

“Characterization of bacteriophages infecting uropathogenic *Escherichia coli* and identification of factors in their genomes with both applied and fundamental scientific significance”

Wojciech Wesołowski, M.Sc.

Urinary tract infections (UTIs) are among the most common bacterial infections worldwide, constituting the second most frequent group of infections after those affecting the respiratory tract. They affect approximately 150–250 million people annually, represent about 30% of all hospital-acquired infections, and impose a substantial medical and socioeconomic burden, with the annual cost of uncomplicated UTIs in the United States alone estimated at approximately \$2 billion. Although UTIs occur in both sexes, women are disproportionately affected, with more than half of women worldwide experiencing at least one episode during their lifetime.

The predominant etiological agent of UTIs is uropathogenic *Escherichia coli* (UPEC), which accounts for approximately 75–85% of uncomplicated and 50–65% of complicated infections, and together with other members of the order *Enterobacterales*, notably *Klebsiella pneumoniae*, constitutes most cases. The predominance of *E. coli* among uropathogens stems primarily from its ability to evade the host immune response and to colonize the environment of the urinary tract. Importantly, UPEC does not constitute a monolithic group, as it comprises strains that differ substantially in virulence and resistance profiles. The genome of *E. coli* can be divided into a conserved core genome, encoding vital functions, and a highly flexible accessory portion comprising mobile genetic elements such as plasmids, prophages, and genomic islands, whose acquisition or loss, mediated by horizontal gene transfer, shapes colonization, pathogenicity, and antibiotic resistance.

The pathogenicity of UPEC is determined by a diverse repertoire of virulence determinants rather than by a single universal factor set. Genomes of these bacteria contain genes involved in adhesion, enabling attachment to epithelial cells and colonization, including type 1, P, and S fimbriae, Dra/Afa adhesins, and antigen 43. They also carry genes encoding toxins that modulate immune responses and damage host cells, such as α -hemolysin, cytotoxic necrotizing factor 1 (CNF1), and the uropathogenic-specific protein (USP), as well as genes involved in siderophore biosynthesis and transport, which facilitate efficient iron acquisition. Additional virulence-associated genes and proteins include invasion-associated factors, autotransporter proteins, type VI

secretion system components, and products involved in capsule and lipopolysaccharide biosynthesis, which contribute to immune evasion and inhibition of phagocytosis.

Biofilm formation constitutes a crucial stage in the pathogenesis of UTIs, enabling UPEC to persist and expand effectively within the urinary tract. Biofilm development is facilitated by structures such as type 1 and P fimbriae, which recognize receptors on the surface of host epithelial cells and initiate production of an extracellular matrix that protects the bacteria from antibiotics and immune cells. The biofilm matrix additionally harbours persister cells, which are capable of surviving immune system attacks and antibiotic therapies and permit immediate regrowth of the bacterial population after antibiotic treatment, rendering biofilm eradication extremely challenging and infection relapses frequent. Biofilms may also form on abiotic surfaces such as urinary catheters, which makes treatment of catheter-associated UTIs particularly difficult, as prolonged catheterization increases the risk of infection by approximately 5% per day. In the hospital environment, bacterial biofilm serves as a reservoir of chronic infection and a site of intensive intercellular gene transfer, promoting the selection and dissemination of antibiotic resistance.

Antimicrobial resistance among uropathogens is a continuously growing problem, driven largely by the incorrect use of antibiotics, and it generates high treatment costs. Reported resistance rates vary markedly by geographic region and antibiotic class. Resistance to trimethoprim-sulfamethoxazole among UPEC isolates in Europe spans approximately from 14.6 to 60%, whereas resistance to amoxicillin-clavulanate ranges from 5.3% in Germany to 37.6% in France, with even higher levels reported in developing countries. Resistance to second- and third-generation cephalosporins is also increasing in Europe, and studies of clinical UPEC isolates have documented resistance to third-generation cephalosporins and amoxicillin-clavulanate exceeding 70–85% in some regions. Extensive use of fluoroquinolones, particularly ciprofloxacin, has contributed to a steady rise in resistance to this class, which is markedly more prevalent in low- and middle-income regions (in the range of 55.5–85.5%) than in highly developed countries (in the range of 5.1–32.0%). On the other hand, UPEC isolates generally remain susceptible to carbapenems and nitrofurantoin, the latter still constituting first-line therapy depending on local resistance patterns and infection type. The scale of the problem is reflected in the inclusion of *E. coli* among the pathogens monitored by the

European Centre for Disease Prevention and Control (ECDC), as well as in data from the WHO Global Antimicrobial Resistance and Use Surveillance System (GLASS), which indicate a substantial burden of antimicrobial resistance among *E. coli* and *K. pneumoniae*, both of which are frequent causes of bloodstream infections. Given the increasing prevalence of antimicrobial resistance among UPEC, there is an urgent need for alternative therapeutic solutions.

Among alternative strategies against infections caused by multidrug-resistant UPEC, bacteriophage therapy is a promising approach. Lytic bacteriophages offer several inherent advantages over conventional antimicrobials: (I) they can destroy bacterial biofilms, which are resistant to antibiotic treatment, (II) they replicate at the site of infection, thereby providing self-amplifying antibacterial agents, and (III) they kill pathogens selectively, sparing the commensal microbiota. Nevertheless, the therapeutic application of phages remains challenging. Only phages with a strictly lytic life cycle are suitable for therapy, and bioinformatic predictions of the phage lifestyle often require experimental verification. Phage activity may be reduced by the host immune response, an aspect that remains rarely investigated. Also, the narrow strain specificity of many phages precludes their broad, universal application.

Importantly, phage genomes also represent a valuable source of genetic information that may reveal proteins and other genomic features with both applied and fundamental scientific significance. On the one hand, gene products encoded in phage genomes can possess useful biological activities such as the lysis of bacterial cells or the degradation of bacterial surface structure and thus serve as a source of new antimicrobial agents. On the other hand, phage genomic sequences provide a rich resource for studying phage diversity and evolution, allowing the identification of genomic features that can improve the reconstruction of evolutionary relationships. Thus, systematic exploration of phage genomes can simultaneously contribute to the discovery of novel application-oriented proteins and to a better understanding of the molecular and evolutionary characteristics of bacteriophages.

In this dissertation, we focused on bacteriophages infecting UPEC bacteria. The study encompassed both previously described phages, whose genome sequences were available in the GenBank database, and the isolation of novel phages followed by their comprehensive characterization, including morphological and physiological analyses,

assessment of their stability and safety toward human cells, as well as genome sequencing, assembly, and detailed genomic analysis. We also performed phylogenetic analyses of the investigated phages to determine their relatedness to other viruses infecting UPEC. The genomic analyses were focused on identifying genes and genomic features of potential scientific and applied significance. Particular attention was given, on the one hand, to the identification of genes encoding enzymatic proteins, including proteins involved in bacterial cell lysis and putative depolymerases, which may represent a source of novel antimicrobial tools. On the other hand, the study aimed to identify genes with a strong phylogenetic signal to improve the quality and reliability of evolutionary relationship reconstructions, even for closely related phages infecting UPEC.

In **article no. 1**, we performed an analysis of 98 bacteriophages, and we demonstrated the considerable therapeutic potential of these viruses for combating UPEC. We found that most of these isolates, specifically 95%, were evaluated for host range, establishing a foundational understanding of their spectrum of activity against various bacterial strains. Our study identified seven phages: vB_Ecos_Golestan, myPSH1131, Killian, 590b, FS2B, UGKSEcP1, and UGKSEcP2, that stand out as well-characterized candidates for further research towards phage therapy development. These specific phage strains underwent detailed physiological and bioinformatic profiling, including assessments of host range, stability, development kinetics, and genome content. Notably, our examination of phage FS2B revealed its strong lytic activity, a high burst size, and exceptional stability across a wide temperature and pH range, alongside a confirmed lytic replication cycle and lack of lysogeny-related genes. Furthermore, we observed that this phage demonstrated synergistic interactions with tested antibiotics, suggesting that with continued refinement, such a well-documented candidate could become an effective, personalized treatment option for recurrent urinary tract infections.

Bioinformatic analyses further demonstrated substantial genomic diversity among UPEC-infecting phages, with genome sizes ranging from 39,091 to 169,905 bp and the number of predicted ORFs ranging from 49 to 305. Genome annotation and comparative analyses were performed to investigate phage protein functions. Notably, eight previously sequenced but unannotated phage genomes were annotated for the first time in this work. Among all 792 unique phage proteins gathered in Supplementary Tables S2 and S3, 79 interesting phage proteins were identified by us in the newly annotated anti-UPEC phage

genomes. A particular focus was placed on the identification of phage-encoded enzymes with potential application in antimicrobial strategies, including proteins involved in bacterial cell lysis and putative depolymerases identified through structure-based bioinformatic analysis. A total of 40 classical endolysins, 75 holins, 40 spanins, and 16 putative depolymerases were identified in the analysed anti-UPEC phage genomes. To further characterize phage developmental potential, life-cycle predictions were performed, providing newly obtained results and revealing several phages requiring further experimental verification because of potentially temperate characteristics. Furthermore, we developed a multi-step approach for the phylogenetic analysis of the phages under investigation. In the first step, the 82 phages were grouped into clades based on their preliminary similarity and evolutionary relatedness, as determined using a proteomic tree generated with ViPTree. Subsequently, orthologous gene groups were identified using OMA. In the next step, individual multiple sequence alignments of the encoded proteins were evaluated based on the number of parsimony-informative sites. On this basis, genes providing phylogenetic information were selected and used to construct multilocus phylogenetic trees. These included, among others, genes encoding tail proteins, portal proteins, and proteins involved in genome replication. Highly conserved sequences were excluded from the analysis.

Collectively, these newly performed analyses substantially expanded the available genomic characterization of anti-UPEC phages and provided new information on their functional potential, developmental characteristics, and evolutionary relationships. Importantly, screening performed using multiple bioinformatic tools consistently indicated the absence of virulence factors and antibiotic-resistance genes in the analysed phage genomes. These findings demonstrate the value of comprehensive genome analysis for assessing phage safety, developmental characteristics, and evolutionary relationships before therapeutic application.

Despite these promising findings, we identified substantial methodological gaps and inconsistencies that currently hinder the clinical translation of phage therapy. Our review (**article no. 1**) revealed that critical safety data is largely missing; notably, none of the 98 analysed phages were tested for cytotoxicity toward eukaryotic cells. Furthermore, only 42% of the phages had their efficiency of plating (EOP) values determined, a critical omission that leads to inaccurate efficacy assessments. In turn, stability studies were

conducted on only 20% of isolates regarding temperature and 19% regarding pH. Additionally, 15% of the phages were not sequenced at all, and others lacked proper annotation or phylogenetic analysis. To address these deficiencies, we conclude that future research must adopt standardized, comprehensive characterization protocols. Specifically, we advocate for mandatory EOP assays, thorough cytotoxicity screening, stability testing under physiologically relevant conditions, and rigorous bioinformatic verification of genomes for toxin or antibiotic-resistance genes to assess the potential of the investigated phages for application in phage therapy.

Building directly upon the limitations we identified in **article no. 1.**, we sought to address these shortcomings through our own experimental work, as presented in **article no. 2.** In that study, we isolated and comprehensively characterized several novel lytic bacteriophages infecting clinical UPEC strains, and we formally proposed five novel phage species: *Hanrivervirus* vB_EcoS-149_4M, *Warwickvirus* vB_EcoS-149_6D, *Nonavirus* vB_EcoS-164_2, *Kagunavirus* vB_EcoS-179_3, and *Vectrevirus* vB_EcoP-179_5. Crucially, we went far beyond the descriptive reporting that dominates the literature: rather than merely recording host range by spot test, we performed EOP analysis and integrated phage characterization with a comparative genomic analysis of the bacterial hosts themselves. Moreover, we performed a broad assessment of our candidates: we confirmed the absence of toxin and antibiotic-resistance genes in the phage genomes, documented the presence of putative lytic protein genes encoding endolysins, holins, and spanins in all newly identified phages.

To determine phylogenetic positions of the tested phages, we refined the approach used in the review **article no. 1.** The genomic distances among the analyzed phages were determined using VIRIDIC, followed by reconstruction of their phylogenetic relationships based on multilocus phylogenetic trees. Orthologous gene groups identified using OMA were used for these analyses. Following multiple sequence alignment and assessment of their phylogenetic signal, sequences with an insufficient number of parsimony-informative sites were excluded, while the remaining sequences were used for tree reconstruction. The genes selected for phylogenetic reconstruction included, among others, those encoding tail proteins, the small and large terminase subunits, and proteins involved in cell lysis, including holins and spanins.

Furthermore, we demonstrated that the biological properties and safety profiles of the investigated phages substantially influence their potential applicability in phage therapy. The phage vB_EcoS-149_4M, characterized by the broadest host range, high stability under the tested conditions, and a favorable safety profile, including the absence of genes encoding toxins and antibiotic resistance determinants, as well as no detectable cytotoxic activity against the human cell lines examined, represents a promising candidate for further application-oriented studies. In contrast, phages vB_EcoS-179_3 and vB_EcoP-179_5 exhibited a very narrow host range, restricted to infection of strain EC179, indicating that they may be more suitable for personalized approaches than for broad-spectrum applications.

Although the identification of five novel species of anti-UPEC bacteriophages expanded the previously described pool of 98 phages capable of infecting these pathogens, our results indicate that the number of characterized lytic phages infecting UPEC remains insufficient relative to the high abundance and substantial genotypic and phenotypic diversity of this bacterial group. A particularly illustrative example is strain EC256, for which no bacteriophage capable of infecting it was identified in the present study.

Next, we extended this multi-aspect characterization strategy even further in **manuscript no. 3** about bacteriophage vB-EcoS_57-3 and its phage-derived endolysin 57_3Lys. In this work, we isolated the phage from urban sewage, characterized its siphovirus-like morphology and plaque phenotype, suggesting depolymerase activity. We fully sequenced and annotated its 45,774 bp genome comprising 91 predicted ORFs and firmly established its taxonomic position as a new representative of the genus *Veterinaerplatzvirus* within the family *Drexelviriidae*. Rather than stopping at descriptive features, we quantified the full phage development cycle: we determined a latent period of 11 minutes, a burst size of approximately 42 PFU per cell, moderately fast adsorption reaching 80% within 10 minutes, and highly efficient lytic development reducing viable cell counts dramatically within 30 minutes. Importantly, we tested our phage under conditions relevant to the intended clinical application: stability and activity in artificial urine at pH 4.5, 5.8, and 6.5, where the phage remained stable and reduced bacterial numbers to as little as 10% of the initial inoculum after 24 hours. Additionally, we

excluded adverse effects of the phage lysate on mammalian cells by demonstrating no cytotoxicity against human bladder (T24) and kidney (HEK293) lines.

Following the characterization of bacteriophages, we turned our attention to the proteins they encode, since phage-derived lytic enzymes represent a complementary and versatile class of antibacterial agents. As we summarized in **article no. 4**, endolysins are peptidoglycan-degrading enzymes produced during the phage lytic cycle, and they offer several advantages such as rapid bacterial killing, a low probability of resistance development, synergistic activity with other antimicrobial agents, and effectiveness against both planktonic cells and biofilms. Through genetic engineering, their spectrum of activity can even be deliberately modified, for instance by fusing them with outer-membrane-disrupting components, cell wall-binding domains for targeting specificity, or antimicrobial peptides. Nevertheless, reports on endolysins directly tested against uropathogens, especially UPEC, are limited. The central obstacle is that the outer membrane of Gram-negative bacteria severely restricts access of externally applied enzymes to peptidoglycan. Additional challenges include the risk of microbiota disruption resulting from an overly broad antibacterial activity spectrum, purification difficulties, and the fact that low concentrations of endolysins may alter bacterial phenotypes in ways that could be unfavorable for the patient, underscoring the need for thorough safety evaluations.

In our search for the endolysins described in **article no. 4**, we analyzed the genomes of 75 lytic anti-UPEC phages available in the GenBank database. This analysis enabled us to identify 39 putative endolysins in 30 phage genomes. The identified enzymes included both proteins previously described by other researchers and proteins reported by us for the first time. The *in silico* analysis revealed considerable diversity among the putative endolysins with respect to their predicted enzymatic activities. The identified proteins were classified into five groups comprising enzymes with lysozyme, amidase, and peptidase domains, as well as proteins with predicted transglycosylase and hydrolase activities. The physicochemical properties of the identified endolysins indicated that their molecular masses ranged from approximately 8.7 to 22.6 kDa, while their isoelectric points indicated a predominantly basic character. The instability index values suggested good stability for most of the analyzed proteins, whereas the high aliphatic index values indicated their potential thermal stability. Negative GRAVY values, in turn,

indicated the hydrophilic nature of these proteins and their predicted solubility in water. In addition, multiple antimicrobial peptide (AMP) sequences were identified within the protein sequences, suggesting their potential lytic properties.

Particularly noteworthy was our finding that the genomes of seven phages contained large proteins ranging from 120 to 145 kDa with endolysin-like domains. These proteins are likely involved in facilitating the injection of phage genetic material into the host cell at the initial stage of infection rather than in the final lysis of the bacterial cell. Furthermore, our phylogenetic and functional analyses revealed issues concerning the correct annotation of several putative endolysins previously described by other researchers. For several proteins originally annotated as endolysins, analysis of the identified domains indicated that they may have been incorrectly classified. For example, proteins derived from phages Killian and ZCEC14 were found to contain domains characteristic of anaerobic ribonucleoside-triphosphate reductase and T4 RNase H, respectively, as well as a protease inhibitor domain associated with bacteriophage head maturation. The presence of these domains, together with the results of phylogenetic classification, suggested that the analyzed proteins may perform functions other than peptidoglycan degradation and are therefore unlikely to represent endolysins. These findings highlight the need for thorough verification of proteins annotated as phage endolysins.

The continuation of this bioinformatic groundwork came in our subsequent experimental study in **manuscript no. 3**. In this work, we performed functional validation of the vB-EcoS_57-3 phage endolysin, called 57_3Lys. *In silico*, we established that 57_3Lys is a small 161-amino-acid protein of 17.55 kDa. Based on 57_3Lys domain analysis, we found that it is a class II SAR endolysin, exemplified by the endolysin R21 of lambdoid phage 21. We identified its N-terminal, moderately hydrophobic 20-residue segment as a signal-anchor-release domain, its catalytic E(X)₈-D(X)₅-T triad at residues E26, D35, and T41, and a cationic, amphipathic antimicrobial peptide-like region between residues 110 and 140. Functionally, we cloned and overproduced the enzyme, purified it via affinity chromatography to the *in silico* predicted molecular weight. Next, using sodium azide (NaN₃), an inhibitor of the Sec translocase in *E. coli*, we demonstrated a concentration-dependent inhibition of 57_3Lys-induced bacterial cell lysis. At a concentration of 5 mM NaN₃, the lytic effect was markedly reduced, whereas at 10 mM it was almost completely

abolished. Combined with the presence of a predicted N-terminal SAR domain and the characteristic E(X)₈-D(X)₅-T motif of class II SAR endolysins, these results provided convergent evidence for the involvement of the Sec pathway in the translocation of 57_3Lys into the periplasmic space. Besides, we showed that purified 57_3Lys, applied externally, significantly reduced both the turbidity and the viable cell count of clinical EC57, a UPEC strain without any added outer-membrane permeabilizer. This activity developed slowly, becoming significant after 24 h of incubation, in contrast to commercially available hen egg-white lysozyme, which showed no such effect. This activity against the Gram-negative envelope offers a plausible functional role for the C-terminal AMP-like region, presumably facilitating electrostatic interaction with lipopolysaccharide and destabilization of the outer membrane. However, it should be emphasized that the involvement of this region in the activity of 57_3Lys remains a hypothesis that requires direct experimental validation.

In summary, the subject of this doctoral dissertation was a comprehensive analysis of bacteriophages infecting uropathogenic *E. coli* strains and the identification of factors within their genomes with potential application-oriented and fundamental research significance. This research was motivated by the growing problem of antimicrobial resistance among uropathogens, associated, among other factors, with the inappropriate use of antibiotics. The scale of this phenomenon varies depending on the geographical region and the class of antibiotics used, highlighting the need to explore alternative approaches to combating infections caused by UPEC. One such approach is phage therapy, whose potential is associated, among other factors, with the ability of lytic phages to eliminate biofilms, replicate at the site of infection, and selectively target pathogenic bacteria while exerting a limited effect on commensal bacteria.

However, our systematic review of 98 anti-UPEC bacteriophages revealed significant gaps in their characterization to date. The most notable limitations concerned safety assessment, including the lack of cytotoxicity studies using eukaryotic cells, as well as incomplete EOP determinations, limited assessment of phage stability, and incomplete genome annotation and analysis. These findings highlighted the need for a more comprehensive and standardized framework for bacteriophage characterization. The results obtained in the experimental part of this dissertation directly addressed the limitations identified in the systematic review. We isolated and comprehensively

characterized novel lytic bacteriophages infecting clinical UPEC strains, leading to the proposal of five new phage species. Particular attention was given to phage vB_EcoS-149_4M, which, owing to its broad host range, high stability, and favorable safety profile, was identified as a candidate for further studies aimed at assessing its suitability for phage therapy. We further extended this approach through an in-depth characterization of bacteriophage vB_EcoS_57-3 and its endolysin, 57_3Lys. An important component of this work was the assessment of phage properties under conditions relevant to its potential application in urinary tract infections. We demonstrated its stability and lytic activity in artificial urine at pH values corresponding to those occurring in the urinary tract, while no cytotoxic effects on human cells were observed. In parallel, we analyzed 75 genomes of lytic anti-UPEC bacteriophages, focusing on the identification of putative endolysins. We identified 39 proteins belonging to five functional groups that differed in their predicted enzymatic activities. This analysis also demonstrated that genome annotation alone does not always enable reliable identification of endolysins. Detailed characterization of 57_3Lys subsequently enabled us to combine bioinformatic predictions with functional experimental data. The results indicate that the endolysin 57_3Lys identified in this study exhibits antibacterial activity and may have potential future applications in the treatment of UTIs. The enzyme was capable of significantly reducing both the turbidity and the number of viable cells of a clinical UPEC isolate in the absence of an exogenous outer-membrane permeabilizer, highlighting its activity against the Gram-negative bacterial envelope and further supporting the therapeutic potential of phage-derived lytic enzymes as a complementary strategy to phage therapy.

Importantly, the findings of this dissertation are relevant not only to the identification of novel antimicrobial agents but also have fundamental research significance for the study of bacteriophage biology and evolution. Analysis of genes influencing phylogenetic tree topology enabled us to identify sequences particularly informative for reconstructing evolutionary relationships among anti-UPEC phages. The selected genes, including those encoding tail proteins, portal proteins, lysis-associated proteins, and proteins involved in genome replication, exhibited strong phylogenetic signals. They may therefore serve as useful genomic markers for distinguishing closely related phages and improving the reliability of their classification and the reconstruction of their evolutionary relationships.

Overall, the results obtained in this dissertation demonstrate that bacteriophages are a source of diverse factors with potential application-oriented and fundamental research significance. Their comprehensive characterization enables not only the identification and assessment of candidates for incorporation into novel antimicrobial strategies but also provides valuable information for understanding the diversity, biology, and evolution of these viruses. In particular, the integration of classical biological approaches with genomic bioinformatics and phylogenetic analyses enables a more comprehensive assessment of both the potential applications of phages and their products and the underlying biology and evolutionary history of bacteriophages.