, cleavage patterns, and mechanisms. Type I RMSs were the earliest to be uncovered and are comprised of multi-subunit enzyme complexes, consisting of three primary proteins: a methyltransferase (typically referred to 14 as HsdM), an endonuclease (HsdR), and a sequence recognition protein (HsdS) [PMID: 34181709]. In our proteomic study we identified three methyltransferases: 1) Escherichia coli O45:K1 AdoMet-dependent methyltransferase, which is a co-factor required for Type I, II, and III restriction modification systems. 2) K/L Escherichia coli (strain K12) DNA methyltransferase Ribosomal RNA large subunit methyltransferase [PMID: 31798540, PMID: 12079349], and Escherichia coli (strain 55989 / EAEC) ModE protein, which is a methyltransferase involved in the Type II R-M system.

5.3) Assembly interference: we have found Klebsiella pneumoniae subsp. rhinoscleromatis peptidyl-prolyl cis-trans isomerase ppiC, which is an enzyme that prevents the phage packaging process.

5.4) Phage-specific immunity through clustered regularly interspaced short palindromic repeats: The CRISPR-Cas bacterial adaptive immune system includes the Cas9 protein, which has two endonucleases accountable for creating double-strand breaks. Specifically, the HNH domain is responsible for cleaving the target strand of DNA duplexes, while the RuvC domain is dedicated to the nontarget strand. In our proteomic investigation, we successfully measured the Streptococcus phage IPP16 HNH endonuclease [PMID: 35420793]. Phage-specific immunity through and abortive infection. We observed the abortive infection bacteriophage resistance protein (Abi) in patient samples. Abi is a part of 'altruistic' cell death systems activated in response to phage infections, which curtail viral replication and safeguard the bacterial population. Although the Abi strategy has a common principle of cell suicide upon viral infection sensing and prior to phage replication cycle completion, it is carried out through a wide array of defense systems within bacterial genomes, each operating through distinct mechanisms. Every Abi system incorporates at least two functional modules: one designed to detect the presence of a phage infection, and the other responsible for triggering cell termination or metabolic shutdown once the phage infection has been detected [PMID: 32559405]. Activated Abi systems are often highly toxic, targeting various key cellular processes to impede phage DNA replication, transcription, and protein synthesis. Abi systems can induce degradation or depolarization of the inner membrane, degrade both phage and host DNA and RNA, including transfer RNAs (tRNAs), cleave essential proteins within the host's translation machinery [PMID: 32559405]. While the majority of Abi systems are typically found in Gram-positive lactococcal strains, some have also been identified in Gram-negative species like Escherichia coli, Vibrio cholerae, and Shigella 15 dysenteriae. We identified a gram-negative Abi family protein which is a protease and triggers cleavage of essential bacterial host proteins [PMID: 32559405, PMID: 19124776]. UTI patients exhibited higher pathogen protein abundance compared to non-UTI controls, as well as higher levels of phage protein abundance. This observation suggests a relationship between phage and pathogen protein abundance in association with UTIs.

6 Phage precipitated from urine using affinity hydrogel particles contain large DNA molecules that can be packaged into T7 phages.

Poly(NIPAm/AA) affinity hydrogel particles were incubated with human urine and then were subjected to DNA extraction (Supplementary Fig. 8). The purified DNA was compared to DNA extracted from phages isolated with the agar overlay method (Supplementary Figure S5). An agar gel demonstrated that the size of DNA extracted from the particle precipitated material is greater than 2000 bp (Supplementary Figure 5C lane 4), similarly to the DNA extracted from the phages isolated and propagated with the agar overlay method (Supplementary Figure 5C Lances 2-3). Although the gel resolution doesn't support a precise estimation of the DNA size, poly(NIPAm/AA) particle derived DNA is greater than the expected range of the most abundant urinary cell free DNA (50-300 bp, PMID: 32325682) and is compatible with enrichment of viral organismal DNA from the urine matrix. The particle-derived DNA was packaged in phage capsules using the Novagen T7 Select packaging system. Packaged phages were plated on T7 phage sensitive *E. coli* BL21 lawn and formed plaques that were similar to the *Podoviridae* phages directly isolated from the urine (1-4 mm diameter, Supplementary Fig. S5C). Packaged phages failed to further lyse bacteria upon subsequent passages. The use of the Novagen T7 select packaging system to test the re-packaging ability of phage DNA into a different capsid is novel, as the system was designed for phage display and protein expression in *E. coli*. Biases in this packaging system towards certain types of phage genome and whether the system is compatible with other phages rather than just *E. coli* phages are unknown. More experimental evidence to pursue this novel indication of the system and to answer these open questions is warranted.

Supplementary Tables

Supplementary Table 1. A combination of phage and bacterial proteins selected through the lasso feature selection algorithm discriminated UTI cases from non-UTI controls.

Accession ID	Description	Organism	Lasso Coefficient
P17315	Colicin I receptor	*Escherichia coli*	5.0409062
A0A376Y6R3	Cysteine desulfurase	*Escherichia coli*	11.1211315
A0A2X1JEL2	Large ribosomal subunit protein uL24	*Escherichia coli*	-0.1040636
A0A7H0XAJ2	Uncharacterized protein	*Xanthomonas* phage Xoo-sp14	-9.3390244
A0A241XSA9	YfdX protein	*Pseudomonas aeruginosa*	1.4304863
DAQ14039.1	MAG TPA: endosialidase chaperone	Bacteriophage sp.	-8.9968805
A0A0H2V8C1	Putative pesticin receptor	*Escherichia coli* UPEC	0.3778550
A0A0K1LKF6	Uncharacterized protein	*Vibrio* phage H188	-1.3544457
QFG06635.1	hypothetical protein CPT_Myduc_012	*Proteus* phage Myduc	4.9282199
A0A377K133	Large ribosomal subunit protein uL2	*Escherichia coli*	-7.1046319
QNL31226.1	hypothetical protein A2_158	*Enterococcus* phage vB_EfaM_A2	0.8432144
WP_172503165.1	cyclic pyranopterin monophosphate synthase MoaC	*Enterococcus faecalis*	8.6098750
A0A7H4N9H1	Cytochrome b561 homolog 2	*Klebsiella* sp.	-0.1388827
Q9AZ61	MHP	*Lactococcus* phage	32.4863461

Supplementary Table 2. Phage proteins selected through the lasso feature selection algorithm discriminated UTI cases from non-UTI controls.

Accession ID	Description	Organism	Lasso coefficient
A0A0K1LKF6	Membrane lipoprotein	*Vibrio* phage H188	-0.02757225
A0A1Y3FEC6	Host factor I for bacteriophage Q beta replication, a growth-related protein	*Acinetobacter seifertii* phage	2.80029532
A0A2H4YGX6	Uncharacterized protein KMC53_gp11	*Raoultella* phage Ro1	3.56490034
A0A564WHB1	Bacteriophage Mu tail sheath	*Candidatus Defluviicoccus seviourii*	0.04912609
DAQ14039.1	MAG TPA: endosialidase chaperone	Bacteriophage sp.	-0.17551566
I6S6X7	Virion structural protein	*Croceibacter* phage P2559S	2.32091329
Q2NP83	dUTP diphosphatase	*Xanthomonas* phage OP2	1.14401082
Q9AZ61	N4-gp56 family major capsid protein	*Lactococcus* phage TP901-1	0.24010885
QFG06635.1	hypothetical protein HWC59_gp12	*Proteus* phage Myduc	0.26950152
QXV77638.1	hypothetical protein bas50_0172	*Escherichia* phage DrSchubert	0.47952972
S5MLP4	Tail tape measure domain-containing protein OS	*Pseudoalteromonas* phage TW1	0.15105967

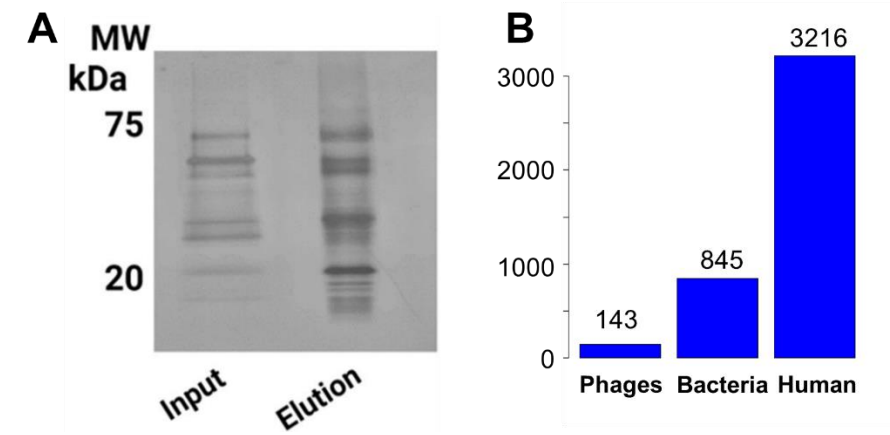
Supplementary Table 3. Bacterial proteins selected through the lasso feature selection algorithm discriminated UTI cases from non-UTI controls.

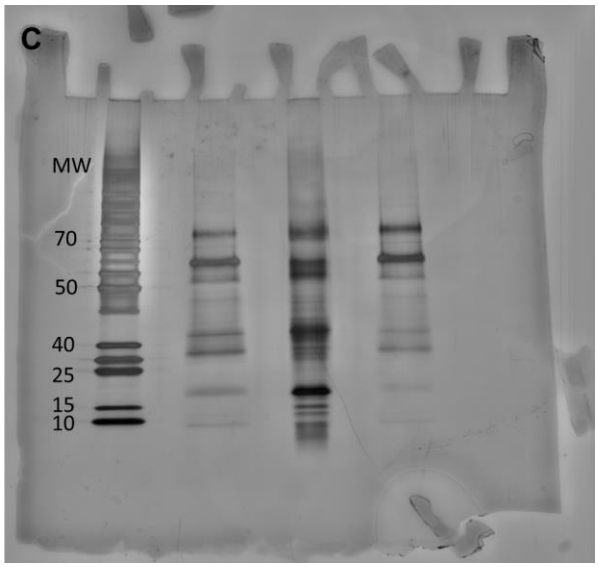
Accession ID	Description	Organism	Lasso coefficient
WP_069988041.1	pneumococcal-type histidine triad protein	*Streptococcus sp.*	-0.47368837
OAO50651.1	beta-D-glucuronidase	*Escherichia coli*	0.92165314
QNQ85507.1	L-fucose isomerase	*Escherichia coli*	-0.67355347
WP_172503165.1	cyclic pyranopterin monophosphate synthase MoaC	*Enterococcus faecalis*	38.88553697
A0A7H4N9H1	Cytochrome b561 homolog	*Klebsiella sp.*	-0.29164063
P0ABU7	Biopolymer transport protein ExbB	*Escherichia coli*	1.12452487
A0A9X4PSK2	Phage major capsid protein	*Pseudomonas aeruginosa*	67.00775933
A0A485BZA0	Betaine aldehyde dehydrogenase	*Klebsiella sp.*	1.88312197
Q9I138	Serine hydroxymethyltransferase 2	*Pseudomonas aeruginosa*	-15.16861891
A0A377K133	Large ribosomal subunit protein uL2	*Escherichia coli*	-59.84742726
P17315	Colicin I receptor	*Escherichia coli*	3.96016328

Supplementary Table 4. T tests on the gene abundances in *E. coli* B and UPEC CFT073 [WAM2267] grown in presence of non-UTI and UTI urine, and in presence or absence of mitomycin C. Gene = gene name, Comparison = experimental condition compared, T statistic = t, p_value = t test p value, Adjusted_p_value = Benjamini–Hochberg adjusted p values, Sig at 0.05 = an asterisk is reported if the adjusted p value is below 0.05. Colors are lined to the bargraph in Fig. 6J.

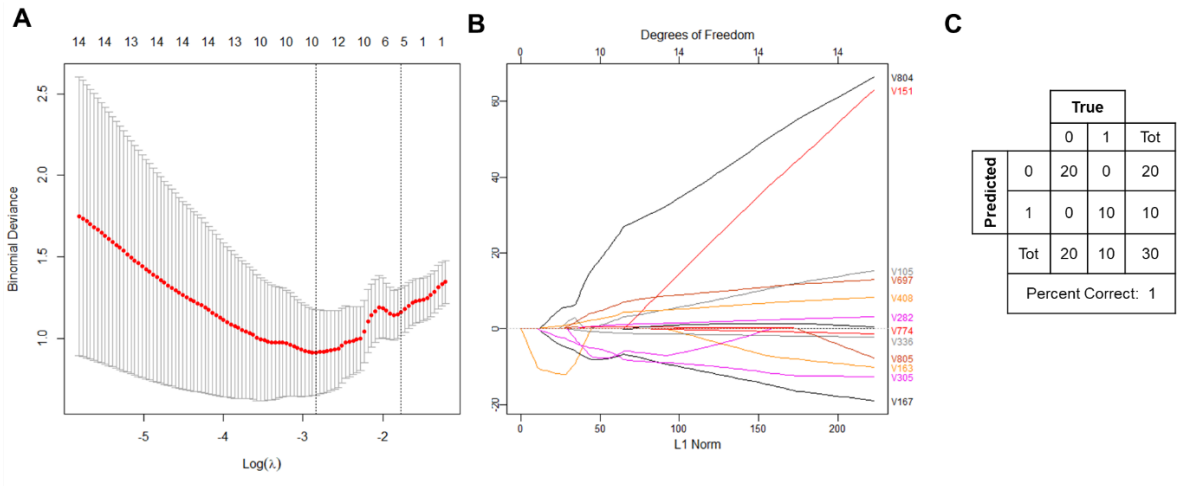
Gene	Comparison	T_Statistic	P_Value	Adjusted_P_Value	Sig at 0.05
RecA	N-E.coli-MitoC vs N-E.coli	8.18	4.58E-05	6.59E-05	*
RecA	U-E.coli vs N-E.coli	22.45	8.77E-10	1.58E-08	*
RecA	U-E.coli-MitoC vs N-E.coli	16.62	2.07E-08	1.24E-07	*
RecA	N-CFT073-MitoC vs N-CFT073	-2.81	0.016	0.018	*
RecA	U-CFT073 vs N-CFT073	1.20	0.256	0.271	
RecA	U-CFT073-MitoC vs N-CFT073	0.82	0.432	0.445	
RecA pr	N-E.coli-MitoC vs N-E.coli	17.71	1.85E-08	1.24E-07	*
RecA pr	U-E.coli vs N-E.coli	16.66	7.91E-07	2.03E-06	*
RecA pr	U-E.coli-MitoC vs N-E.coli	9.71	3.81E-05	5.71E-05	*
RecA pr	N-CFT073-MitoC vs N-CFT073	8.27	1.40E-05	2.51E-05	*
RecA pr	U-CFT073 vs N-CFT073	9.37	2.48E-05	3.94E-05	*
RecA pr	U-CFT073-MitoC vs N-CFT073	11.35	7.21E-07	2.00E-06	*
Trans	N-E.coli-MitoC vs N-E.coli	13.59	1.53E-08	1.24E-07	*
Trans	U-E.coli vs N-E.coli	22.79	3.19E-11	1.15E-09	*
Trans	U-E.coli-MitoC vs N-E.coli	12.16	2.69E-07	8.80E-07	*
Trans	N-CFT073-MitoC vs N-CFT073	11.05	7.64E-06	1.45E-05	*
Trans	U-CFT073 vs N-CFT073	15.37	1.01E-06	2.42E-06	*
Trans	U-CFT073-MitoC vs N-CFT073	6.12	0.0004	0.0005	*
XerC	N-E.coli-MitoC vs N-E.coli	11.35	3.40E-06	6.81E-06	*
XerC	U-E.coli vs N-E.coli	16.40	2.64E-07	8.80E-07	*
XerC	U-E.coli-MitoC vs N-E.coli	12.64	3.08E-06	6.53E-06	*
XerC	N-CFT073-MitoC vs N-CFT073	10.27	3.01E-07	9.04E-07	*
XerC	U-CFT073 vs N-CFT073	13.19	2.70E-08	1.39E-07	*
XerC	U-CFT073-MitoC vs N-CFT073	1.86	0.088	0.096	
XerD pr	N-E.coli-MitoC vs N-E.coli	9.12	2.10E-05	3.60E-05	*
XerD pr	U-E.coli vs N-E.coli	18.04	6.86E-08	3.09E-07	*
XerD pr	U-E.coli-MitoC vs N-E.coli	19.90	3.57E-09	4.29E-08	*
XerD pr	N-CFT073-MitoC vs N-CFT073	6.61	2.52E-05	3.94E-05	*
XerD pr	U-CFT073 vs N-CFT073	10.66	1.84E-07	7.35E-07	*
XerD pr	U-CFT073-MitoC vs N-CFT073	4.89	0.0004	0.0005	*
CysG	N-E.coli-MitoC vs N-E.coli	-6.05	5.74E-05	7.94E-05	*
CysG	U-E.coli vs N-E.coli	2.82	0.022	0.024	*
CysG	U-E.coli-MitoC vs N-E.coli	0.18	0.864	0.864	
CysG	N-CFT073-MitoC vs N-CFT073	4.36	0.001	0.002	*
CysG	U-CFT073 vs N-CFT073	9.43	1.43E-06	3.21E-06	*
CysG	U-CFT073-MitoC vs N-CFT073	-3.68	0.003	0.004	*

Supplementary Figures

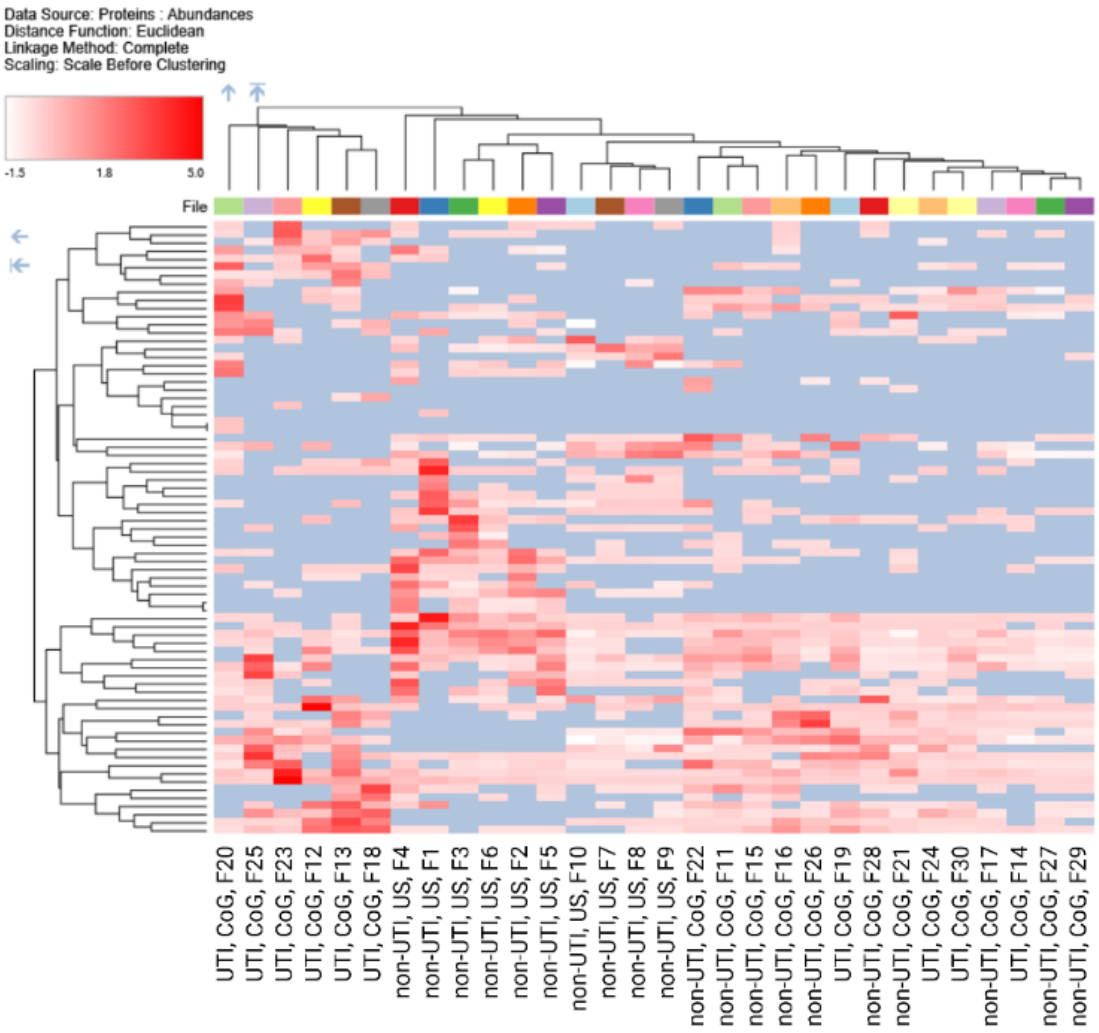




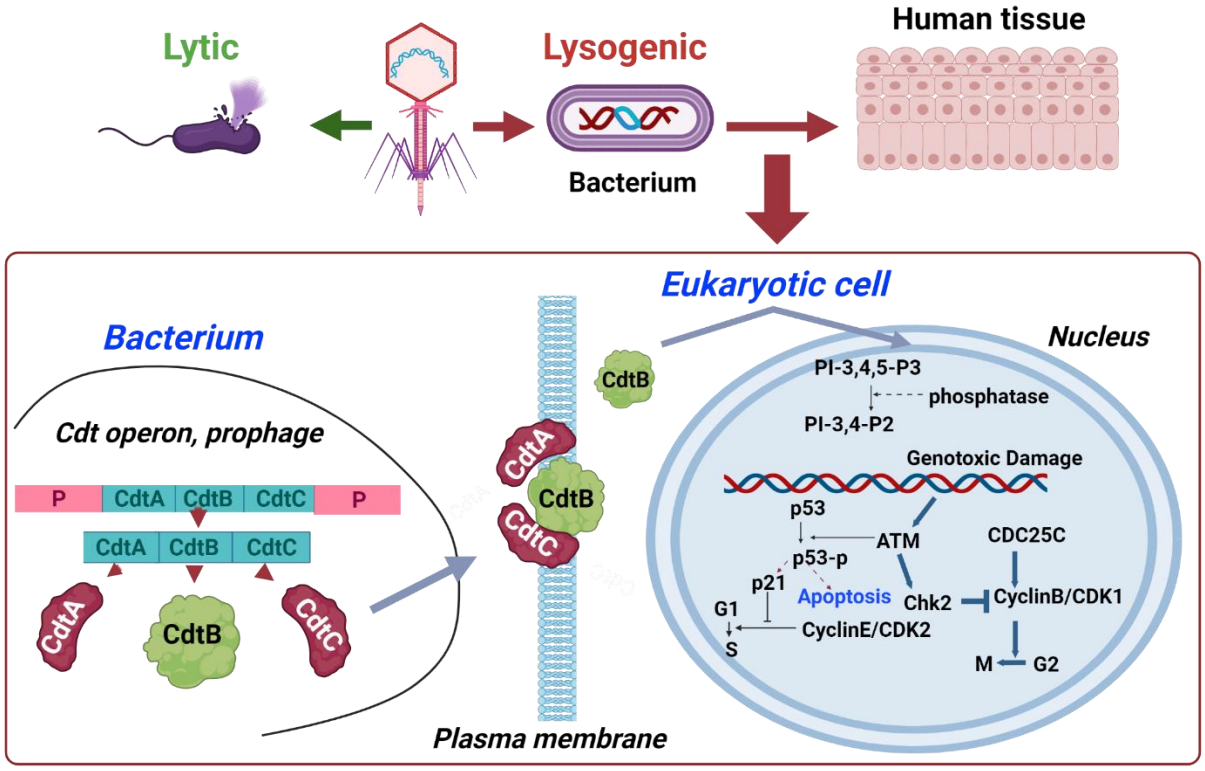
Supplementary Figure 1. SDS PAGE and mass spectrometry analysis of urinary proteins. (A) Silver stained SDS PAGE analysis of proteins in the urine (Input) and proteins captured by the poly(NIPAm/AA/RB221) affinity hydrogel particles (Elution). The image shows protein enrichment at the low molecular weight. (B) Bar graph of the number of phage, bacteria and human proteins identified by affinity enhanced mass spectrometry analysis of urine. (C) Uncropped image of the SDS PAGE gel in panel (A).



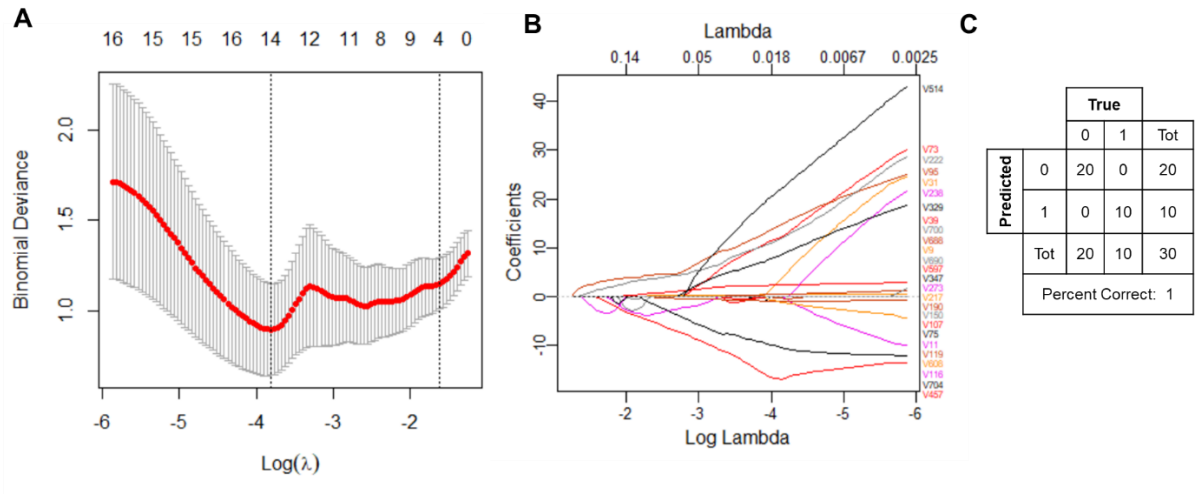
Supplementary Figure 2. Lasso model using bacteria and phage combined proteins. The Lasso algorithm was used to select variables that best discriminate between UTI cases and non-UTI controls. **A)** Log-transformed regularization parameter lambda is plotted against the binomial deviance, which represents the goodness of fit of the model. The vertical dashed line represents the value of lambda that minimizes the cross-validated binomial deviance. **B)** L1 norm (X axis), which is the sum of the absolute values of the protein vector components, is plotted versus the coefficients of the model (Y axis). Each curve represents the trajectory of a coefficient as lambda changes. As lambda increases, more coefficients shrink towards zero, performing variable selection. The points where the trajectories hit the x-axis indicate that the corresponding variables are excluded from the model for higher values of lambda. **C**) Confusion matrix for discrimination performance of the model (lambda = 0.02) on the training set.



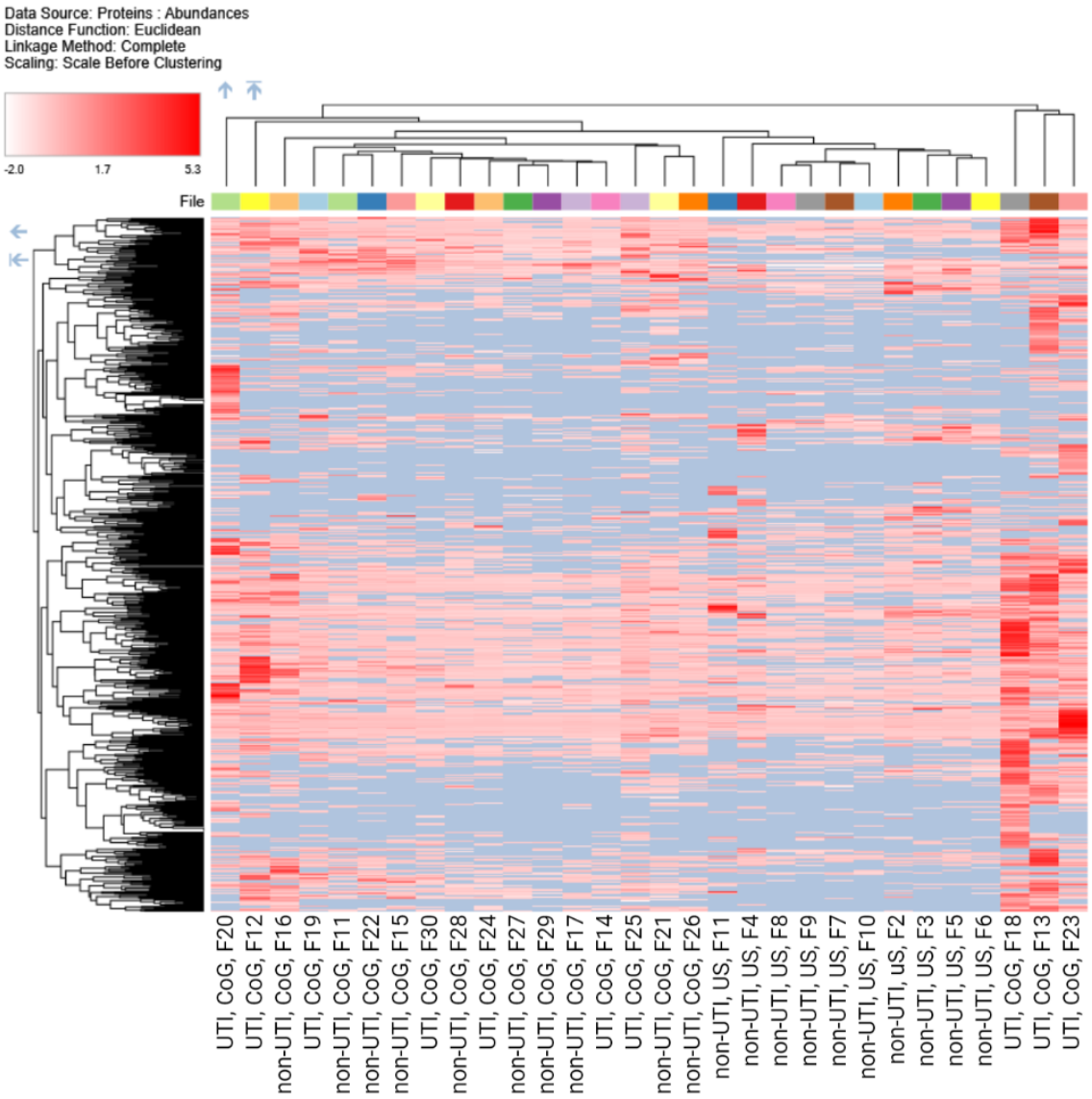
Supplementary Figure 3. Heatmap and two-way hierarchical clustering of bacteriophage protein abundances in the patient urine samples.



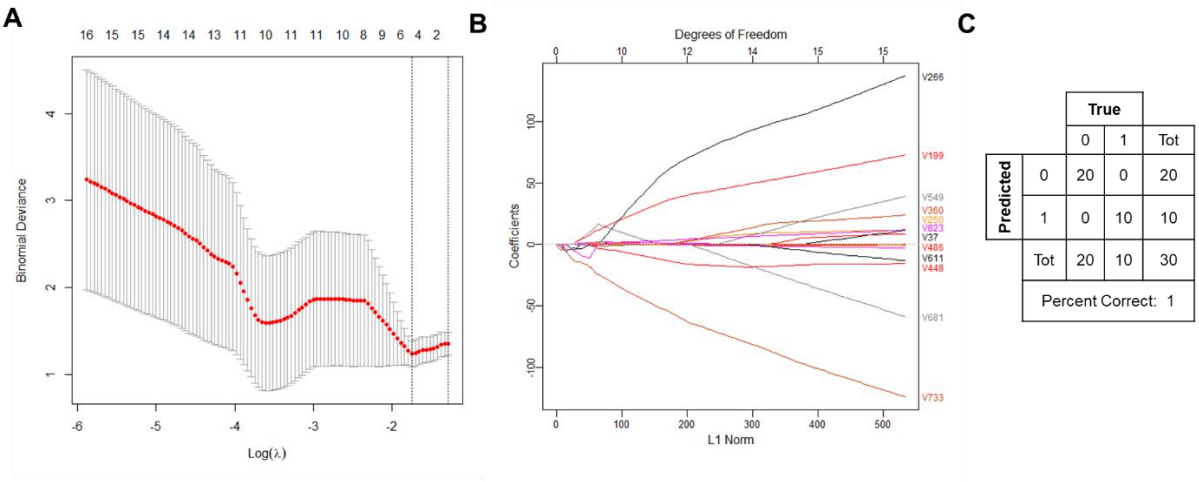
Supplementary Figure 4. Example of a symbiotic phage-bacteria mechanism revealed by urinary proteomics. The cdt toxin gene encodes for the cytolethal distending toxin (CDT), a protein toxin that exhibits cytotoxic effects on mammalian cells. CDT is a potent virulence factor that can cause mammalian cell cycle arrest, DNA damage, and cellular distension. The genes responsible for Cdt production are typically located within the bacterial chromosome and plasmids, but also in bacteriophages like in the present case.



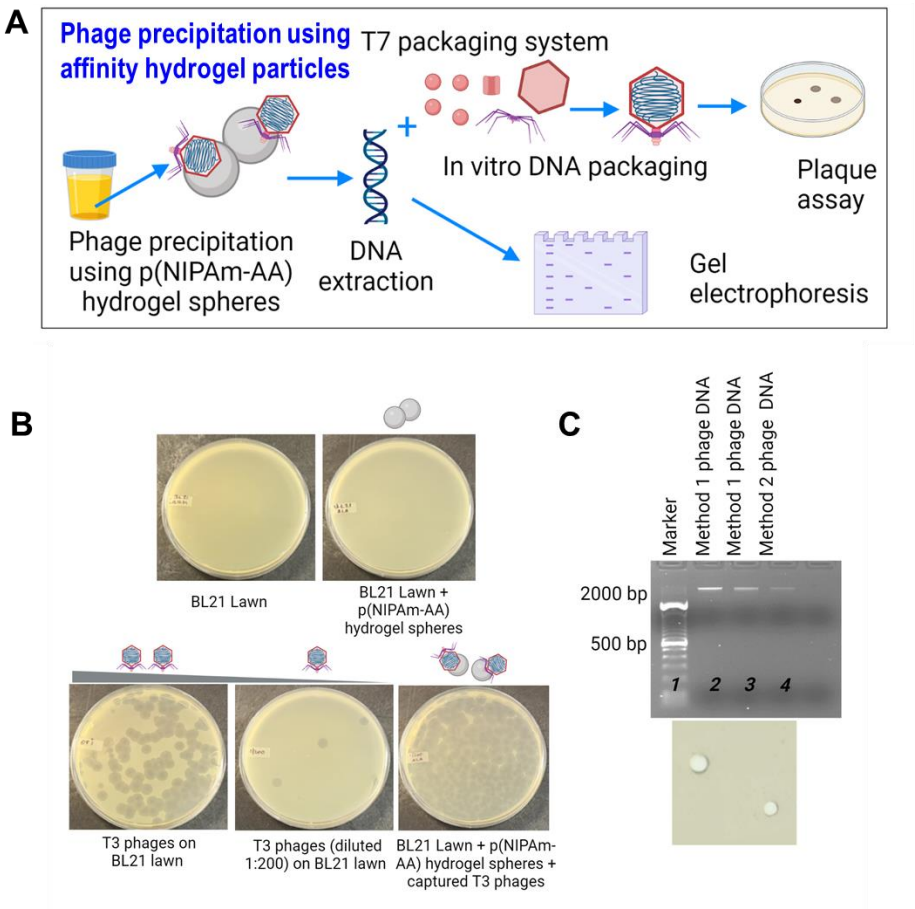
Supplementary Figure 5. Lasso model on phage proteins. The Lasso algorithm was used to select variables that best discriminate between UTI cases and non-UTI controls. **A)** Log-transformed regularization parameter lambda is plotted against the binomial deviance, which represents the goodness of fit of the model. The vertical dashed line represents the value of lambda that minimizes the cross-validated binomial deviance. **B)** L1 norm (X axis), which is the sum of the absolute values of the protein vector components, is plotted versus the coefficients of the model (Y axis). Each curve represents the trajectory of a coefficient as lambda changes. As lambda increases, more coefficients shrink towards zero, performing variable selection. The points where the trajectories hit the x-axis indicate that the corresponding variables are excluded from the model for higher values of lambda. **C**) Confusion matrix for discrimination performance of the model (lambda = 0.02) on the training set.



Supplementary Figure 6. Heatmap and two-way hierarchical clustering of UTI pathogen protein abundances in the patient urine samples.



Supplementary Figure 7. Lasso model on bacterial proteins. The Lasso algorithm was used to select variables that best discriminate between UTI cases and non-UTI controls. **A)** Log-transformed regularization parameter lambda is plotted against the binomial deviance, which represents the goodness of fit of the model. The vertical dashed line represents the value of lambda that minimizes the cross-validated binomial deviance. **B)** L1 norm (X axis), which is the sum of the absolute values of the protein vector components, is plotted versus the coefficients of the model (Y axis). Each curve represents the trajectory of a coefficient as lambda changes. As lambda increases, more coefficients shrink towards zero, performing variable selection. The points where the trajectories hit the x-axis indicate that the corresponding variables are excluded from the model for higher values of lambda. **C**) Confusion matrix for discrimination performance of the model (lambda = 0.02) on the training set.



Supplementary Figure 8. Phage isolation using affinity hydrogel particles achieves 82% phage recovery efficiency. (A) Aliquots of 200 µL of poly(NIPAm/AA) particles (10 mg/mL polymer dry weight) were mixed with 10 mL of urine and allowed to incubate for 45 mins at room temperature. Particles were collected by centrifugation at 16.1 rcf for 20 minutes, washed in PBS and resuspended in 100 µL of sterile water. Phage DNA was extracted by incubating the particles with 100 µL of phenol/chloroform/isoamyl alcohol (ratio 25/24/1); samples were mixed by inverting the tube few times and allowed to incubate on ice for 5 minutes. After incubation, the samples were spun at 16.1 rcf for 20 minutes. The top aqueous solution was placed in a new tube and the procedure was repeated once more to ensure complete removal of protein contaminants. DNA from aqueous solution was precipitated with equal volume of 5 M ammonium acetate and two volumes of ice-cold 200 proof ethanol at -20°C for 1 hour. Tubes were spun at 16.1 rcf for 20 min, and DNA pellet was washed with 75% ice-cold ethanol, spun 16.1 rcf for 10 minutes and pellet was air-dried and resuspended in 20 µL of water. DNA concentration was measured using NanoDrop® ND-1000 UV-Vis Spectrophotometer (Thermo). The DNA was re-packaged in phage shells using the T7Select packaging system (Novagen). Briefly, 5 µL of extracted DNA at concentration 0.5 µg/ µL was mixed with 25 µL of T7Select Packaging Extract and allowed to incubate at room temperature for 2 hours. The reaction was stopped by adding 270 µL of LB medium. The mixture was plated for plaque assay using the E. coli BL21 bacterial strain. (B) Hydrogel particle capability to capture phages in suspension was confirmed using T3 phages spiked in LB medium. Representative plates are reported for: LB21 lawn, clean hydrogel particles plated alone, T3 phages plated alone, T3 phages diluted 1:200,

hydrogel particles allowed to incubate with T3 phage suspension for 45 minutes. (C) An agar gel shows DNA from phage clinical isolates. Lanes:1-marker, 2- and 3- clinical isolate phage DNA (Figure 5), 4- DNA precipitated from urine (Figure 5). In the bottom panel, a plaque assay of urine DNA packaged into T7 phage shells on susceptible E. coli BL21 lawn yielded plaques.