Isolation and characterization of phages for enterotoxigenic *E. coli*

Francisca Rodrigues Gonçalves

UMinho | 2023

Universidade do Minho
Escola de Engenharia

Francisca Rodrigues Gonçalves

Isolation and characterization of phages for enterotoxigenic *E. coli*

April 2023

Universidade do Minho
Escola de Engenharia

Francisca Rodrigues Gonçalves

Isolation and characterization of phages for enterotoxigenic *E. coli*

Master's Dissertation
Master's degree in Biotechnology

Work conducted under the guidance of:
Doctor Ana Cristina Afonso Oliveira
Doctor Luís Daniel Rodrigues de Melo

April 2023

Direitos de autor e condições de utilização do trabalho por terceiros

Acknowedgements

Estando prestes a terminar mais uma etapa do meu percurso académico e a cumprir mais um dos objetivos que tracei para mim, não posso deixar de agradecer a todas as pessoas que fizeram parte desta jornada e me possibilitaram chegar até aqui.

Em primeiro lugar, gostaria de agradecer aos meus orientadores, Doutora Ana Oliveira e Doutor Luís Melo. Pela oportunidade que me deram ao ingressar neste projeto, por toda disponibilidade e atenção, pelos conselhos, pela constante partilha de conhecimento e pela paciência, o meu mais sincero obrigada. Foi um prazer ter trabalhado convosco nos últimos meses e ter o vosso apoio durante esta etapa, não tenho dúvidas do quanto vocês me ajudaram a crescer a nível académico. Queria também deixar uma palavra de agradecimento a todos os meus colegas do LPhage. Obrigada pelo acolhimento, pela interajuda e pelos momentos de convívio. No entanto, queria deixar um obrigado especial aos meus colegas de mestrado e laboratório, Pedro, Valentina e Mariana, desejo-vos um futuro brilhante.

Aos meus pais, um obrigado nunca será suficiente para retribuir tudo o que fazem por mim. É graças a vocês, e por me permitirem alcançar os meus sonhos, que estou aqui hoje. Obrigada por me ensinarem a lutar pelo que quero e por acreditarem sempre em mim e naquilo que sou capaz. À minha irmã, da mesma forma que aprendo contigo todos os dias, espero ter sido um bom exemplo durante este nosso percurso. Obrigada por todo o apoio e principalmente pela paciência, fico-te a dever um cornetto! A toda a minha família, agradeço todo o carinho e preocupação!

Não posso deixar de agradecer de forma especial ao Luís. Sem dúvida que foste das pessoas mais importantes durante esta etapa da minha vida. Acima de tudo, quero agradecer por estares sempre presente, por me incentivares a não desistir e por todo o carinho e compreensão. Obrigada por fazeres parte da pessoa que me tornei, por me ensinares a crescer e a ser paciente. Serei sempre grata por te ter na minha vida e por todo o teu apoio, infinitos!

Por último, queria deixar uma palavra de agradecimento a todos os meus amigos, sem dúvida que tornaram este caminho muito mais divertido! Aos do costume, obrigada por serem família. À Sara, obrigada pelas palavras de incentivo. À Krypton, Andreia, Sérgio, Tec, Pinto e Potássio, obrigada por terem caminhado comigo e por terem tornado Braga um pouco mais bonito, não podia ter pedido melhor companhia. A toda a grupeta maravilha de Guimarães, obrigada pela amizade, apoio e momentos de descontração, que muito me fizeram falta durante este percurso!

Statement of Integrity

I, Francisca Gonçalves, hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

Resumo

A ocorrência de diarreias está relacionada com infeções causadas por *Escherichia coli*, sendo a *E. coli* enterotoxigénica (ETEC) frequentemente identificada em casos de diarreia do viajante e de diarreia pós-desmame (PWD) em suínos. A vasta utilização de antibióticos para tratar infeções de ETEC em animais alertou as autoridades de saúde pública devido ao desenvolvimento bacteriano de resistência a múltiplos medicamentos. Assim sendo, é urgente encontrar soluções eficazes para substituir os antimicrobianos convencionais. Os bacteriófagos (fagos), vírus que infetam bactérias, são uma abordagem de tratamento promissora devido ao seu potencial para tratar infeções bacterianas. No entanto, algumas bactérias resistem à infeção fágica através de sistemas de defesa anti-fagos (APDS). O presente estudo tem como objetivo isolar e identificar fagos que possam contornar as principais APDS da ETEC, alargando o seu alcance lítico e possibilitar a sua utilização na terapia. Foi possível isolar três fagos, EcoSus34, EcoSus42 e EcoSus65. Atribuídos à classe *Caudoviricetes*, os fagos apresentam duas morfologias distintas: EcoSus34 e o EcoSus65 são myovirus e o EcoSus42, podovirus. Através de um processo de caraterização, todos os fagos revelaram um baixo espetro de hospedeiros e, num único caso, uma baixa eficiência de infeção. A análise genómica dos fagos indicou que todos exibem características típicas de ciclos de vida exclusivamente líticos e não possuem quaisquer proteínas associadas à virulência. No entanto, o EcoSus65 destacou-se como sendo o único fago que codificou uma possível contra-defesa para os APDS, nomeadamente a molécula Dmd, que interfere com o rácio toxina-antitoxina (TA) da bactéria. A ETEC H10407 (do GenBank) e a EC43 foram submetidas a anotação funcional e tinham um total de nove e quatro proteínas relacionadas com a APDS, respetivamente. Entre as proteínas identificadas, a incidência de sistemas TA foi maior, mas as bactérias também codificaram sirtuins associados à defesa (DSR), sistemas de exclusão de superinfeção (Sie) e sistemas de restrição-modificação (RM). Além disso, a presença de um mecanismo capaz de causar a morte celular após a adsorção fágica foi verificada *in vitro*, utilizando as estirpes EC40 e EcoSus42 da ETEC. A um nível geral, o trabalho desenvolvido confirmou a importância da ocorrência de APDS em estirpes bacterianas e a urgência de encontrar fagos que consigam ultrapassar os mecanismos de defesa, como o EcoSus65, sendo o mesmo selecionado como o candidato mais adequado para potenciais aplicações terapêuticas.

Palavras-chave: Bacteriófago, *E. coli* enterotoxigénica, terapia fágica, sistemas de defesa anti-fagos

Abstract

Frequent occurrences of diarrheal illnesses are often related to infections caused by various pathotypes of *Escherichia coli*, with enterotoxigenic E. coli (ETEC) being the most frequently identified pathogen in cases of travelers' diarrhea and post-weaning diarrhea (PWD) in pigs. The massive use of humans' last resort antibiotics to treat ETEC infections in animals caught the attention of public health authorities due to bacteria's ability to develop multidrug resistance. Therefore, it became urgent to find efficient solutions to replace conventional antimicrobials. Bacteriophages (phages), viruses that specifically infect bacteria, represent a promising treatment approach because of the enormous potential to overcome bacterial infections. Nevertheless, some bacteria can resist to phage infection using a wide range of anti-phage defense systems (APDS). The present study aims to isolate and identify phages that can effectively address the primary APDS of ETEC, widening their range of lytic activity for possible use in therapy. Firstly, the isolation of three ETEC-infecting phages EcoSus34, EcoSus42 and EcoSus65 was successfully accomplished. As part of *Caudoviricetes* class, the phages represented two distinct morphotypes: EcoSus34 and EcoSus65 are myovirus and EcoSus42, podovirus. Through a characterization process, among a selection of 95 ETEC strains, all phages revealed a low host range and, in a single case, a low efficiency of infection. The phages' genomic analysis indicated that all exhibit typical features of exclusively lytic life cycles and did not have any virulence-associated proteins. However, EcoSus65 stood out as the single phage that encoded a possible counter-defense for bacterial APDS, this being a Dmd molecule that interferes with toxin-antitoxin (TA) bacterial ratio. ETEC H10407 (from GenBank) and EC43 underwent functional annotation and had total of nine and four proteins possibly related to APDS, respectively. Among these identified proteins, the incidence of TA systems was higher, but the bacteria also encoded defense-associated sirtuins (DSR), superinfection exclusion (Sie) systems and restriction-modification (RM) systems. Additionally, the presence of a mechanism capable of causing cell death upon phage adsorption was supported *in vitro*, using ETEC strain EC40 and EcoSus42. Overall, the work developed confirmed the importance of APDS occurrence in bacterial strains and the urgence of finding phages that can overcome defense mechanisms, such as EcoSus65, being selected as most suitable candidate for potential therapeutic applications.

Keywords: Bacteriophage, enterotoxigenic *E. coli*, phage therapy, anti-phage defense systems

Table of Contents

List of Abbreviations and Acronyms

ETEC	Enterotoxigenic *Escherichia coli*
DNA	Deoxyribonucleic Acid
PWD	Post-weaning Diarrhea
EPEC	Enteropathogenic *Escherichia coli*
EHEC	Enterohaemorrhagic E*scherichia* coli
EAEC	Enteroaggregative *Escherichia coli*
EIEC	Enteroinvasive *Escherichia coli*
DAEC	Diffusely Adherent *Escherichia coli*
VTEC	Vero-toxin Producing *Escherichia coli*
STEC	Shiga-toxin Producing *Escherichia coli*
CFs	Colonization Factors
LT	Heat-labile Enterotoxin
ST	Heat-stable Enterotoxin
GSα	Guanine Nucleotide Protein
AC	Adenylate Cyclase
cAMP	Cyclic Adenosine Monophosphate
PKA	Protein Kinase A
CFTR	Cystic Fibrosis Transmembrane Conductance Regulator
GC-C	Guanylate Cyclase C
cGMP	Cyclic Guanosine Monophosphate
RS	Rehydration Salts
Cu	Copper
ZnO	Zinc Oxide
EU	European Union
DFM	Direct Fed Microbials
TMP	Trimethoprim
SMZ	Sulfamethoxazole
PBPs	Penicillin-binding Proteins
CS	Colistin Sulfate
MDR	Multidrug Resistance
WHO	World Health Organization
HGT	Horizontal Gene Transfer
ARG	Antibiotic-resistant Gene
RNA	Ribonucleic Acid
RBP	Receptor Binding Proteins
LPS	Lipopolysaccharide
mRNA	Messenger Ribonucleic Acid
APDS	Anti-phage Defense Systems
EPS	Exopolysaccharide
Sie	Superinfection Exclusion
RM	Restriction-modification

CRISPR - Cas	Clustered Regularly Interspaced Short Palindromic Repeat – Associated
AdoMet	S-adenosylmethionine
GTP	Guanosine triphosphate
NAD/NAD$^+$	Nicotinamide Adenine Dinucleotide
Abi	Abortive infection
TA	Toxin-antitoxin
CBASS	Cyclic Oligonucleotide-based Anti-phage Signaling System
Pycsar	Pyrimidine Cyclase System for Anti-phage Resistance
Agos	Short prokaryotic Argonautes
DSR	Defense-associated Sirtuins
LB	Lysogeny Broth
LB 2x	Double Concentrated LB
rpm	Rotation *per* Minute
OD	Optical Density
OD_{600}	Optical Density at 600 nm
PFU	Plaque Forming Units
EOP	Efficiency of Plating
LFW	Lysis from Without
MOI	Multiplicity of Infection
NC	Negative Control
epp	Eppendorf
K_a	Adsorption Constant
CFU	Colony Forming Units
TEM	Transmission Electron Microscopy
ORF	Open Reading Frames
tRNA	Transfer Ribonucleic Acid
bp	Base Pairs
CDS	Coding Sequences
VICTOR	Virus Classification and Tree Building Online Resource
GBDP	Genome Blast Distance Phylogeny
ICTV	International Committee on Taxonomy of Viruses

List of Figures

List of Tables

1. Introduction

1.1. Enterotoxigenic *Escherichia coli*

Having an outer membrane composed of lipopolysaccharides (LPS), *Escherichia coli* is a Gram-negative bacterium usually with a rod-shaped morphology (measuring 1 - 3 μm *per* 0.4 - 0.7 μm). Its motility is attributed to a peritrichous flagellar arrangement, and optimal growth is observed at a temperature of 37°C. It is part of the natural intestinal flora, primarily as a major facultative anaerobe residing in the large intestines of both humans and warm-blooded animals [1,2].

In the absence of genetic elements that carry virulence factors, these bacteria remain harmless to organisms, contributing positively to the human microbiome by aiding in digestion and defending against opportunistic pathogens. However, when containing harmful genes, they become virulent and have the potential to induce disease, such inflammatory dysentery [3].

Frequent occurrences of diarrheal illnesses in both farm animals and humans are often related to infections caused by various pathotypes of *E. coli*. These strains have developed specific characteristics through horizontal gene transfer, which have effectively endured within the host. Among these pathotypes, enterotoxigenic *E. coli* (ETEC) stands out as a significant contributor to this type of diseases, particularly affecting young animals like weaned pigs [1,4].

In humans, ETEC is the most frequently identified pathogen in developing regions of the world and in cases of travelers' diarrhea [5,6]. This pathogen is also responsible for causing moderate-to-severe diarrhea in children under the age of 5, which can be linked to a higher likelihood of stunted growth, increasing the risk of mortality from other infectious diseases. In 2015, more than 40 000 deaths were reported due to cases of diarrhea caused by ETEC infection [7].

Regarding farm animals, studies show that ETEC is the most widespread pathotype responsible for post-weaning diarrhea (PWD) in pigs. This condition commonly arises within two weeks post-weaning, marked by severe diarrhea, dehydration, notable mortality, and weight loss in surviving pigs. This development stage is crucial for pig health as the immature intestinal immune system, combined with the cessation of sow milk consumption, heightens susceptibility to microbial infections. Being the most prevalent illnesses in swine industry, PWD infected pigs can reach a mortality rate of 20% to 30% over the course of one to two months [8–10].

Pathogenesis, virulence factors, and mechanisms of ETEC

The genome sizes of *E. coli* can vary by a million base pairs between symbiont and pathogenic strains, with this additional genetic material potentially containing virulence and fitness-related genes. Comparative genomics has revealed that *E. coli* genomes can be divided into two main categories: a common, conserved set of genes referred to as the core genome and a dynamic gene pool [11]. Despite the fact that this chromosomal flexibility from the dynamic gene pool expedites *E. coli*'s adaptation to diverse environments, it also enables several concurrent and precise evolutionary routes through gene gain and loss, ultimately leading to similar phenotypes [12].

A total of 11 distinct pathotypes of *E. coli* have been classified into two main categories, namely intestinal and extraintestinal pathogenic *E. coli* [13]. Among the intestinal pathogenic *E. coli*, which can also be characterized as diarrheagenic *E. coli* due to specific combinations established based on the group of acquired virulence factors, the pathotypes exhibit variations in terms of their favored host colonization sites, mechanisms of virulence, and the resulting clinical symptoms and outcomes [1,14].

Escherichia coli main intestinal pathotypes, aside from ETEC, are enteropathogenic *E. coli* (EPEC), enterohaemorrhagic *E. coli* (EHEC), enteroaggregative *E. coli* (EAEC), enteroinvasive *E. coli* (EIEC), diffusely adherent *E. coli* (DAEC) and Vero- or Shiga- like toxin-producing *E. coli* (VTEC or STEC) [8].

ETEC enters the human or animal body through ingestion when exposed to contaminated food and drinking water. In livestock, it can also occur from asymptomatic carrier piglets or sows to uninfected piglets. When bacteria are present in significant quantities, it establishes colonies in the small intestine. This colonization starts with the attachment of bacteria to the specific receptors expressed on the small intestinal epithelium or within the mucus layer covering the epithelium, facilitated by specific fimbrial adhesins that are surface proteins called colonization factors (CFs). As soon as the colonization is achieved, the ETEC spreads quickly and produces one or more different forms of enterotoxins, leading to intestinal fluid secretion and common clinical symptoms associated with ETEC-triggered diarrhea [8,15,16].

The excreted enterotoxins are categorized into two groups: the heat-labile enterotoxins (LTs) and the heat-stable enterotoxins (STs). ETEC strains can express either LT, ST, or both [17]. STs are relatively small enterotoxins and are present in approximately 75 to 80% of ETEC isolates, with around 45% consisting of ST alone. It is worth noting that STs are more commonly associated with

severe human diseases compared to strains with LTs only. Regarding LTs, they are bigger toxins and share approximately 80% similarity with the cholera toxin (expressed by *Vibrio cholerae*) [18].

The mentioned excreted enterotoxins promote the secretion of water and electrolytes (Na^+ and Cl^-) in the small intestine, thereby decreasing fluid absorption, resulting in dehydration and acidosis. This secretion is possible because LT, after being transferred to the endoplasmic reticulum and cytoplasm of the host cells, triggers the activation of guanine nucleotide protein (GSα) through ADP-ribosylation. This activation initiates the activity of adenylate cyclase (AC), leading to an increase in cyclic adenosine monophosphate (cAMP) levels. Following this, cAMP-dependent protein kinase A (PKA) becomes active that engages the opening of the cystic fibrosis transmembrane conductance regulator (CFTR). Consequently, electrolytes and water are secreted into the intestinal lumen.

As for the ST enterotoxin, they are divided into two categories according to their structure and function, referred to as STa and STb. The functional effect of STa is achieved by stimulating the guanylate cyclase C (GC-C) signal transduction pathway, resulting in the intracellular buildup of cyclic guanosine monophosphate (cGMP). The accumulation of cGMP contributes to diarrhea in two ways, one involves the opening of the CFTR channel, resulting in the substantial release of chloride and bicarbonate into the intestinal lumen and the second mechanism involves the indirect inhibition of the sodium-hydrogen antiporter, which diminishes sodium reabsorption. Regarding STb, its uptake results in extracellular calcium ions entering the cell through a calcium channel that is activated by receptor-dependent ligands. The elevation in intracellular calcium levels results in the activation of a calcium-dependent chloride channel, permitting the release of chloride ions from the cell into the lumen. This increased in calcium levels can additionally trigger the production of prostaglandins, which can have a role in the regulation on the movement of water and electrolytes. All this infection process is represented in **Figure 1** [19–22].

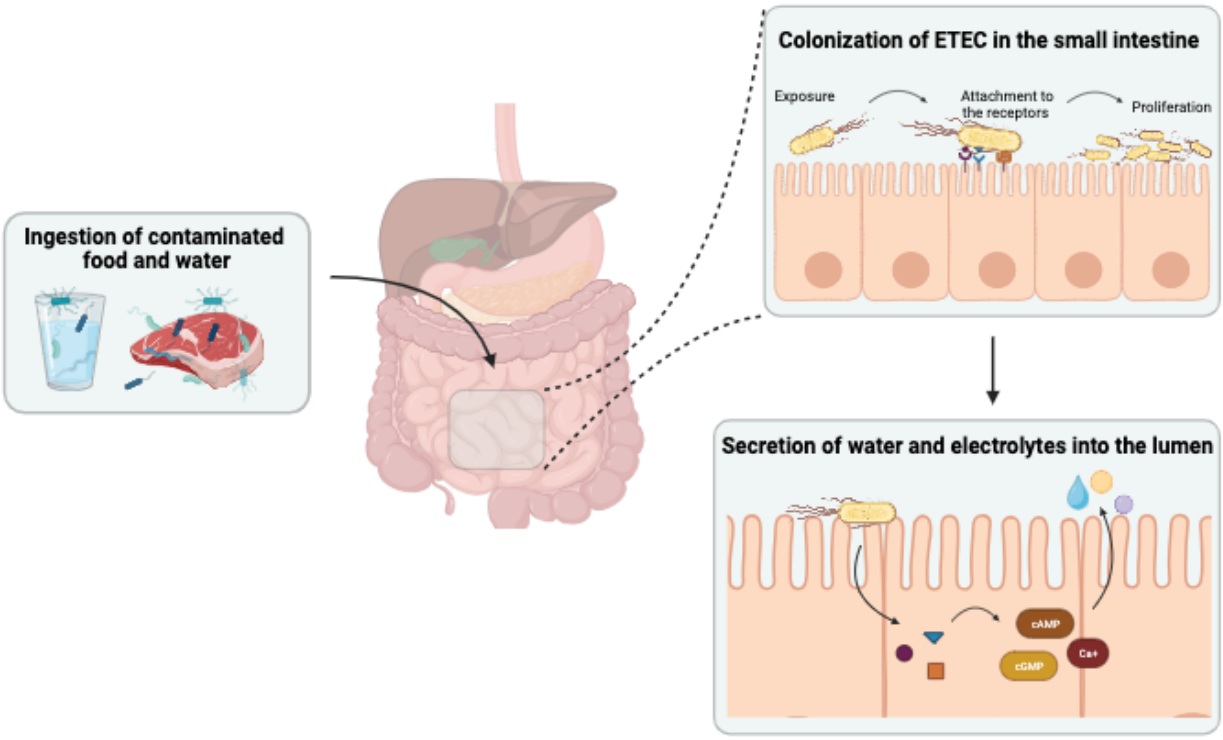

Figure 1 - Pathogenesis of ETEC. ETEC infections result from the consumption of contaminated food and water. Then it enters the gastrointestinal tract, eventually establishing colonization in the small intestine. Once within the small intestine, adheres to intestinal epithelial cells through CFs, leading to its proliferation on the intestinal surface. ETEC produces and releases LT and ST enterotoxins to induce their toxic effects, releasing water and electrolytes into the lumen to further develop diarrheal illnesses. Image created with BioRender.

Currently used measures against ETEC infection

The approach to treating diarrheal disease caused by ETEC is identical to that used for cholera or any other form of acute secretory diarrhea. The primary focus in all cases remains on a nutritional management, which may involve dietary adjustments, complemented with rehydration therapy [23]. The choice between oral or intravenous rehydration depends on the extent of dehydration, and it is advisable to use rehydration salts (RS) solution, to compensate the water and electrolyte loss. In cases of severe dehydration caused by diarrhea, initial management may require intravenous fluids, succeeded by a transition to oral RS solution for the ongoing correction of fluid losses [24].

Aside from humans, rehydrating pigs presents a unique challenge due to the impracticality of the intravenous route and subcutaneous administration. Intraperitoneal injection is an option but has limitations, including a restricted infusion volume and unpredictable uptake, so it is replaced by oral administration of electrolyte solutions with glucose to treat every type of dehydration and metabolic acidosis [25].

Among food-producing animals, it is, thus, crucial to treat the infection caused by the bacteria so that the outbreaks can be minimized and the disease controlled. This can be achieved by treatments such as vaccination, antibiotic, specific antibodies or the usage of feed supplements. There are also other alternative therapies for PWD, that include breeding of resistant pigs, the use of bacterial probiotics or proteolytic treatment [15].These alternatives therapies will not be discussed in this document.

Vaccination

Rapid implementation of sanitation systems and the provision of safe drinking water, which could prevent ETEC and other enteric pathogen-related diarrhea infections, are often challenging to achieve in many countries with limited resources. For that reason, vaccination is regarded as the most feasible and efficient approach for preventing ETEC diarrhea [26].

Nonetheless, a significant obstacle in the development of successful ETEC vaccines lies in the diversity found within ETEC strains, making it difficult to establish durable mucosal immunity within the host against strains expressing more than 25 distinct CFs antigens and two enterotoxins. Additionally, there is a deficiency of an ideal challenge model to conclusively assess the protective effectiveness of vaccine candidates and production of vaccines that are both efficient and affordable. For these reasons, up until 2020, no vaccines for ETEC had received official licensing [27].

The current frontrunner, ETVAX, is an oral inactivated whole-cell vaccine that merges four inactivated genetically modified *E. coli* strains, which overexpresses common CFs along with a recombinant subunit protein (incorporating the binding subunits of LT and cholera toxin) and it is currently undergoing an expanded field trial among adult Finnish travelers (aged 18-64) [28,29]. Another advanced vaccination candidate is ACE527, composed of three live attenuated strains of ETEC with deleted enterotoxin genes and antibiotic resistance determinants. The vaccine induced immune responses against CFs expressed on each of the three strains in most subjects during a phase one trial with healthy adult volunteers, where ACE527 was well-tolerated at high dosage [30].

Despite not as progressed, other vaccines addressing ETEC treatment are under development, being the case of MecVax. This polyvalent candidate triggers the production of antibodies capable of neutralizing the enterotoxic effects of STa and LT and prevents the binding of seven adhesins, additionally proved effective in shielding rabbits from ETEC colonization and preventing ETEC-induced diarrhea in pigs [31].

Concerning swine-related treatment against ETEC, three categories of vaccines have been tested. The first type involves intramuscular injectable vaccines, that trigger systemic immunity and elevate circulating antibodies, maintaining low levels of intestinal bacteria to prevent pathogenicity. The second approach involves orally administering non-enterotoxigenic *E. coli* strains with pigs' fimbrial adhesins, that will stimulate their intestinal colonization, leading to the secretion of intestinal antibodies and ultimately hindering ETEC adherence. The third method involves orally administering purified fimbriae, rather than the entire bacteria. This type of vaccine triggers a mucosal immune response and leads to substantial reduction of pathogenic *E. coli* in feces [32].

Although these candidates are a promising start to find an optimal ETEC vaccine, they wouldn't effectively block the colonization of the host's small intestines by all virulent pathotypes and neutralize the toxic effects of both ETEC toxins, but creating such a vaccine for ETEC appears unreachable [33,34].

Specific antibodies

Considerable research has explored the use of specific antibodies to prevent diarrhea triggered by similar CFs antigens. People who were orally administered hyperimmune bovine serum immunoglobulins targeting an entire ETEC strain demonstrated protection against both moderate and severe diarrhea [35,36]. This preventive measure is derived from the safeguarding effect offered by colostrum (first milk produced by mammals after birth) and breast milk to newborns across diverse animal species. This substance exhibits a potential antimicrobial action to neutralize endotoxins within the digestive tract due to the presence of significant amounts of immunoglobulins to confer passive immunity following birth so, it is thought to play a substantial role in alleviating gut inflammation, enhance mucosal health and aid tissue repair [37,38].

Some approaches of this type of treatment are based on targeting CFA/I, that are the most frequently occurring CFs expressed by ETEC strains, and its minor adhesin subunit (CfaE) located at the tip of the fimbria. Since this subunit can trigger anti-adhesive immunity against ETEC infection, in human trials, the oral administration of bovine IgG (primary class of immunoglobulin found in colostrum) antibodies targeting CfaE provided protection for over 60% of the participants [39]. Given this, the success of passive immunization using bovine antibodies necessitates the consumption of substantial quantities or simultaneous use of buffering agents to overcome the acidic environment of the stomach, before reaching the small intestine [40].

In farm animals it is used egg yolk antibodies, derived from immunized laying hens, that target specific bacterial fimbrial antigens, consequently decreasing the attachment of ETEC to the mucosal epithelium of the small intestine providing a cost-effective measure [41]. The usage of blood plasma demonstrated to enhance weight gain and decrease the incidence of ETEC-associated PWD due to the presence of specific anti-ETEC antibodies found in the blood plasma [42]. However, despite the promising results, the researchers conclude that the use of specific antibodies is a temporary risk reduction measure instead of a treatment one [23]. Also, it is important to consider the future challenges related to specific antibodies, which encompass aspects such as stability, affordability, and accessibility.

Feed supplements

In addition to all types of treatments, there are several non-antibiotic feed supplements and supportive measures that can be used in the treatment and management of ETEC infections in livestock, particularly in animals such as pigs, by enhancing the animal's capacity to resist the colonization of pathogenic bacteria in their intestinal system, often achieved through an enhanced immune response to pathogens [43]. Among the supplements used, the ones that stand out in the treatment of PWD caused by ETEC are:

- Minerals, such as copper (Cu), that have beneficial impact on pigs' diet due to its bacteriostatic and bactericidal properties, attributed the ability to decrease bacterial populations in the intestine, potentially influencing the growth and structure of microorganisms in the cecum and colon [44]. As for zinc oxide (ZnO), it was highly used as mineral supplement in swine industry, however, in 2022, the European Union (EU) banned the inclusion of pharmacological levels of ZnO due to the accumulation of this microelement in the environment [45];
- Acidifiers, which regulate the pH levels in the gastrointestinal tract and manage bacterial growth in both the stomach and intestine, also a study showed that adding to the diet lactic acid or citric acid proved to be effective in preventing PWD [46];
- Direct fed microbials (DFM), which are products containing live microorganisms that support gut microbiota, aid digestion, and boost the host's immune system. Some bacteria strains revealed efficiency in reducing both pathogenic ETEC intestinal infections and associated intestinal inflammatory responses [47].

It is important to mention another form of feed supplements called phytogenic which are natural compounds sourced from plants and added to animal diets to enhance livestock productivity. These types of additives have antioxidative, antimicrobial, growth-enhancing properties and have the potential to ultimately lower the presence of intestinal pathogens, due to the production of intestinal mucus inhibiting their adhesion to the mucosa [48].

Antibiotics

The use of antibiotics has historically been effective in the prevention and treatment of bacterial infections, so the initial approach to treating ETEC infections involved the use of antimicrobials. Since then, research in this field has evolved to identify the most effective antibiotic solutions for combating this pathotype.

Research had demonstrated that antibiotics like trimethoprim (TMP) alone or used in combination with a sulfonamide family antibiotic called sulfamethoxazole (SMZ), decreased duration of abdominal cramps and loss of appetite, along with the reduction in both the duration and quantity of diarrhea in patients undergoing treatment [49,50]. Previously, another study showed that the use of doxycycline, an antibiotic of the tetracyclines family, prevented gastrointestinal symptoms and, consequently, the lack of detectable ETEC in stool samples from individuals using the medication [51,52]. Then, a few years later, it was proved that quinolone antibiotics, like ciprofloxacin, provided an 84% protection rate, surpassing the 51% protection rate achieved with TMP/SMZ, serving as a highly efficient and secure antimicrobial when utilized as a preventative measure against travelers' diarrhea [53,54]. The β-lactam antibiotics were another treatment proposal undergoing analysis because of its capability of causing *E. coli* death through the inactivation of enzymes known as penicillin-binding proteins (PBPs) [55,56]. However, β-lactam antibiotics are generally not the first-line choice for treating ETEC infections because these bacteria are typically more susceptible to other classes of antibiotics.

One of the most used last-resort antibiotics among food-producing animal, that hasn't been discussed yet, is colistin, a member of the polymyxin family. In addition to its use in treating Gram-negative bacterial infections in humans, colistin sulfate (CS) was the sole approved product for managing intestinal infections in pigs caused by *Enterobacteriaceae* due to the decrease in ETEC discharging and lower diarrhea ratings during the treatment period [57,58]. Despite being banned in many countries, including EU, since around 2017, it is still used nowadays for treatment by low- and middle-income countries [59–63].

1.2. Concerns in ETEC treatment

Even though that are some possibilities of treatment that aim to reduce the incidence of ETEC, there is no approach that efficiently controls the disease. This challenge is due to many factors, including durability, cost-effectiveness, and availability of therapy methods or, because of the broad spectrum of CFs found in pathotypes. However, the primary concern lies in the bacteria's ability to develop resistance to antibiotic treatments, the most effective and manageable treatments used, especially for PWD. Between 2001 and 2004, antibiotic resistance in ETEC showed significant raising trends, with nearly 70%, 60% and 50% of human patient isolates resistant to TMP/SMZ, doxycycline and ampicillin, respectively [64].

The World Health Organization (WHO) has reassigned colistin, ciprofloxacin and ampicillin as critically vital for human medicine [65]. For that reason, the escalated utilization of this types of antibiotics in livestock production, has captured the focus of health authorities about its effectiveness in humans, as it is the leading factor in the proliferation and transmission of colistin resistance [59].

Antibiotic resistance: Development and spreading

In modern times, we can identify a growing number of organisms in both healthcare facilities and the general community that pose an overwhelming challenge for treatment due to multidrug resistance (MDR). This term includes the phenomenon of resistance to more than one antibiotic in any microorganism [66].

The excessive utilization and improper administration of antibiotics have given rise to MDR in numerous human pathogens, lowering the effectiveness of the intended treatment. In 2004, it was estimated that over 70% of pathogenic bacteria had developed resistance to at least one of the antibiotics that were in use. The development of antibiotic resistance primarily centers on two key mechanisms: genetic mutations and the resistance genes acquisition through horizontal gene transfer (HGT) [67,68].

Under a specific antibiotic concentration in a microbial system, is susceptible the appearance of mutant strains within a bacterial population due to survival adaptation. This phenomenon enables these bacteria to not only endure but also propagate as antibiotic-resistant bacteria containing antibiotic-resistant genes (ARGs) [69].

The involvement of mobile genetic elements, such as plasmids, transposons and prophages (bacteriophage genome that is integrated into the bacterial chromosome), further accelerate the spread and facilitation of genetic recombination of ARGs into non-resistant bacteria through HGT common methods of genetic interchange [70]. These methods of horizontal gene transfer include transformation, which occurs when a donor cell is damaged and releases its plasmid or chromosomal DNA into the environment for a competent recipient cell to absorb and, possibly, exhibit the characteristics imparted by the donor DNA (**Figure 2A**). Another methos is transduction, that employs a prophage as an intermediary and can mistakenly encapsulate a segment of donor cell DNA instead of the intended viral genome. When the donor cell ruptures, the bacteriophage containing the bacterial genes can infect and integrate the DNA of new recipient cells (**Figure 2B**). The third method is conjugation, that involves a direct exchange of genetic material between two bacterial cells. The donor cells must establish stable physical contact with a recipient cell to facilitate the transfer or swapping of genetic components as plasmids or transposons (**Figure 2C**) [71,72].

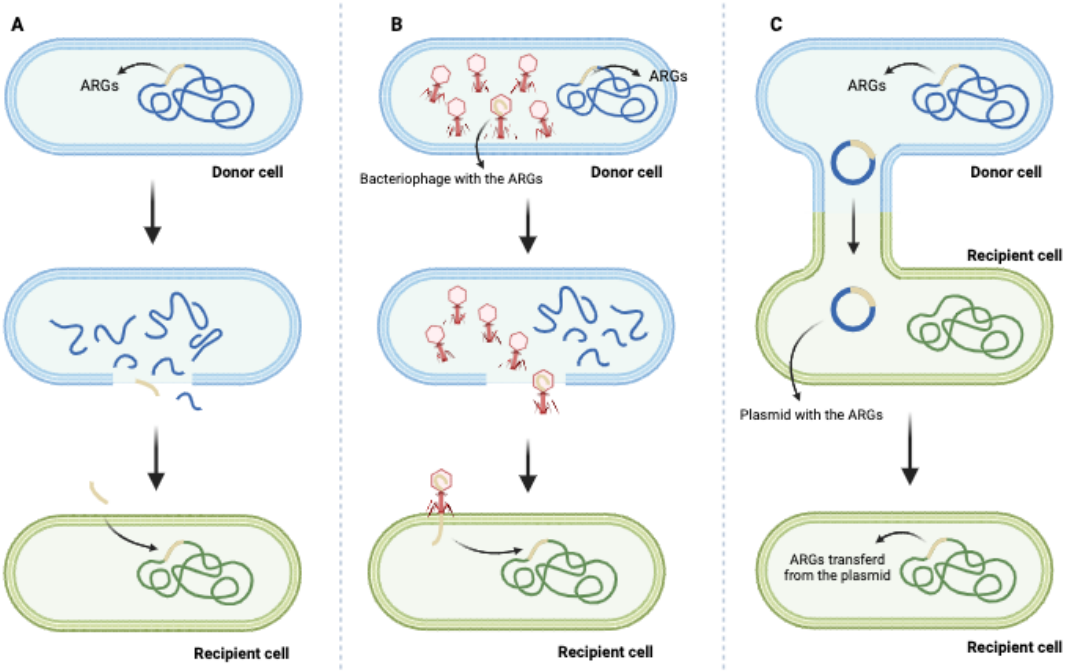

Figure 2 - Common methods of HGT regarding the exchange of ARGs between bacteria. In (**A**) is represented transformation, the absorption of exposed environmental DNA; In (**B**) is represented transduction, a transfer facilitated by bacteriophages; In (**C**) is represented conjugation, which is bacterial conjugation and exchange of plasmids. Image created with BioRender.

Besides development and transfer of ARGs, the mechanisms leading to resistance alter depending on the class of antibiotic and its genetic composition. These mechanisms can act by reducing the antibiotic uptake or enhancing antibiotic export, inactivating the antibiotic target or introducing new targets that are antibiotic-resistant and hydrolyzing or modifying the antibiotic [73].

It is equally important to take account the spreading of MDR bacteria though the environment, due to its favorable impact on mutations and the occurrence of HGT, which in turn leads to the proliferation of antibiotic-resistant bacteria. The humans high demand for vital antimicrobials has encouraged unprescribed and unconventional use of antibiotics and, consequently, our planet is saturated with these harmful agents [74]. A diagram illustrating the spreading of MDR bacteria and their exposure to humans is represented in **Figure 3**.

Antibiotic-resistant bacteria are introduced into water and land environments through sewage discharge, industrial production, animal farming, and landfill effluents. Contamination sources also include runoff from farm fields with livestock manure, irrigation with treated wastewater in agroecosystems, and the use of animal manure as fertilizer [75]. So, it is possible to affirm that human activities play a central role in the spread of ARGs, with livestock farming emerging as a primary contributor [76] (**Figure 3**). Moreover, the challenges in reducing antibiotic concentrations and ARGs in manure through composting, along with the lack of global regulations on the release of antibiotic-contaminated livestock wastewater, delay the efforts to prevent the transmission of antibiotic resistance from the environment to humans [59]. Research has confirmed that animal wastewater can potentially lead to soil contamination and affect nearby water systems with ARGs, thereby manifesting a risk to both human and animals due to the capability of entering the food chain [77]. Also, exposure to antibiotic contamination is closely linked to alterations in the composition of the intestinal microbiome that could lead to various pathologies. This is primarily attributed to antibiotics' wide-ranging impact on the entire microbial colony [78,79].

The massive misuse of last resort antibiotics in food-producing animals that are intended for humans and were considered critically vital by the WHO, such as colistin, is causing a severe contamination of antibiotic residues on the environment. This contamination leads to the propagation and spreading of ARGs in bacteria which can affect human health by influencing the effectiveness of those antibiotic treatments in ETEC infections. Besides that, forecasts for the future are concerning, projecting that in 2050, approximately 10 million lives may be lost to resistant microorganisms if no interventions occur [80]. So, as a result, is being promoted the creation of viable

and environmentally friendly alternatives for addressing these pathogens, such as bacteriophage therapy.

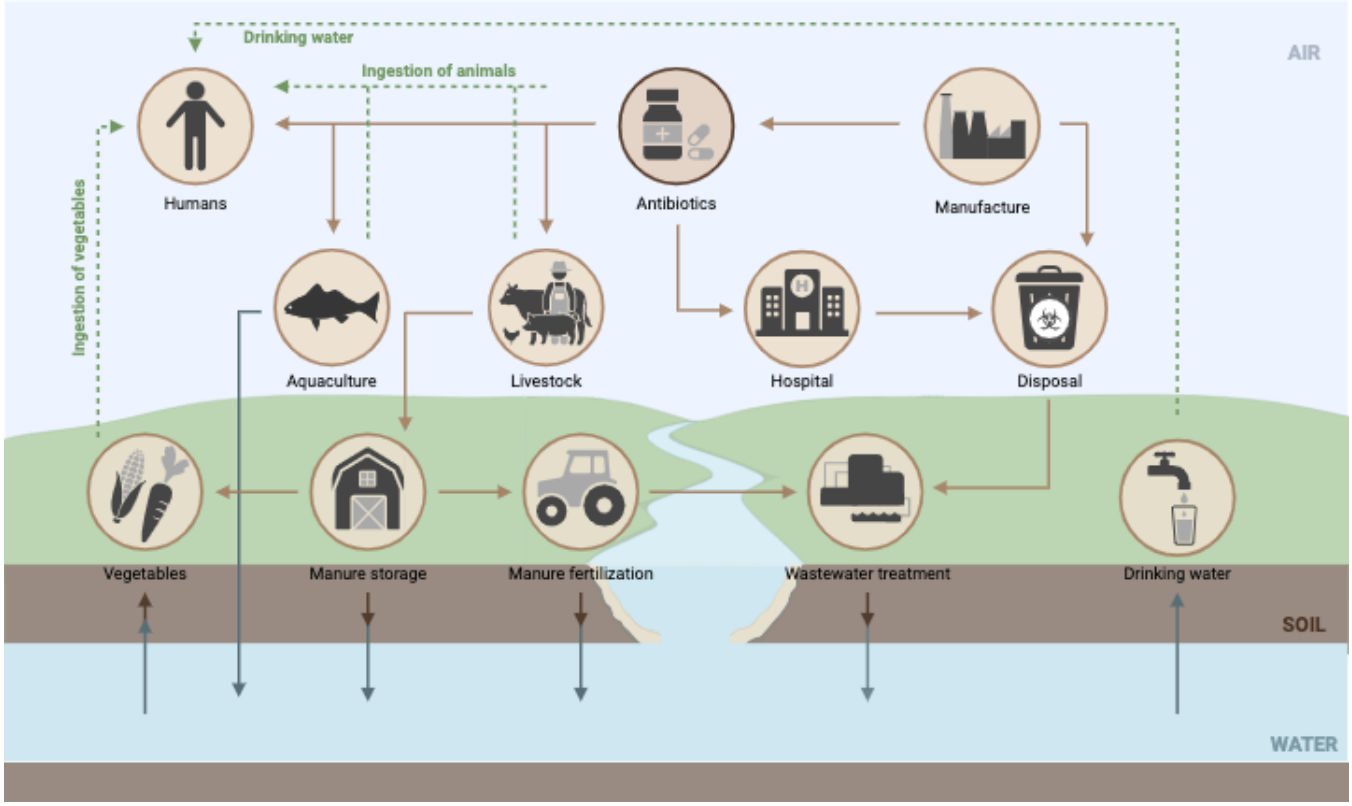

Figure 3 - Diagram representing the spreading of MDR bacteria and their exposure to human which is associated with antibiotic residues in the environment. Image created with BioRender.

1.3. Bacteriophage therapy

As strict rules and regulations are being imposed for the use of conventional antimicrobials in animal production (Regulation (EU) n° 2019/6) due to the rise of bacterial resistance events both in animals and humans, the search for alternatives methods with antimicrobial properties has gained significant popularity [81,82].

The growing curiosity for bacteriophage therapy is evident through the rising amount of research and review articles published over the years, which some have studied into specific areas, including the use of bacteriophages in agricultural production, as well as in animal farming and in human health [83]. Given their inherent characteristics, bacteriophages appear to be promising options for antibacterial treatment. This is due to their high degree of specificity towards bacterial species,

their non-threatening properties to animals and plants, and their widespread dispersion and autonomous reproduction as they infect their targeted microorganisms [84,85].

Bacteriophages (phages) have been recognized as potential antibacterial agents for more than a century, ever since their initial discovery in the 1920s and use though the 1940s to treat bacterial infections in humans. Although the subsequent success of antibiotics diminished research on phages as prospective antimicrobial agents, the growing problem of antibiotic resistance has sparked a renewed interest in phage therapy [86,87]. Throughout the previous century, countless patients, possibly numbering in the millions, have received treatment involving bacteriophages [80].

Recently, the IDSA (Infectious Disease Society of America) released an article affirming that, in cases where bacteriophage therapy was applied, the positive safety record and successful results provide strong encouragement for expanding the use of this type of treatment in future clinical trials[88].

Bacteriophage characterization

In the past 25 years, it has become evident that phages are the Earth's most abundant organisms, with approximately 10^{31} phage particles existing globally. So, these microorganisms can be found in every bacterial habitat and, it is presumed that each bacterial strain is infected by at least one type of phage, and likely, multiple types [89].

Phages are viruses that exclusively infect bacteria, and, if virulent, can always induce complete lysis and death of susceptible bacterial strains. They contain genetic information exclusively for replicating their nucleic acid and producing the proteins required for production of new phages. As obligatory parasites, these viruses can't multiply without living host cells [90].

Phages are differentiated based on the type of nucleic acid present in the genome, structure of the capsid (also called head), their life cycle and bacterial target, although, nowadays, phages' classification also depends on genomic analysis. Phage genomes consist of either DNA or RNA (ribonucleic acid) and it can be double-stranded or single-stranded. The respective genetic material is enclosed within a capsid which can take various forms, including polyhedral, filamentous, or pleomorphic, and it can be attached to a tail (*Caudoviricetes*). Nearly all isolated phages have tailed structures and double-stranded DNA genomes and can be subdivided into three morphotype groups, namely, podovirus, siphovirus and myovirus, depending mostly on tail structure (**Figure 4**) [91]. The connection between the capsid and its tail, via connector mechanism, typically consists of a portal

protein linked to the head or connector proteins. The portal protein creates a channel through which viral DNA can pass in two opposing directions, allowing it to enter and exit the viral capsid [92,93].

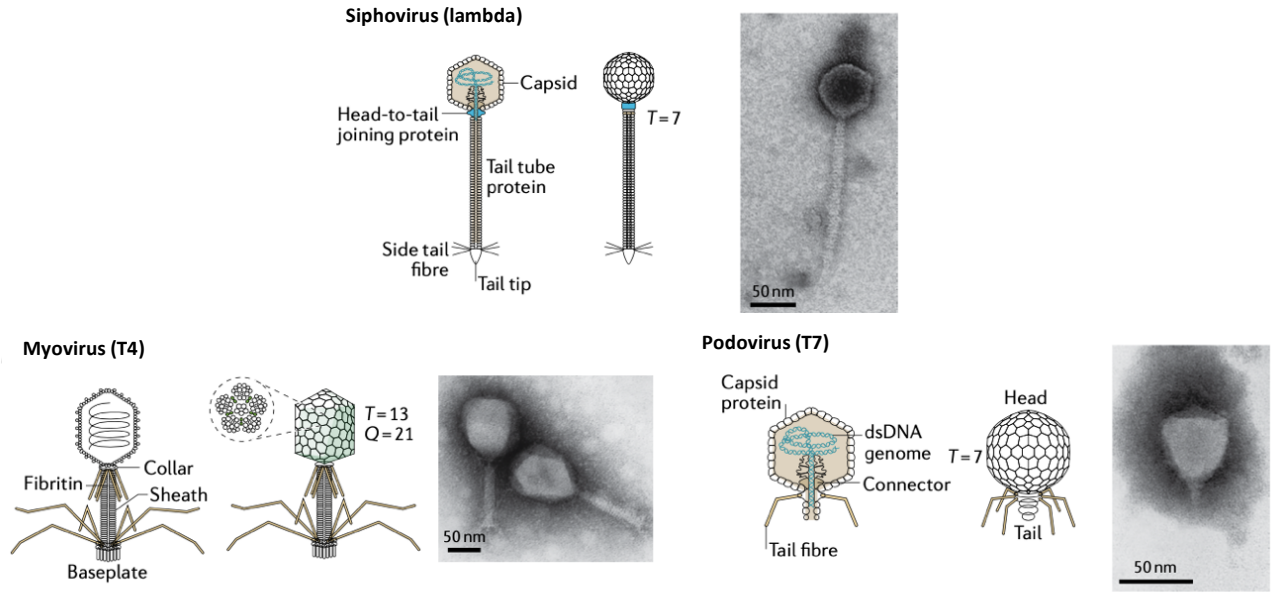

Figure 4 – Morphology groups of dsDNA tailed phages according to their structural characteristics and a transmission electron micrograph (TEM) of a phage example is exhibited in each group. Siphovirus-like phages have long non-contractile tail (TEM of phage lambda is shown), Myovirus-like phages have long contractile tail (TEM of T4 is shown), and Podovirus-like phages have short non-contractile (TEM of T7 is shown). Image adapted from Dion *et al.* [91].

Bacteriophage life cycles

Phages can exhibit two different life cycles that determine their impact on bacterial biology and classify them according to their virulence. The life cycles are called lytic cycle, characterized by virulent or productive phages, and lysogenic cycle, characterized by temperate or dormant phages (**Figure 5**) [94,95].

The first stage of both phages' life cycles is the adsorption of the phage into their host. This stage represents a critical step in virus recognition of a susceptible host cell because phages must successfully entry the bacteria to continue their life cycle [96]. The phage attaches itself to the bacterial cell, through receptor binding proteins (RBP), connecting to specific receptors situated on the bacterial cell (e.g. lipopolysaccharide (LPS) for Gram-negative bacteria) [97]. The phages' host range is primarily defining by the match between RBP and bacterial receptor [98].

Nevertheless, adsorption does not automatically translate in successful phage propagation. If a significant quantity of phages is adsorbed by the bacteria, there is a potential for cell-wall damage at vulnerable points, leading to subsequent lysis prior to viral DNA transfer to bacteria. Additionally,

if the adsorption does not occur, lysis intermediated by external factors is still possible due to the release of phage-derived lysins (lytic enzymes that degrade cell walls), that can be employed to induce lysis without the need for adsorption [99,100].

Following adsorption, the penetration (also common in both life cycles) involves the entry of the genetic material of the phage into the bacterial cell through various mechanisms, determined by the virus's configuration. Usually, it consists of tail contraction and break down of the bacterial wall by enzymes that are produced by the phages [101]. Within the bacterial cell, the phage's genetic material is transcribed by RNA polymerases to generate mRNA (messenger RNA) and produce new proteins to take control of the host cell's machinery. In lytic cycle, the metabolic processes of the host will be directed to produce structural phage proteins for new virions and lysis proteins to eventually burst/lyse the bacteria [97,102].

While virulent phages exclusively follow the lytic cycle of replication, temperate phages can follow both lytic and lysogenic cycles of replication. The lysogenic cycle involves the integration of phages into their host genome, as prophages, or as plasmids within the host cell, rather than instantly leading their hosts into cell death [103].

Lysogeny typically advances in three stages, the first is the establishment, which is the incorporation of the phage DNA in the host bacterium's genome and inception into lysogeny. The choice to engage in lysogeny is influenced by genetic compatibility, the physiological condition of the host, and the abundance of phages. Eventually, in a second stage called maintenance, the incorporated phage genome will undergo replication alongside the host's genetic material. As a result, subsequent generations of bacteria will inherit this viral DNA due to vertical gene transfer. The prophage can coexist in the bacteria for generations in a dormant state, however, a third stage called induction can occur if, under adverse environmental conditions and external stressors, the bacterial genetic material is harmed. When this happens, the prophage genes will shift back to the lytic cycle, resulting in the synthesis of fully assembled phages and cell lysis [90,104–106].

The use of phages for therapy requires them to be virulent, the ones that undergo strictly the lytic cycle causing immediate cell death. The application of temperate phages would imply potential adverse effects of lysogenic conversion in bacteria, such as the achievement of new genetic features encrypted by phages that can potentially be pathogenic, including toxins that increase their virulence, or possibly causes for antibiotic resistance [87,107–109].

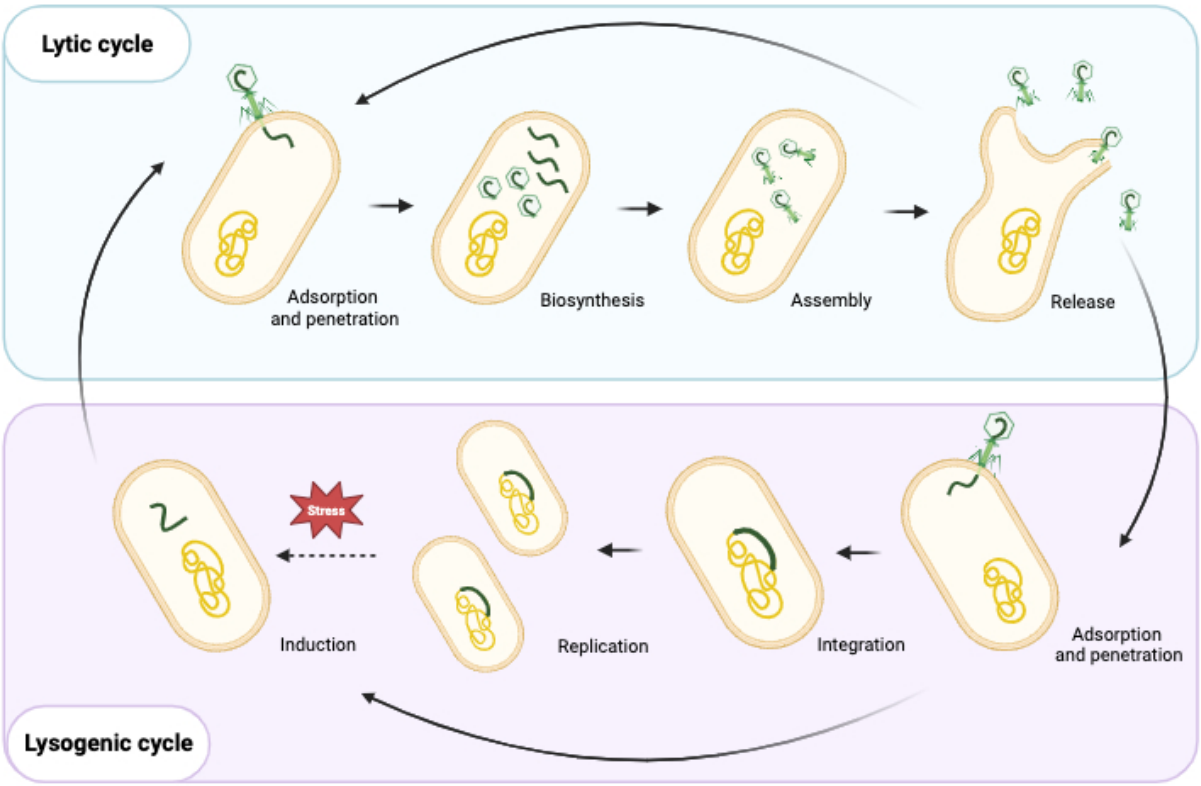

Figure 5 – Representation of two primary life cycles of phages. In the **lytic cycle**, phage replication occurs right after infection, involving the assembly and release of virions, which eventually leads to cell lysis. Each virion can then initiate a new lytic cycle, resulting in a burst of 'productive' infection. In the **lysogenic cycle**, phages have the ability to integrate into the bacterial chromosome and replicate alongside it as prophages until a lytic cycle is initiated by external stressors. Image created with BioRender.

The use of phages towards ETEC infections

Phage therapy is known as direct application of virulent phages to an organism with the aim of causing the lysis of the bacterial pathogen responsible for a clinically significant infection [110]. It is expected that the phages that target a specific bacteria can be found in the same environment, since phages depend on hosts for their survival. Consequently, phages targeting intestinal bacteria in mammals, such as ETEC, can be easily obtained from fecal matter, sewage from wastewater, and runoff from farms [95,111].

Experimental administrations of phages to livestock have demonstrated a positive outcome, resulting in a notable decrease on pathogenic *E. coli* levels or the complete eradication of these strains of animals' microbiome. One of the pioneering case studies showed that, a phage mixture against an enteropathogenic strain of *E. coli* in calves, neonatal pigs and lambs, efficiently lysed their respective hosts to prevent bacteria from establishing in significant numbers in the small intestine shielding them from diarrhea and mortality [112].

In the succeeding years, the exploration of phage administration in feed surges and is proven that delivering phages orally reduced *E. coli* pathotypes (including ETEC) discarding in sheep, ruminants, and pigs. The protection of phages during intestinal passage could enhance this strategy's efficacy, however, the phages were excreted and there were no adverse effects [113,114]. Besides feed administration, the use of suppositories containing a probiotic in combination with phages targeting pathogenic *E. coli* in young calves experiencing diarrhea, resulted in a shortening period of calf diarrhea, effectively eradicating it within 24 – 48 hours following treatment and promoted the activation of immune mechanisms to increase resistance to infections [115].

In a recent study, a phage cocktail (mixture of a certain number of phages), targeting a specific MDR *E. coli* strain (isolated from diarrheal pigs), was administrated to a group of weaned piglets and it occurred a decrease in fecal *E. coli* counts after seven days of phage treatment. It was also proved that, 24 hours post final phage administration, the normal gut microbiome was restored and, in the initial weeks post treatment, the piglets that received a higher dosage of phages, demonstrated growth improvement [116]. A similar outcome was early observed in a research where a combination of phages were administrated orally in weaned pigs with ETEC infection. The treatment notably reduced duration and severity of diarrhea and the presence of ETEC in feces, without affecting piglets' normal *E. coli* flora [117].

Regarding the evaluation of phage therapy targeting pathogenic *E. coli* in humans, Alam *et al.* [118] administered a mixture of T4-like phages, into 15 healthy adults from Bangladesh, at different dosages in order to anticipate phage safety. The results revealed lack of phage replication due to the absence of the targeted bacteria and did not reduce the normal microbiota in feces, indicating that, even phages' high dose did not show adverse effects in the healthy adults that participated on the study.

Similar outcome was reported by Febvre *et al.* [119] where adults with self-reported gastrointestinal distress were administered with a mixture of *E. coli*-targeting phages. It was reported reductions in fecal *E. coli*, even so, there were no significant changes to normal microbiota. Additionally, after treatment, the results showed a significant increase in CO_2 (normally diarrhea is associated with low blood CO_2 levels), and the aspartate aminotransferase and alanine aminotransferase levels were lower compared to placebo (these enzymes levels increase after exposure to systemic inflammation and tissue damage). Despite this promising results, the reduced efficacy of phage titers after passing through gastric acid was recognized as an additional potential cause for possible treatment failure [120]. On that matter, further extensive clinical trials are necessary

to establish phage therapy as a routine and widely accepted form of treatment for ETEC, rather than merely an experimental approach [121].

Based on current research, phage therapy is considered as a valuable approach for controlling and treating infectious disorders caused by significant pathogens, especially in the swine industry. Nonetheless, there is still a need to understand the dynamics involving phages and its hosts, so that an effective, safe and non-toxic phage therapy will be a possibility for human bacterial infections treatment [122].

Benefits and challenges of bacteriophage therapy

Several benefits associated with phages make them a viable alternative therapy method. Those assets include the minimal intrinsic toxicity and reduced impact on the typical microbiota, the auto "dosing" effect (phages' capability of multiply near hosts and decrease as bacterial cells are eliminated), the lower risk of resistance development, the easy access and quick detection of new phages, the versatile formulation and application and, most important, the clearance of structured microbial communities (biofilm) [123].

The most appealing attribute of phages regarding therapy application lies in their specific mode of action, allowing them to target and eliminate only the recognized pathogen. This advantage mitigates the elimination of potentially beneficial bacteria, the proliferation of secondary pathogens, and the development of antibiotic-resistant strains, those being primary concerns linked to antibiotic administration [124]. Although host specificity imparts numerous benefits, is important to consider that the effectiveness of phage therapy might be reduced when dealing with polymicrobial infections. The use of phage cocktails can potentially address this concern, if the target bacteria is known, which is frequently not the case in treatment strategies [125].

Regardless all the positive outcomes and benefits of phages, like all other therapeutic approaches, this type of therapy also comes with its limitations, some of them of an economic character. Multinational pharmaceutical companies may hesitate to invest in phages due to high costs associated with clinical trial and unclear regulatory requirements, potentially causing significant delays in phage therapy becoming mainstream [126,127].

Furthermore, despite existing regulation and guidelines for bacteriophage usage as therapy for animals (Regulation (EU) nº 2019/6, Commission Delegated Regulation (EU) 2021/805 and EMA/CVMP/NTWP/32862/2022), the poor regulation for humans' phage usage as biological

medicinal product, limits their accessibility for health purposes. For that reason, efforts are crucial to establish best techniques for safe phage research and development. This, in turn, will be key to make these experimental treatments a widely accepted complement to antibiotics [127–129]. Likewise, the potential of phages to enhance swine production will remain incomplete without full comprehension of phage resistance, interactions between phage and host and the microbial environment [85].

However, the main concern rises upon the development of bacterial resistance mechanisms against phages, which can cause a restricted host range [130,131]. This problem affects the primary objective of phage therapy to optimize a quantity of phages that reach and infect a wide range of bacteria, leading to a reduction of their levels to clinically insignificant, all while avoiding any undesired side effects [132]. Besides this limitation, their high bacterial specificity may prevent phage infections in clinical cases triggered by multiple pathogenic bacteria and, phage degradation while interacting with human metabolism (gastrointestinal system and liver) and the extraction of phages from the circulatory system, may lead to difficulty in maintaining sufficient phage concentrations at the target sites for bacterial infection [133].

1.4. Anti-phage defense systems

Roughly, it is believed that 10^{23} phage infections happen every second, creating substantial selective pressure on bacteria. Consequently, in response to the challenge of phage infection, bacteria have developed multiple antiviral protection strategies that can be referred to as the "bacterial immune system" working to shield themselves from impact of phage infections [134,135].

Despite anti-phage defense systems (APDS) in bacteria being often acquired through HGT, and expected to accumulate in genomes, they are often lost relatively quickly in evolution due to potential adverse effects and fitness costs, in the absence of virus pressure. So, the frequent acquisition and shedding of APDS leads to an irregular pattern in microbial genomes, even among closely related strains and, consequently, becomes hard to predict the type of APDS that can be present in bacteria [136]. So, a thorough characterization of APDS can contribute to a deeper comprehension of the obstacles in the practical use of phages and facilitate the improvement of phage technology and therapy [137].

Defense systems ensuring cell survival

There are APDS operating throughout almost every phase of the lytic cycle [138]. In adsorption, bacteria can modify or downregulate the configuration of their cell surface receptors, for instance, in *E. coli*, it was reported the use of lipoproteins to hinder phage adsorption by a plasmid-encoded outer membrane protein (TraT) that conceals or alters the conformation of outer membrane protein A (OmpA), a receptor for numerous *E. coli* phages [139]. This phenomenon was also observed in phage-encoded lipoproteins (Llp), synthesized by the host cell, that prevents binding to its own receptor (FhuA) [140].

Additionally, to prevent phage adsorption, bacteria also can produce extracellular polysaccharides to create a physical barrier that separates phages from their receptors. It is the case of the K1 extracellular polysaccharide capsule of some *E. coli* strains, that has been demonstrated to directly impede the attachment of phage T7 to its LPS receptor [141]. Certain bacterial strains have capsules formed by exopolysaccharides (EPSs) and plasmids carrying EPSs can be acquired though HGT, leading to the development of a bacterial phenotype that blocks phage adsorption [142].

Another adsorption-related defense mechanism is through synthesis of competitive inhibitor molecules that outcompete phages for receptors. Within *E. coli*, FhuA serves as both an iron transporter and phage receptor (T1 and T5), so, under nutrient deficit, an antimicrobial peptide (MccJ25) is generated that competitively obstructs FhuA, thereby impeding the initiation of T5 adsorption [143].

In the subsequent stage, the penetration of viral DNA into the bacteria can be blocked by proteins anchored or associated with the cell membrane that involves a superinfection exclusion (Sie) system, usually dependent on proteins encoded by prophages from a previous lysogeny infection. These systems are believed to offer a significant advantage to the bacterium due to extending safeguard to the surrounding population [144]. An example of Sie systems, encoded by phages T4, are Imm and Sp proteins and their mode of action is represented in **Figure 6**. Through alteration of the conformation of the injection site, Imm protein blocks the transfer of phage DNA into the bacterial cytoplasm. However, although the Imm function accounts for roughly 80% of Sie, the expression of the Sp protein is necessary for complete phage exclusion. The membrane protein Sp inhibits the activity of T4 lysozymes, potentially preventing the degradation of the host cell wall and the subsequent penetration of phage DNA [145].

Despite these strategies, if the phage genome still enters the bacterial cytoplasm, these microorganisms can cut the genomes of the invading virus using mechanisms such as restriction-modification (RM) and CRISPR (clustered regularly interspaced short palindromic repeat) - Cas (CRISPR-associated) systems [146].

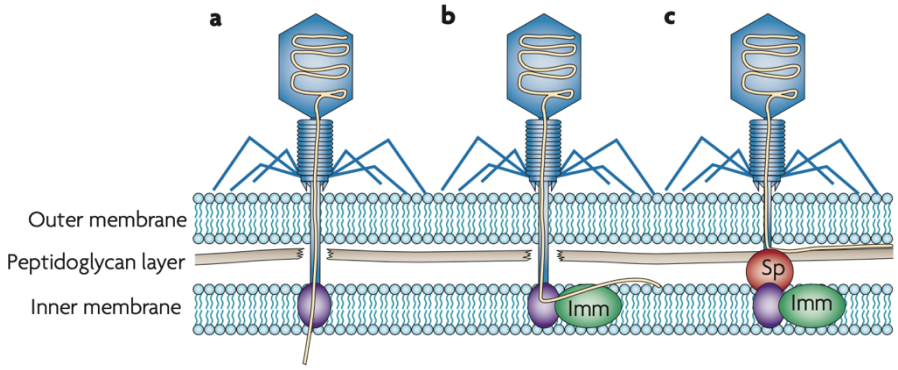

Figure 6 – Representation of a Sie system involving proteins Imm and Sp encoded by coliphage T4. In (**a**) is represented the DNA injection during penetration phase on a normal phage infection. In (**b**) is represented the effect of protein Imm on phage infection, which inhibits the movement of phage DNA into the cytoplasm. In (**c**) is represented the Sp protein, which obstructs the degradation of peptidoglycan, restraining the DNA between the peptidoglycan layer and the outer membrane. Image adapted from Labrie *et al.* [138].

Regarding RM systems, it contains enzymes with dual functions, a methyltransferase which chemically alters the hosts' DNA bases, at specific recognition sites, by adding a methyl group in adenine or cytosine and, a corresponding restriction endonuclease which identifies and cleaves into harmless fragments the DNA sequence that lacks methylation (viral DNA) [144,147].

RM systems fall into four types, represented in **Figure 7**. Type II features simple protein structures and endonuclease subunit only needs Mg^{2+} for activity, although a subset of Type II was created (Type IIS) due to the necessity of S-adenosylmethionine (AdoMet) by the methyltransferase enzyme. Conversely, Type I (described in *E. coli* strains) and Type III are multifunctional proteins capable of cleaving and methylating unmodified DNA, differing in subunit number (three for Type I, with an additional specificity subunit, and two for Type III) and AdoMet requirement (essential for Type I and accelerating for Type III). Ultimately, Type IV differs in structure, due to the fusion of both enzymes (into one endonuclease protein), the endonuclease activity is stimulated by AdoMet and also, it is the only type that cleaves the DNA that has been methylated [148,149]. Type I RM system are the most common in prokaryotes and, the most analyzed example is EcoKI, encoded in *E. coli* K-12, which has a capacity to restrict phage propagation by factors ranging from 10^3 to 10^8. In the same bacterial strain is encoded the best studied Type IV system, namely, McrBC which is the singular nuclease to use guanosine triphosphate (GTP) for cleavage [147].

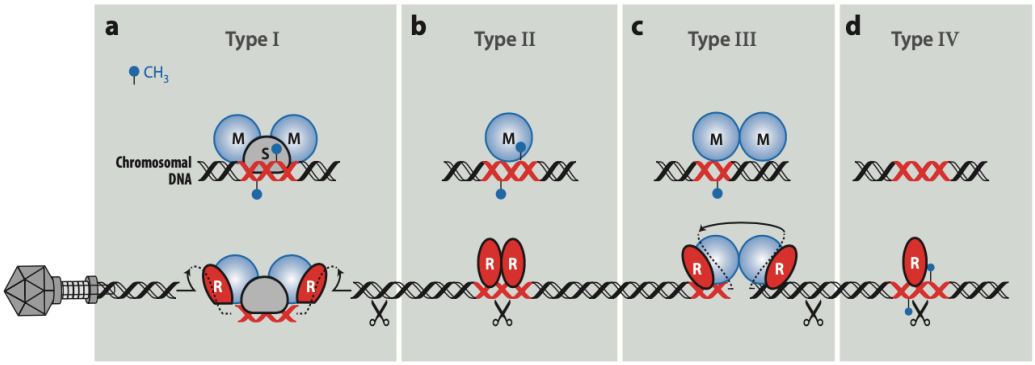

Figure 7 – Representation of different types of RM systems, including chromosomal and viral DNA, restriction (R), modification (M) and specify (S) subunits, and a methyl group (CH_3). In (**a**) is represented a Type I RM system, M and S subunits are required for binding and methylation, and R subunits bond for posterior cleavage. In (**b**) is represented a Type II RM system, a single M subunit binds to DNA for methylation and, separate R subunits later cleave the unmethylated DNA. In (**c**) is represented Type III RM system, two M subunit binds to DNA for methylation and, after, a complex with R subunits is formed for cleavage. In (**d**) is represented Type IV RM system, a "fused" R subunit is formed to recognize and cleave methylated DNA. Image adapted from Dy *et al.* [146].

As for CRISPR-Cas systems, it is a prokaryotic adaptive immune system that shields bacteria against phage infections by preserving a trace of viral DNA within their chromosomes. This system consists of a CRISPR array, which includes spacers that match foreign DNA and is surrounded by repeats, and an operon containing *cas* genes responsible for the processing of the CRISPR array and the cleavage of DNA targeted by the spacers. The CRISPR-Cas operate through a process involving at least two primary stages, first is the adaptation, during which new spacers are incorporated at the leader end of the CRISPR locus and, the second stage is the interference, where the CRISPR-Cas system can target invading DNA or RNA [134,150–152].

The CRISPR-Cas activity against bacteriophages was proved in a study where the CRISPR loci of *Streptococcus thermophilus* strains was altered during the natural generation of phage-resistant mutants. Upon developing resistance to bacteriophages, the CRISPR locus underwent modifications through the integration of new spacers, seemingly originating from phage DNA (proto-spacer). The resistance pattern against phages appeared to be linked to the spacer composition, where strains possessing CRISPR spacers that exhibited 100% identity to conserved proto-spacer in the infecting phages, demonstrated resistance to those phages. Moreover, with the insertion of multiple spacers into the CRISPR loci, the levels of phage resistance increased. These observations suggest that the CRISPR locus undergoes dynamic and rapid evolutionary adjustments influenced by exposure to phages [153]. Although the CRISPR-Cas system remains fallible, since these phage-resistant bacteria mutants remain susceptible to phages lacking the specific proto-spacer in their genomes, also, phage mutants with a single mutation or deletion in their proto-spacer can avoid CRISPR activity and complete their lytic cycles successfully (**Figure 8**) [138].

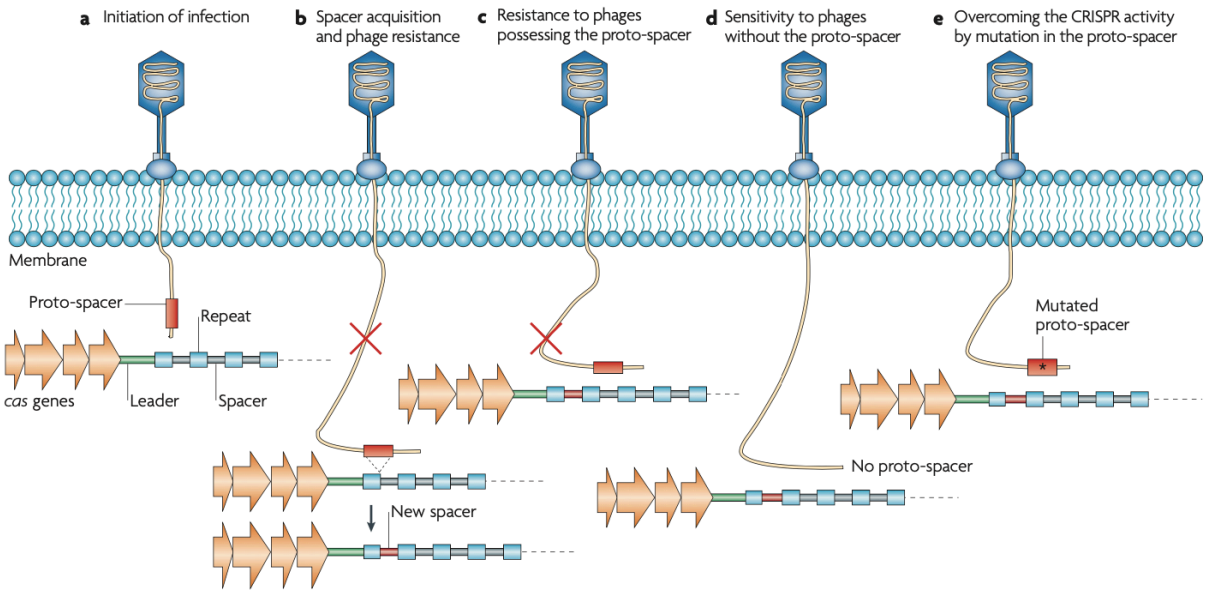

Figure 8 – Representation of a CRISPR-Cas mode of action (related to occurrences in *Streptococcus thermophilus*). In (**a**) is represented the initiation of a phage infection that undergoes lytic life cycle. In (**b**) is represented the formation of bacteria mutants that endure the infection. The CRISPR locus in such phage-insensitive mutants harbors an extra repeat (duplicated from the CRISPR locus), along with a novel spacer (similar to proto-spacer) acquired from the genome of the infecting phage. In (**c**) is represented the phage resistance by phage-insensitive bacteria mutants upon recently acquired repeat-spacer unit. The phage with 100% nucleotide identity to the new spacer-repeat unit will be deactivated and phage infection process is blocked. In (**d**) is represented the sensitivity of the bacteria mutants to phages that lack the proto-spacer in their genome, leading to a successful phage infection. In (**e**) is represented the sensitivity of the bacteria mutants to phages that carry a mutation in their proto-spacer that are also able to complete their lytic cycle. Image adapted from Labrie *et al.* [138].

Despite the APDS mentioned above being the main systems studied that ensure cell survival, recent research has revealed that the defense arsenal within the prokaryotic genome is far more extensive than previously thought and capable of forming "defense islands". New defense systems with diverse mechanisms have been revealed, employing antiviral molecules that inhibit transcription (Viperin), or membrane integrity to delay cell lysis (Dynamins), or even viral DNA recognition and cleavage (Argonaute). Nonetheless, a lot of these “new” defense systems remain unclassified due to unknown function [135,137].

Defense systems promoting cell death

Before defense mechanisms in prokaryotes targeting foreign DNA come into play, phage-infected bacteria can induce premature cell death through a process called abortive infection (Abi). Unlike most strategies for APDS, that ensure the bacterial cell’s survival when confronted with a viral threat, Abi systems cause infected cells to die before the phage completes its replication cycle. This acts as a sacrificial measure, by preventing the production of phage offspring and safeguarding

nearby uninfected cells. Abi systems are commonly found in mobile genetic elements such as prophages and plasmids (**Figure 9**) [137,154].

Abi systems are biologically relevant when the initial defensive mechanisms have proven ineffective and the infected cell's prospects of survival are already minimal, consequently, Abi systems typically start functioning when the phage are at late stages of its infection cycle, because Abi is triggered as a final defense measure. Generally, Abi are latent proteins that become active in response to phages and disturb crucial cellular metabolic processes [146,155].

The Abi impact can be accomplished by means such as membrane depolarization or the inhibition of translation and both of those mechanisms were proven to affect phages that targeted *E. coli* [154]. One of the well-studied Abi mechanisms in *E. coli* is the Rex exclusion system originating from prophage lambda. The Rex first component protein triggers the creation of ion channels, leading to cytoplasmic membrane depolarization through Rex second component protein [156].

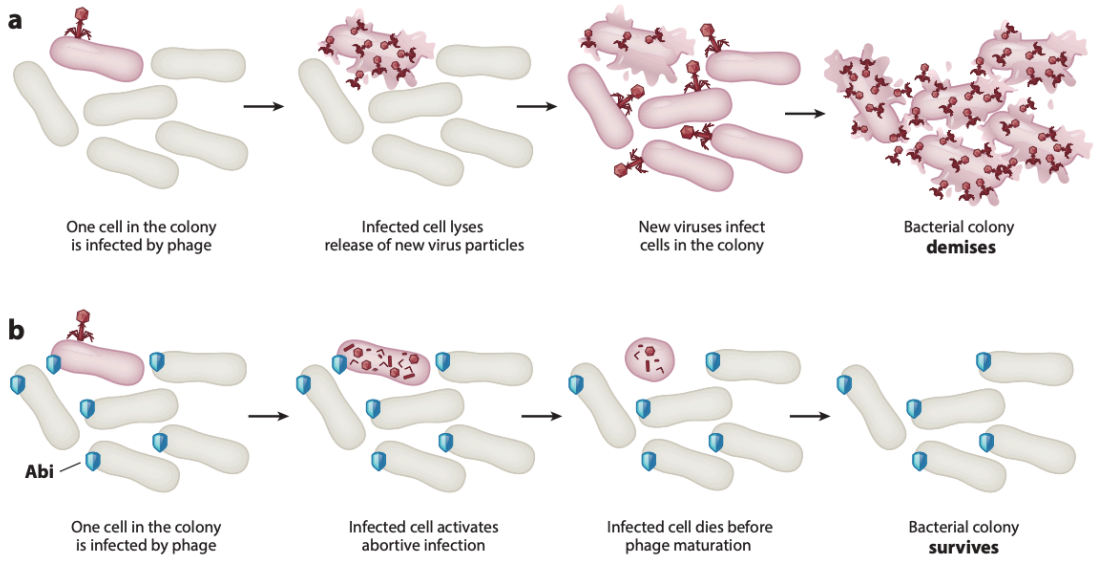

Figure 9 – Representation of a general Abi strategy in bacteria. In (**a**) is represented an infection of a bacterial culture lacking an Abi system. The phage infection of the colony is successful, and the bacteria is eliminated. In (**b**) is represented an infection of a culture possessing an Abi system. The infected bacteria committed "suicide" after phage infection without jeopardizing the colony's livability. Image adapted from Lopatina *et al.* [155].

Toxin-antitoxin (TA) systems are a form of abortive infection found in bacteria resulting in bacterial demise or dormancy upon activation triggered by phage infection. TA systems consist of a toxin and a counteracting antitoxin, which prevents the toxin from causing harm during regular bacterial growth. However, during a phage infection (stress situation), the antitoxin synthesis is disrupted, allowing the unbound toxin to inhibit bacterial translation through RNA-mediated mechanisms, causing either inactivity or cell fatality [154,157]. It has been identified four primary

categories of these genetic elements, distinguished by the characteristics of the antitoxin and how it restricts the activity of the respective toxin protein [158]:

- Type I - antitoxins are RNA molecules that control the levels of active toxin protein by obstructing the translation of the toxin mRNA;
- Type II - antitoxins are proteins that actively attach to and suppress the toxin protein;
- Type III - antitoxins are RNA molecules that control the levels of active toxin protein by immediately suppressing the toxin protein;
- Type IV - antitoxins are proteins that deactivates the toxin protein without direct engagement (reversing its effects on the targets).

It is also notable to mention that, in some research, that has been evidences of certain TA systems association with mobile genetic elements, such as prophages, making them especially susceptible to HGT. A result of that could be the inherence of TA systems between strains [159]. Also, it has been identified a group of abortive infection proteins encoded by prophages that effectively protect bacterial populations against phage outbreaks by, presumably, interact physically with phage DNA to inhibit replication by blocking a replication initiation site [160].

In addition to Abi and TA systems, recent studies reported other types of similar defense mechanisms that trigger cell death in order to protect the remain bacteria population from phage infection. These findings have brought to light new defense systems with diverse mechanisms, including retrons, which they unite with RecBCD proteins to maintain dormancy and, upon phage infection and suppression of RecBCD functions, the retrons become active, inducing cell death [161].

Another type of Abi-like defense systems are CBASS (Cyclic Oligonucleotide-based Anti-phage Signaling System) and Pycsar (Pyrimidine Cyclase System for Anti-phage Resistance), that generate a secondary messenger molecule (cyclic di- and tri- nucleotides for CBASS, and cyclic nucleotide monophosphates for Pycsar) to trigger effectors and initiate cellular lysis through membrane disruption, DNA degradation, or alternative mechanisms [162,163].

Short prokaryotic Argonautes (Agos), defense-associated sirtuins (DSR) and Thoeris systems can efficiently reduce the cellular NAD levels by identifying foreign plasmid or phage DNA, or even phage tail tube proteins, disrupting NADase activity, consequently, the NAD depletion in the cells inhibits phage infection [164–166]. These are only a few of the recently identified and examined APDS, associated with abortive infection activity, however, the diversity of these systems is vast, and

research persists to characterize the functions of new types of defense systems emerging within bacterial cells.

Counter-defenses of phages

In response to all possible APDS that bacteria may encode due to phages' exerted pressure, phages also undergo co-evolution to overcome these obstacles, leading to an ongoing and unpredicted molecular competition [144]. Nevertheless, considering the variety of defense systems across various strains, even among identical species, it is evident that phages must encompass a range of distinct counter-defense mechanisms to achieve a wide host compatibility [136]. Currently, phages have been discovered to possess genes capable of circumventing defense mechanisms such as RM systems, CRISPR-Cas systems, toxin-antitoxin systems, Thoeris, Pysar, and CBASS [18]. Despite that, phages can also bypass defenses that inhibit adsorption by adapting their tail fibers to identify new or modified receptors and obtain hydrolyzing enzymes, allowing them to penetrate the extracellular matrix or to degrade capsules, or even to improve the surface properties of RBP. These enzymes can either be located within the tail fibers or dispersed following phage burst to assist in infecting neighboring bacteria with new viral offspring [167–170].

In a practical experiment, it was observed that recently isolated phages exhibit superior antibacterial properties compared to those identified earlier, through extended evolutionary interactions between phages and their hosts. Consequently, co-culturing bacteria and phages for the purpose of isolating evolved phages could prove to be an efficient method for obtaining more effective phages [171].

1.5. Aims

Infections resulting from bacterial pathogens can lead to severe human health issues and ETEC stands out as a significant contributor of diarrheal illnesses in both farm animals and humans.

The misuse of last resort antibiotics, intended for bacterial treatment, is causing the emergence of antibiotic resistance. This phenomenon caught the attention of public health authorities and now is urgent to find sustainable and efficient solutions to combine with or replace conventional antimicrobials.

Bacteriophage therapy, due to their high specificity, non-threatening properties to animals and plants, and widespread dispersion and autonomous reproduction, has proven to be an optimal candidate to replace antibiotics in bacterial treatments. However, the rise of APDS seems to be in the way of the full potential that phage therapy can achieve.

The main goal of this study was to isolate and characterize ETEC-infecting phages, investigate the potential effect of hosts' APDS in their lytic activity and identify counter-defense proteins whitin the phages genome.

With this purpose in mind, it was intended to isolate new phages that target ETEC strains and characterize them though lytic spectra, morphology and genome analysis. At last, the objective was to comprehend the main APDS present in ETEC strains, in order to identify phages that encode proteins counteracting these defense mechanisms. Based on this understanding, the plan was to select the most suitable phages for potential therapeutic purposes, which could then undergo further examination.

2. Materials and methods

To achieve our objectives with this study, a process named **phage enrichment** was performed to isolate phages. This procedure is characterized by the inoculation of probable host bacteria with an external source of phages (samples of sewage from wastewater). In cases where putative phages were detected, **phage isolation** was conducted and, if successful, they were able to undergo phage production to storage proper quantities of the phage for intended analysis. **Phage characterization** process was done to determine the infectious potential of the isolated (phages' host range, efficiency of infection, and adsorption rate). Meanwhile, it was also performed an assay to **monitor anti-phage defense systems**. Aiming to understand better the phages' suitability for therapy and the phage-bacteria interaction the genome of phages and bacteria were subjected to ***in silico* analysis**.

2.1. ETEC strains and growth conditions

In the present study, a total of 98 diarrheagenic *E. coli* strains were used, originated from clinical cases from Spanish[172] and Portuguese pig farms (provided by the company ALS). All the ETEC strains used are mentioned in **Table A** (Annex 1).

For cultivation, all strains of ETEC were grown in sterile medium containing lysogeny broth (LB) from Nzytech, in liquid cultures (25 g/L), or in plates with LB agar (25 g/L and 1.5% (w/v) of agar). The incubation was done at 37° C and, in the case of liquid cultures, under agitation, respectively 120 rotation per minute (rpm). These procedures were done considering *E. coli* 's optimal growth conditions [173].

To prepare bacterial lawns, LB agar plates were used and, the pre-grown bacteria was poured into the plate after inoculating in NaCl (0.9% w/v). Finally, the top agar (LB and 0.5% (w/v) of agar) was layered over to solidify.

All the procedures were done under sterile conditions as well as the utilization of sterilized reagents and material, given the significance of preventing contamination to ensure trustworthy outcomes.

2.2. Phage enrichment

For phage enrichment, 11 strains of ETEC grown overnight were incubated until reaching exponential phase, which, according to the growth curve of *E. coli* presented in **Figure A** (Annex 1), it happens approximately at an optical density at 600 nm (OD_{600}) of 0.5.

Following that, 200 mL of wastewater samples were inoculated in 200 mL of double concentrated LB (LB 2x) and this mixture was divided into four batches (1 and 3 with 150 mL; 2 and 4 with 50 mL). The bacterial suspensions in log-phase were added to the batches following the scheme: Batch 1: EC16 + EC22 + EC36 + SP130 + SP143; Batch 2: EC65 + SP127; Batch 3: EC52 + EC53 + EC34; Batch 4: EC42. This division was based on the results of a previous competitive assay among strains (to ensure no prophages are released), these results are represented in **Table B** (Annex 1). The batches were incubated overnight under agitation.

Following complete incubation, the four enriched samples were collected and centrifuged with refrigeration (9000 × *g*, 4°C) for 10 minutes. The supernatant of each sample was then filtrated (PES filters of 0.2 µm) and collected for further analysis, while being preserved at 4°C. And that concludes the phage propagation stage [174].

2.3. Phage isolation

During the phage isolation process, we performed a spot test of the four different enrichment samples with the strains used in each batch. After matching the phage that propagated with its corresponding host, it was conducted a phage plaque isolation and phage production (soft-agar overlay technique)

Spot test on bacterial lawns

Initially, the bacterial lawns of the strains used in phage enrichment were prepared. Subsequently, 10 µL of the four different supernatants were poured into the corresponding plates containing the bacterial lawns. Lastly, all the plates were incubated overnight [175].

Phage plaque isolation

For phage isolation, each plate with positive results for the presence of phage plaque formation (inhibition zones in the bacterial lawn), was picked with a sterile toothpick in those inhibition zones. In a new sterile plate with the bacterial lawn, the toothpick was stuck multiple times in a horizontal line. Using a sterile paper strip, the line done with the toothpick was streaked downward along the plate, changing the paper strip in each streak and ensuring that it touches the previous streak. The plate was incubated overnight. This procedure was repeated two more times until equal phage plaque morphology, ensuring the suspensions had only a single type of phage (**Figure 10 - A**) [176].

Phage production

Initially, bacterial lawns were made from each bacterial hosts, on five new sterile plates. Each isolated phages were picked by a sterile toothpick (in an area with substantial inhibition zones) and stuck in the new plates. In all plates, three to four horizontal lines were perforated with the toothpick, picking the plaques again every time it was done a new line. A sterile paper strip was used to spread the phage across the plates, ensuring that it passed through the lines already formed. After that, all the plates were incubated overnight (**Figure 10 - B**). Five mL of sterile SM Buffer (5.8 g/L NaCl; 2 g/L MgSO4·7H2O; 50 mL/L 1M Tris-HCl pH 7.5) were added into all the plates and they were incubated for five hours at 4°C under slight agitation (50 rpm). After incubation, the SM buffer, from the plates that corresponded to the same phage, was collected into the same sterile falcon tube and chloroform was added to a final concentration of 10% (v/v). The falcon tubes were centrifuged with refrigeration (9000 × *g*, 4°C) for 10 minutes. Following centrifugation, the supernatant was collected, filtered (PES filters of 0.2 µm) and stored in 4°C for further analysis. Once this process was concluded, we had stocks of different isolated phages. With these isolated phages, we were able to move on into the phage characterization process [177].

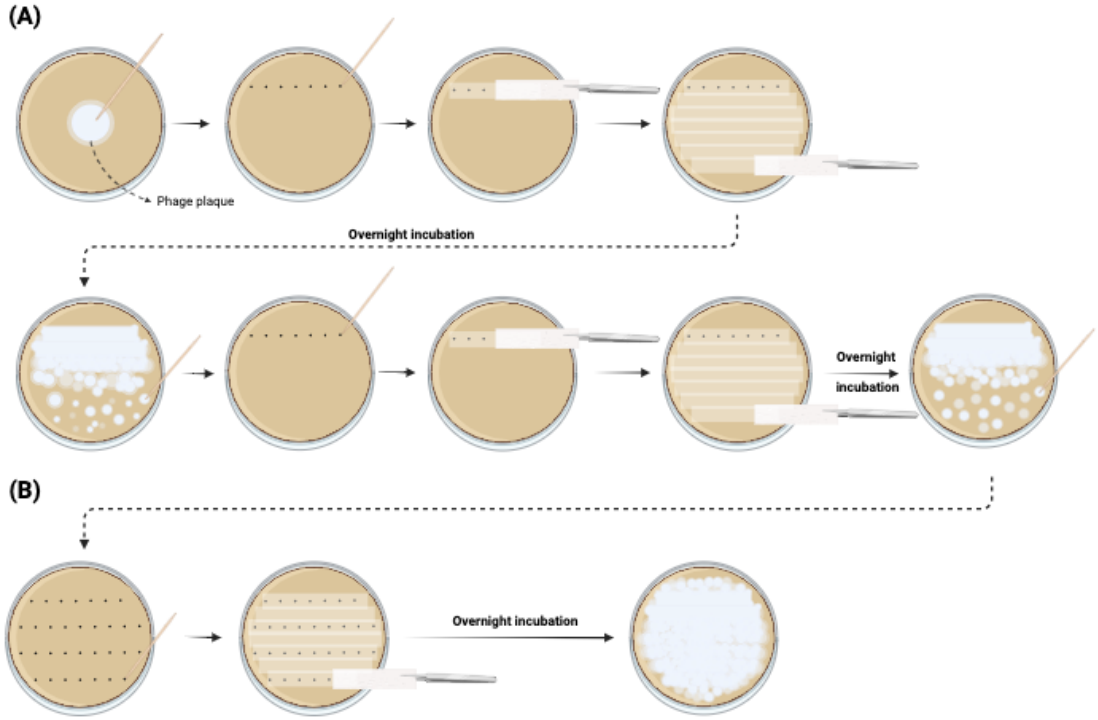

Figure 10 – Representation of the methods used during the phage isolation and production stages. In **(A)** is illustrated the steps done for isolation, until only one morphology of phage plaques was shown. In **(B)** is illustrated the initial steps of production, before the addition of SM buffer, of a single phage plaque morphology. Image created with BioRender.

Phage titration

To assess phage concentration, an adaptation of the small drop plaque assay was used [178]. Firstly, we prepared the bacterial lawns of the hosts from each phage. Subsequently, each isolated phage was subjected to serial dilutions (1:10) in sterile SM Buffer. In the plates containing the hosts' bacterial lawns, it was added 10 µL of the last four dilutions of the respective phage and the plate was tilted vertically so it would form swab along the plate (facilitate plaque counting). As soon as all the plates dried, we left them to incubate overnight [179].

Following incubation and formation of phage plaques, they were counted in a single dilution that would contain, approximately, between three to thirty plaques. The phage titration was obtained following the subsequent **equation (1)**.

$$Phage\ titer\ (PFU/mL) = \frac{N°\ of\ phage\ plaques \times Dilution\ factor}{Volume\ of\ phage\ plaqued} \qquad \textbf{Equation (1)}$$

2.4. Phage characterization

Throughout this stage, phages were characterized according to their lytic spectra in a wide spectrum of hosts, their efficiency of plating (EOP) and their adsorption rate.

Lytic spectra analysis

For the host range determination, a set of 95 strains of ETEC were acquired for this characterization study, which included eight of the strains employed in phage enrichment mentioned above (**Table A** - Annex 1). Each isolated phage (≈ 10^8 PFU/mL) was spotted (10 µL) onto freshly prepared bacterial lawns. Following that, as soon as the plates dried, they were incubated overnight. After incubation, the plates were analyzed for the formation of clear spots in the bacterial lawn [180].

Efficiency of plating assay

The positive results observed in the lytic spectra analysis were then submitted to the EOP assay. In this technique, serial dilutions (1:10), in sterile SM buffer, of each of the phage suspension were spotted (10 µL) on top of the bacterial lawns of the phage-susceptible strains observed in the lytic spectra analysis. The plates were then tilted vertically to create an even spread across the surface, facilitating the counting of plaques. The plates were incubated overnight as soon as they were dried [181,182]. Following incubation, the concentration was assessed as described before (**Equation 1**). The EOP for a given strain is defined by the comparison of phages' titer in each phage-susceptible strain to the phage titer of their respective host according to **Equation (2)** [183].

$$EOP = \frac{Phage\ titer\ of\ susceptible\ strain}{Phage\ titer\ of\ phage\ host} \qquad \textbf{Equation (2)}$$

Results were categorized as follows: highly efficient (1.0 = EOP > 0.5), moderately efficient (0.5 ≥ EOP > 0.2), lowly efficient (0.2 ≥ EOP > 0.001), inefficient (EOP ≤ 0.001) and lysis from without (LFW) [184].

Morphological characterization

The isolated phage suspensions were taken to a transmission electron microscopy (TEM) for morphological observation [185]. Briefly, phage particles were collected after centrifugation (25,000 × *g* at 4 °C, for 1 hour). The pellet was washed twice with tap water before centrifugation. Furthermore, phage was deposited on copper grids with carbon-coated Formvar films, stained with 2% uranyl acetate (pH 4), and analyzed using a Jeol JEM 1400 transmission electron microscope[186].

Adsorption assay

To improve knowledge on the infectious process of the phages, an adsorption assay was performed using two selected strains that showed LFW with the same phage (EC42 and EC70). The pre-inoculum of the selected strains was prepared, separately, in 5 mL of sterile LB and then placed in incubation under agitation overnight.

In the following day, the pre-inoculums were diluted in LB (1:100) each and submitted to agitation during incubation until it reached the exponential phase, which corresponded to an OD_{600} of approximately 0.5 (**Figure A** - Annex 1). After that, the bacterial suspensions were infected at a multiplicity of infection (MOI) of 0.001 (calculated according to **Equation 3**), knowing the concentration of the bacteria according to its calibration curve (**Figure B** – Annex 1), and incubated under agitation for 10 minutes. During the incubation time, samples of 100 µL of the suspensions were collected at 0, 5 and 10 minutes, after the infection with the phage, immediately diluted in SM Buffer (1:10), treated with the same volume (100 µL) of chloroform and vortexed to ensure well-distributed mixing. Subsequent centrifugation was done in the samples (8000 × *g* for two minutes) and the top aqueous phase was collected into a new tube. Immediately after, the proper volume of SM buffer was added to obtain serial dilutions (1:10) and the last four dilution were plated in plates containing only the phages' host to evaluate phage titer. Finally, all the plates were incubated overnight [187,188].

$$MOI = \frac{Phage\ concentration}{Bacterial\ concentration} \qquad \textbf{Equation (3)}$$

Additionally, the assay had a positive (C1) and negative control (C2), where a phage suspension with the phage host (C1) and a phage suspension with no bacteria (C2) were subjected to the steps mentioned above. The adsorption rate was calculated in accordance with **Equation 4** and the results were compared using t-test and considered statistically different if p - value ≤ 0.05 or statistically similar if p - value > 0.05 [172]. The adsorption rate of the phages, in terms of probabilities of adsorption *per* unit time, was also estimated though the calculation of adsorption constant (K_a) in *per* mL *per* minutes using the **Equation 5**, where P represents the phage titer at the end of the time interval used in the assay (10 minutes), P0 is the initial phage titer, B is the host cell concentration in colony forming units *per* mL (CFU/mL) and t represents the time in which the majority of phages were adsorbed, in minutes [189].

$$\%\ Adsorption = \frac{(Phage\ titer\ of\ C2 - Phage\ titer\ of\ sample)}{Phage\ titer\ of\ C2} \qquad \textbf{Equation (4)}$$

$$K_a\ (mL^{-1}min^{-1}) = \frac{-\ln(P/P0)}{B \times t} \qquad \textbf{Equation (5)}$$

2.5. Abortive infection assay

In order to investigate the event of abortive infection mechanisms against phages, one bacterial strain was selected after its positive results in the adsorption assay done earlier. The host of the phage and a strain that did not show any adsorption were also used for comparison. The assay was done in each strain individually and adapted from Vassallo *et al.* [190]. The progress of the bacterial concentration (by measuring the OD_{600}) was monitored under three different scenarios: 1 - without the addition of phage (control); 2 - with addition of phage in high concentrations (high MOI); 3 - with the addition of phage in low concentrations (low MOI).

Firstly, a pre-inoculum of the bacteria strain was prepared in LB and left at incubation with agitation overnight. The next morning, the pre-inoculum was diluted (1:100) in LB and returned to incubation until exponential phase was reached (OD_{600} ≈ 0.5 according to **Figure A** - Annex 1). The bacterial suspension was divided into three eppendorf tubes (epp) to evaluate the different scenarios

mentioned before, in epp number 1 bacterial suspension and SM Buffer were added, in epp number 2 bacterial suspension and phage titer reflecting a high MOI (50) were added and, in epp umber 3 bacterial suspension and phage titer reflecting a low MOI (0.001) were added. To achieve the desired MOI (calculated according to **Equation 3**) bacterial and phage concentration were adjusted. After that, the resulting suspensions were vortexed and transferred to a 96-well plate (200 µL) in triplicate, and placed in a microplate reader (Cytation3, Biotek) with previously determined settings: OD_{600}, 37°C, vertical vigorous agitation, reading every 15 minutes for five hours. The assay is schematically represented in **Figure 11**. The results of the assay were compared using t-test and considered statistically different if p - value ≤ 0.05 or statistically similar if p - value > 0.05 [172].

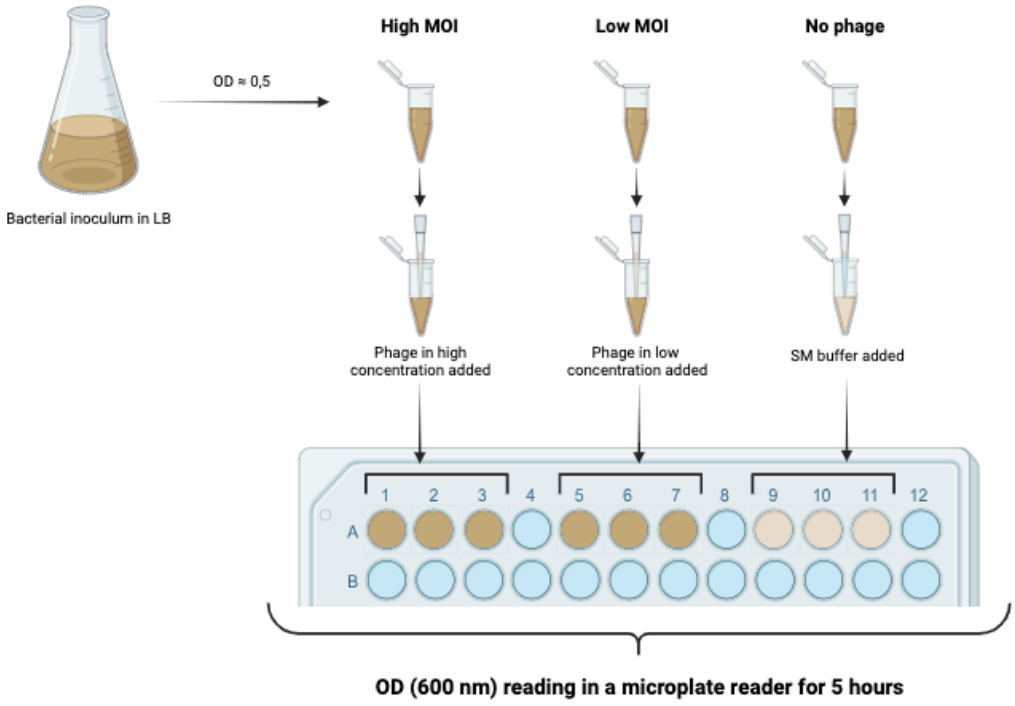

Figure 11 – Representation of the Abi assay (for a single strain). It is represented the main steps after the dilution of the bacterial strain pre-inoculum. Image created with BioRender.

2.6. DNA extraction

The DNA of all phages was extracted using either the ZR viral DNA kit™ (ZYMO Research), following the manufacturer's instructions, or through the phenol/chloroform method, depending on the efficiency of the methods used for a given phage. The bacterial DNA purification was accomplished by utilizing the Quick-DNA™ Fungal/Bacterial Miniprep Kit (ZYMO Research) also in accordance with the manufacturer's guidelines. The purified viral DNA was analyzed in an agarose (1% w/v) gel electrophoresis for quality assessment. All DNA retrieved (including bacterial) was analyzed in a nanodrop (ND-ONE-W, Thermo Scientific) for quantification and purity evaluation, through absorbance ratios.

Phenol/chloroform method

A volume of 25 µL of $MgCl_2$ (0.5 M) was added, separately, to 750 µL of each isolated phage suspension and mixed gently. Then, one µL of DNAse I (10 mg/mL) and 10 µL of RNAse A (10 mg/mL) were added. The resultant suspension was vortexed briefly and incubated for two hours (37°C). After that, the mixture was heated up at 70°C for 15 minutes to inactivate enzymes and, subsequently, 10 µL of Proteinase K (10 mg/mL) was added to the mixture. The temperature of the incubation was decreased to 56°C for the remaining incubation (overnight). After incubation, the suspension was set to cool at room temperature.

Once the suspension was cooled, 650 µL of phenol were added, vortexed, centrifuged (13000 × *g*, 10 minutes) and, after that, the top aqueous phase was extracted to a new tube. Equal parts of phenol and chloroform (350 µL each), respectively, were added to new tubes containing the aqueous phase, vortexed, centrifuged (13000 × *g*, 10 minutes) and, again, the top aqueous phase was collected into a new tube. Lastly, to the obtained aqueous phase, 600 µL of chloroform were added, vortexed, centrifuged (13000 × *g*, 10 minutes) and the top aqueous phase was extracted to a new tube once more.

A solution of 1:4 proportion of sodium acetate (3 M) and isopropanol was added to the top aqueous phase collected and gently mixed. Then, the mixture was putted on the freezer for 40 minutes and, afterwards, was centrifuged at 4°C (14000 × *g*, 15 minutes). The supernatant was discarded and the pellet was left to dry in room temperature for 2 hours. Once the pellet had dried,

it was added 35 µL of nuclease-free water and incubated at 56°C for 5 minutes. The pellet containing the DNA was stored at -20°C to further use [191,192].

All the viral and bacterial DNA extracted, according to the procedures above, were sequenced by the NovaSeq Illumina platform and assembled using a bioinformatics tool called Unicycler for further *in silico* analysis [193,194].

2.7. *In silico* analysis

The *in silico* analysis was done to evaluate the phages and bacteria genome and understand their characteristics and potential for usage in phage therapy against ETEC.

In silico analysis of phages

The structural analysis of the phages was done by homology-based predictions and manual correction of start sites and direction of the genome using Geneious software (version 9.1.4). myRAST [195] software was used to identify open reading frames (ORFs) and the identification of tRNA sites was done with tRNAscan-se [196] software [188,197,198].

Subsequently, software BlastN [199] was used to classify the taxonomy of each phage through the closest homolog of the *Caudoviricetes* class with a nucleotide-level similarity of over 90% across almost the entire genome [198,200].

Through BlastP [199] and HHpred [201] softwares, functional annotation of the phages was achieved. By homology and conserved domains analyses, the attribution of functions to the protein sequence integrating the phages ORFs was accomplished [191,202].

In silico analysis of bacteria

The prophage screening and functional annotation in the bacterial genomes was focused on identifying proteins related to APDS and compare the two strains genomes.

Regarding the bacteria analysis, the initial achievement was to export ETEC genomes from GenBank (27 in total) to do a prophage screening using Phaster [203] and Phastest [204] softwares. Afterwards, a whole well-characterized strain was selected (ETEC H10407, accession number: FN649414.1) to do a functional annotation, by protein homology and conserved domains evaluation, operating with BlastP and HHpred [205]. That same prophage screening and functional annotation was done to the bacteria strain (EC43) from which the DNA had been previously extracted.

Figure 12 illustrates a workflow with all information cited above, to synthetize and clarify the methods used during this project.

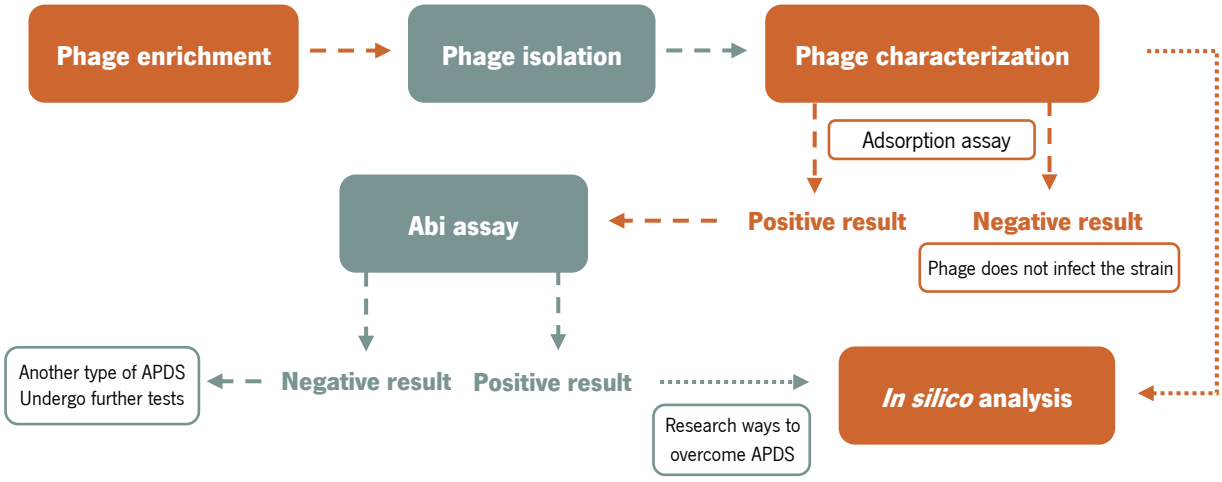

Figure 12 – Flux gram of the general methods executed during this study to achieve the main goal. Some of the processes within the primary methods were represented in order to enable the understanding of the relation between each stage of the methods that were used.

3. Results and Discussion

3.1. Isolated phages against ETEC

With the purpose of identifying phages with potential therapeutic applications against ETEC, sewage from wastewater was incubated with potential hosts for phage propagation. By engaging the methods mentioned in the previous chapter, we successfully isolated three distinct phages (**Figure 13**). These phages were named EcoSus34, EcoSus42 and EcoSus65, according to their respective bacterial host (EC34, EC42 and EC65, respectively). It is crucial to note that these phages were isolated with strains found in clinical cases in swine, which does not necessarily imply that they will infect ETEC strains in human clinical cases.

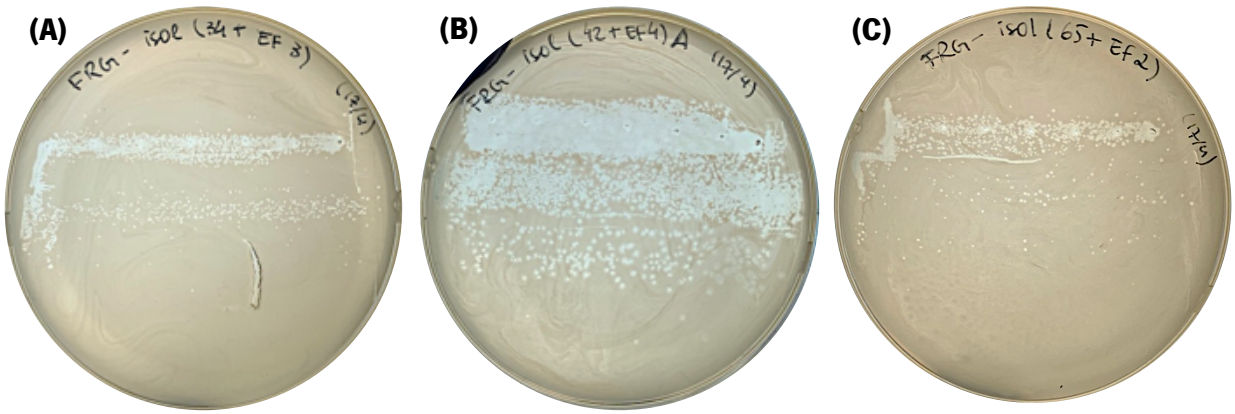

Figure 13 – Plates containing the isolates phage plaques after phage enrichment with potential hosts. **(A)** Plaques from phage EcoSus34 on the respective host (EC34) lawn; **(B)** Plaques from phage EcoSus42 on the respective host (EC42) lawn; **(C)** Plaques from phage EcoSus65 on the respective host (EC65) lawn.

After isolation and production of the phages, we were able to acquire the following concentrations (in PFU/mL): EcoSus34 at 3.6×10^{8}, EcoSus42 at 4.7×10^{10} and EcoSus65 at 7.5×10^{8}. Phage infection is facilitated when occurs the implementation of elevated phage titers at the targeted site, for instance, in the range of 10^{8} PFU/mL or higher, so, it is possible to state that the isolation and production of the discovered phages was successful, and their concentration is viable for bacterial infection [206].

3.2. Characterization of isolated phages

In this subchapter, all the outputs from the analysis of the phages' characteristics, including, phages morphology, host spectrum and efficiency of infection (EOP), will be referred.

Morphology of the phages

Based on the TEM microphotographs, we were able to include the isolated phages in the *Caudoviricetes* class, which remains the predominant class in phages, encompassing 96% of all known phages [207]. Within *Caudoviricetes* class, the phages represented two distinct morphotypes. EcoSus34 and EcoSus65 were identified as having linear contractile tails, recognized as myovirus, and EcoSus42 as having a shortened and thick non-contractile tail, also defined as podovirus (**Figure 14**) [208]. With regard to EcoSus34, phage heads were 105 nm × 114 nm (height × width) and the tails length was 152 nm (**Figure 14 - A**), while EcoSus42, phage heads were 73 nm × 67 nm and the tails length was 27 nm (**Figure 14 - C**). Lastly, for EcoSus65, phage head was 105 nm × 84 nm and the tail length was 116 nm (**Figure 14 - B**).

Myovirus-like and podovirus-like morphologies represent, combined, around 40% of the *Caudoviricetes* phages, with the first type being most frequently found than the latter [209]. Nevertheless, both phage morphologies were reported to have virulent effects on *E. coli* strains [182,210].

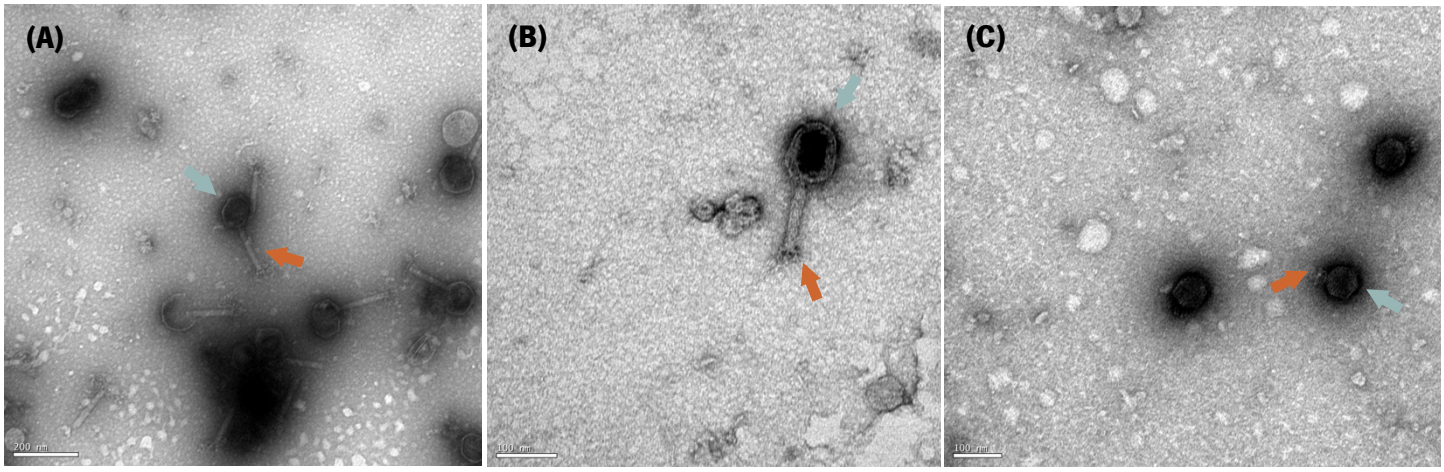

Figure 14 – Transmission electron micrograph of the isolated phages of this project. In **(A)** and **(B)** are represented phages EcoSus34 and EcoSus65, respectively, both categorized as myovirus-like. In **(C)** is represented phage EcoSus42 that was categorized as podovirus-like. The orange arrows indicate the tail and the blue arrows the head of the phages. Scale bar represents 200 nm on image **(A)** and 100 nm on images **(B)** and **(C)**.

Host spectrum and efficiency of infection

In order to evaluate the phages' lytic spectra, each phage was tested in 95 ETEC strains and the results of the spot test are represented in **Figure 15**.

Examining the outcome of this test, it is possible to affirm that phages' lytic spectra is substantially low, with 85.3% of the strains being resistant to phage activity due to absence of phage plaques. EcoSus34 and EcoSus42 showed clear phage plaques on their respective host and formed more or less turbid phage plaques in four other strains, those being distinct for both phages. As for the EcoSus65, it formed clear phage plaques in a strain other than its host and generated somewhat turbid phage plaques in two additional strains. The absence of phage plaques originating from two different phages in any strain highlights the considerable diversity and specificity among the isolated phages.

However, not all formation of plaques represents propagation and infection of phages (lysis from within), in some cases it can happen what is called lysis from without (LFW), referring to a bacterial lysis that occurs either through externally provided agents that damage the cell wall, or triggered by high multiplicity virion adsorption without phage production. Consequently, in these cases, the phage plaques only appear in high dilutions without progression [107,211]. To comprehend this phenomenon and to predict phages infectivity range, the EOP assay was performed and the results are represented in **Figure 15**. The outcome of the analysis is equal for phages EcoSus34 and EcoSus42 that, four out of the five infected strains represent LFW, reveling incapability of infecting other strains besides their host. On the other hand, in phage EcoSus65, LFW represented two out of four infected strains, since it was capable of infecting other strain, EC43, besides the respective host. However, EOP calculation of EC43 infection by EcoSus65 is equal to, approximately, 0.0027, which indicates that this phage has low efficiency.

Summing up the collected data, it is reasonable to state that all phages have a low host range, which it is not an ideal characteristic in phages intended for therapy [212]. Nevertheless, other studies have been reporting narrow lytic spectra in ETEC-infecting phage. Akindolire *et al.* [175] demonstrated that 60% of the phages studied only had the ability to lyse two or three *E. coli* O157:H7 strains and, being in accordance with our results, a substantial percentage (86.9%) of the environmental strains exhibited immunity against the tested phages. Kulikov *et al.* [189] stated that a N4 phage and N4-related phage, tested in 50 *E. coli* strains, only infected the respective host. This confirmed a typical trait of N4 phages by exhibiting a low lytic spectrum in *E. coli*, as has already been shown in other bacteria.

The same lytic spectrum was reported by Wei *et al.* [194], where the isolated phage against enterotoxigenic *E. coli* K88 did not inhibited any other bacteria, from the 22 tested, except its host. Upon these results, it is becoming crucial a better knowledge of the *E. coli* phages characteristics to achieve effective phage therapy of ETEC.

It was possible to realize that most results of EOP assay were reported as LFW or, in a single case, a low efficiency phage propagation. This outcome has been reported previously by Ferreira *et al.* [177] where four out of five phages, against ETEC strains, showed a higher LFW rate compared to the lysis from within rate, which also showed low EOP scores. A similarity has been observed between EcoSus65 and a previously characterized phage by Ferreira *et al.* [172] denominated FJ1, where both phages successful propagated in a unique *E. coli* strain besides their host, EC43. But, unlike EcoSus65, phage FJ1 had a high efficiency of propagation (EOP > 100%). Although, the low efficiency of EcoSus65 can be associated with bacterial abortive infection systems, that can interfere with phages' efficiency, without necessarily inhibiting virulent phages from producing progeny and infect other bacteria cells [213].

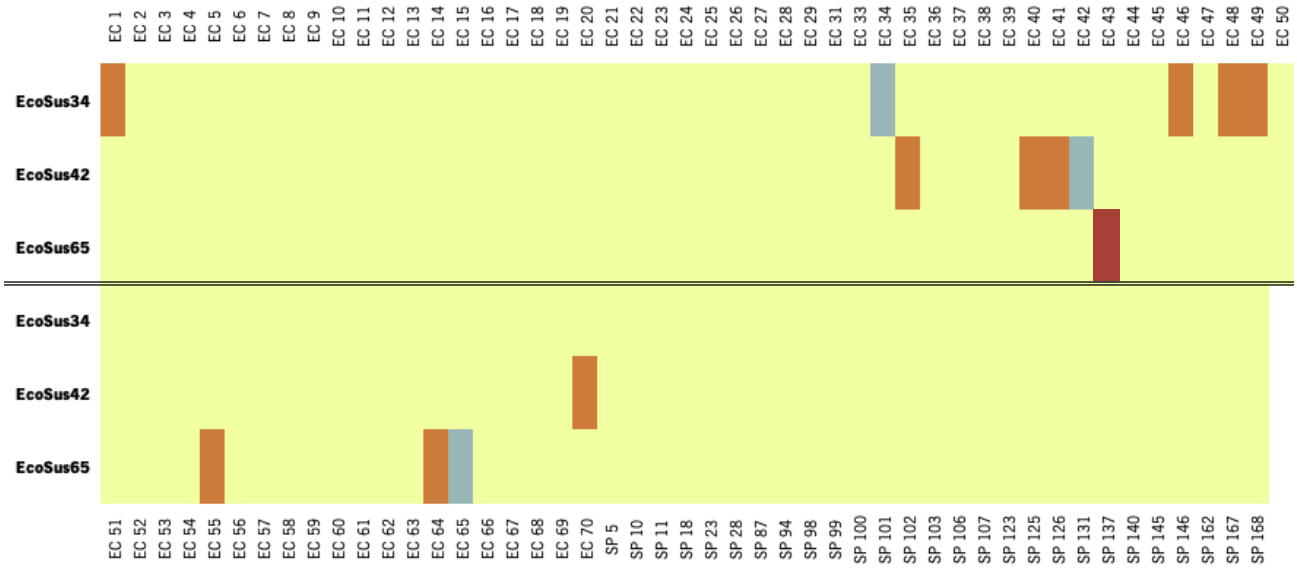

Figure 15 – Heat map representation of the results from the lytic spectra and EOP analysis of phages EcoSus34, EcoSus42 and EcoSus65 with 95 strains of ETEC. Orange indicates LFW, blue indicates the respective phages' hosts and red indicated lysis from within at low efficiency (0.2 ≥ EOP > 0.001).

3.3. Bacteriophage adsorption assay

The adsorption assays were conducted to investigate whether the events of lysis without phage propagation relied on phage infection with adsorption to cell wall receptors, which was subsequently inhibited by the immunity mechanisms of the bacteria or were directly affected by extracellularly supplied agents due to high MOI, independently of phage infection [211].

For the evaluation of this adsorption rate, one phage, EcoSus42, and two different *E. coli* strains, EC40 and EC70 (which lysis was scored as LFW), were selected. The results of the adsorption assay are present in **Figure 16**.

Analyzing the results, it was concluded that only the EC40 strain exhibited successful phage adsorption, aligning with the adsorption rate of the positive control (C1) (p - value > 0.05), proving that successful phage adsorption does not automatically result in successful phage infection [206]. In this case, high multiplicity virion adsorption could have occurred, that is when a significant number of phages adhere, enough cell-wall damage at vulnerable areas can cause bacterial lysis. Although, abortive infection should be taken into consideration as possible bacterial defense mechanism once phage adsorption occurs, causing the inhibition halo on the bacterial lawn [99,130].

The strain EC70 did not allow phage adsorption, which is perceptible by the increase values of phage titer, in accordance with the negative control (C2) values (p - value > 0.05), suggesting that it happens external liberation of phage-encoded lytic enzymes (lysins), leading to cell death without the need for adsorption. However, Gram-negative bacteria are known to resist the effects of externally applied lysins because of their protective outer membranes [214]. Nevertheless, a small subset of phage-encoded lysins demonstrates capability to eliminate Gram-negative bacteria, an example includes LysAB2 that, when administered externally, showed bacteriolytic activity against Gram-negative bacteria, including *E. coli* [215].

These results were confirmed by the proximity of adsorption constant (K_a) in both EC40 and C1, which were 3.7×10^{-10} mL^{-1} min^{-1} and 1.5×10^{-10} mL^{-1} min^{-1} respectively. Conversely, K_a value of EC70 equals zero, being the same value as C2, indicating ineffective binding of phages to the bacterial cell. Additionally, the K_a of EC40 and C1 revealed that phage EcoSus42 is slowly adsorbed, since K_a values of higher adsorption rates are around 10^{-9} mL^{-1} min^{-1} and K_a values of slower adsorption rates are around 10^{-10} mL^{-1} min^{-1} [216]. Also, the difference between K_a values may be due

to slightly alteration of bacterial concentration that disturbs the phage-bacteria ratio. Therefore, when the bacteria concentration is higher (causing a lower MOI) consequently, the K_a value is lower [217].

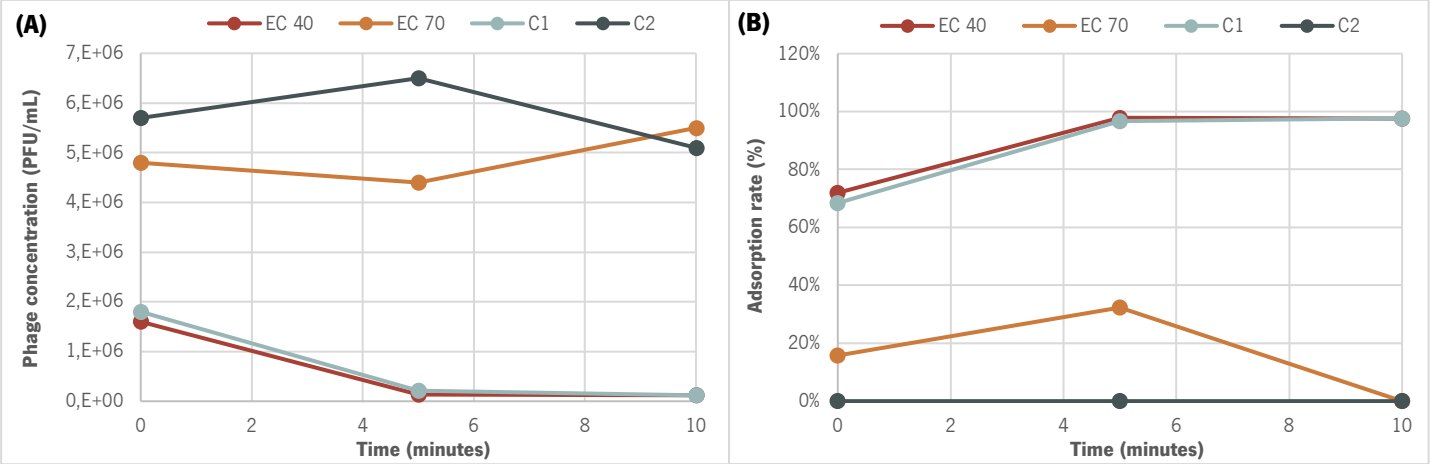

Figure 16 – Results of the adsorption assay on ETEC strains using phage EcoSus42. In **(A)** the value of phage concentration (PFU/mL) in function with time (minutes). In **(B)** is represented the adsorption rate (%) in function with time (minutes). All the data was obtained under the same condition and the adsorption rate of both positive and negative controls are also represented as C1 and C2, respectively.

3.4. Abortive infection as defense mechanism

The *E. coli* strain EC40, which exhibited phage adsorption without infection, underwent an Abi assay and, additionally, strain EC70, that showed no phage adsorption, and the strain EC42 (phage host) were used as terms of comparison. The variations of bacterial OD_{600} in function with time (minutes), after addition of the EcoSus42 phage, are illustrated in **Figure 17**. The respective OD_{600} reading translates the concentration of bacteria available, so, when the OD_{600} rises, the bacteria are under multiplication which increases their concentration. In contrast, when the OD_{600} decreases, the bacterial concentration also lowers, indicating cell death [218].

Regarding EC40, there was an initial decline observed in OD_{600} when the higher MOI (50) was used, followed by a subsequent and gradual increase. On the other hand, the low MOI (0.001) curve was similar ($p > 0.05$) to the bacterial growing patterns of the control curve (solution without phage) (**Figure 17 - A**). As for strain EC42, the high MOI curve had a significant initial bacterial propagation but, halfway through the assay, the OD_{600} decreased, reflecting cell death in a faster rate than the control curve. In contrast, the low MOI curve does not indicate bacterial propagation due to cell death in the initial stage of the assay, which maintains the low bacterial concentration

until the end (**Figure 17 - B**). Concerning strain EC70, the curves corresponding to high and low MOI settings were expected to align with the control curve, however, this alignment did not occur and is perceptible that the values are lower. Despite that, the low MOI is statistically similar to controls' bacterial growth rate (p - value > 0.05). Conversely, this similarity is not observed in high MOI values (p - value ≤ 0.05), however, there is a noticeable increase in the OD_{600} values, which was expected (**Figure 17 - C**).

Based on these results, it was possible to conclude that, EC40 genome probably encodes for proteins responsible for abortive infection mechanisms. That can be deduced because, in the event of abortive infection, the OD_{600} of the high MOI curve should decline over time, because, given the superior phage titer over bacteria, it is anticipated that all available bacteria adsorb the phages and be lysed after a certain duration, maintaining the OD_{600} at low values. Although, the increased of OD_{600} after initial decline could be possibly attributed to the lack of phage adsorption by some bacteria due to their low concentration, causing target bacteria to be widely dispersed [105]. Regarding the low MOI curve, it is expected that the OD_{600} should follow the pattern of the control curve, consistently increasing, given that the limited phages available will be adsorbed but phage propagation is inhibited by the defense system of the bacteria infected at an early stage and no phages are released for lysing neighbor cells, which was observed in the results.

Conversely, strain EC70 does not exhibit manifestation of Abi systems, which was expected since it does not adsorb the phage (Section 3.3.). This can be suggested since, without the occurrence of adsorption and possible Abi, the bacteria cells are not subject to phage infection and subsequent death, both in high MOI and in low MOI conditions, resulting in an OD_{600} increase of all the curves in line with the control curve, reflecting an unaltered bacteria growth.

As expected, the results confirmed that EC42 was susceptible to the phage and shows patterns of virulent infection. Reflecting phages' lytic life cycle, the OD_{600} in both high MOI and low MOI scenarios should firstly increase, reflecting phage adsorption, penetration and assembly taking place inside the infected bacteria and, subsequently, the phage is released causing bacterial lysis and drastic decrease of the OD_{600}. The only difference to note between high and low MOI curves is that, with a higher phage titer, bacterial death occurs earlier, resulting in a faster stabilization of the growth curve at values close to zero.

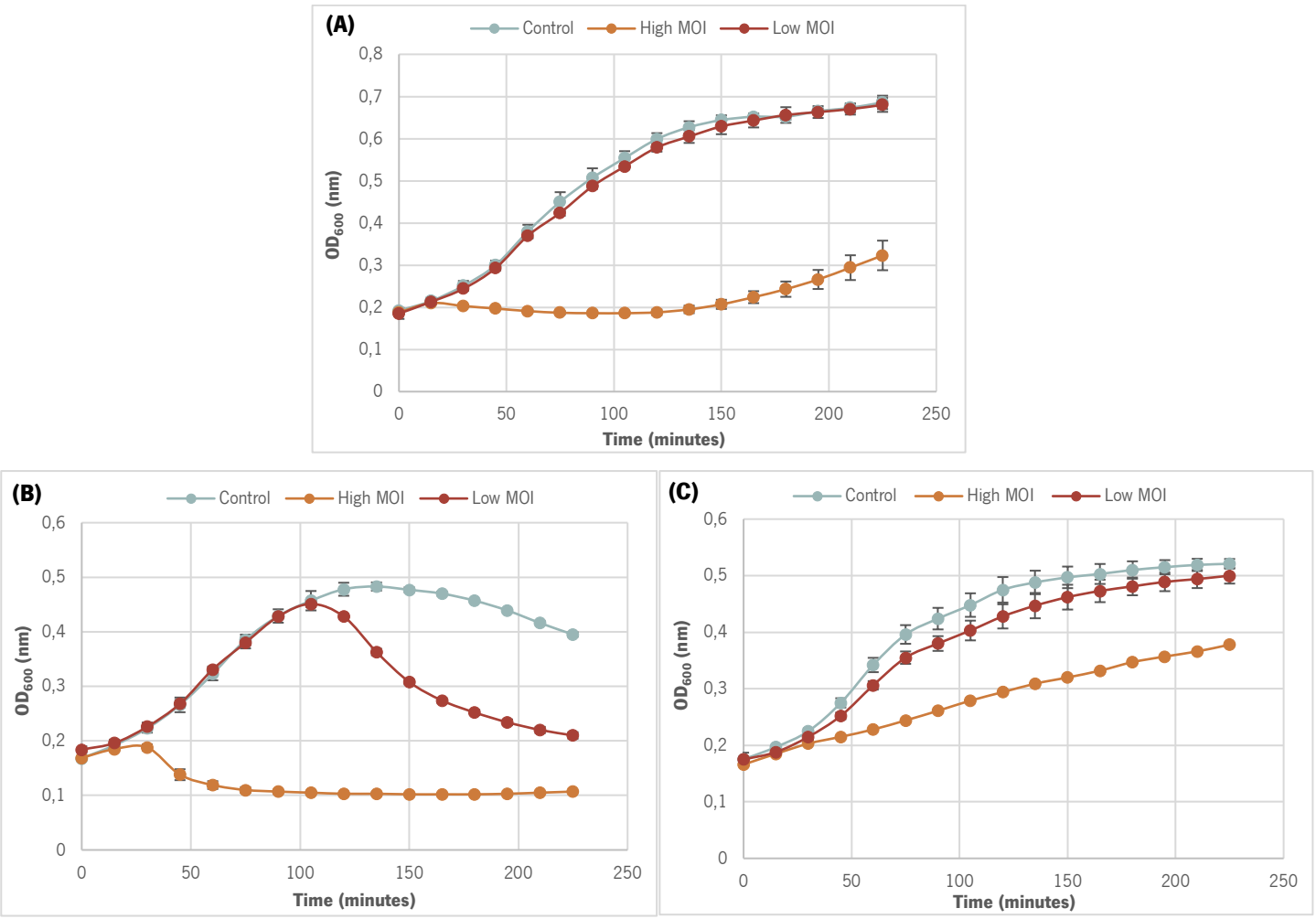

Figure 17 – Abi assay results of ETEC strains and phage EcoSus42. The value of OD_{600} is in function with time (minutes). In **(A)** is represented the strain undergoing analysis, EC40. In **(B)** and **(C)** are represented the strains used for comparison, *E. coli* EC42 (host) and *E. coli* EC70 (no adsorption) respectively. All the data was obtained under the same condition and error bars represent standard deviations of the samples taken in triplicate.

3.5. Genomic analysis

The genomic information of the three phages (EcoSus34, EcoSus42, EcoSus65) and a selected bacterial strain (*E. coli* EC43) were isolated and analyzed and, besides, the genomes of additional ETEC strains deposited on GenBank database were retrieved for subsequent *in silico* analysis. The extracted DNA underwent quantity and quality assessment and a structural annotation (genome length, identifying ORFs and percentage of GC) and, subsequently, it was done a functional annotation.

The functional annotation will ensure that the phages are suitable for treatment by discarding the ones that have virulence factors and established toxin-related proteins [219]. Furthermore, the

primary concern is the search for proteins that encoded APDS, this annotation will allow the prediction of those proteins inserted into the bacterial genome and possible counter-defense systems incorporated in phages' genome [135].

Genomic quantification and structural annotation

To extract phages' DNA the first approach was using the ZR viral DNA kit™ due to its simplicity and fastness [220]. However, this approach only demonstrated efficacy in the extraction of DNA from EcoSus42, which may be due to the higher concentration (4.7×10^{10} PFU/mL) in comparison with the titer of EcoSus34 (3.6×10^{8} PFU/mL) and EcoSus65 (7.5×10^{8} PFU/mL) since phage concentration during the DNA extraction procedure is crucial in determining the efficiency of the process [221]. So, an alternative for DNA extraction of EcoSus34 and EcoSus65, the phenol-chloroform method was used. This is a strong standard method that exhibits high yield, cost-effectiveness and high purity of the resulting DNA samples [222].

The DNA collected from all the phages went through a quantity and quality assessment. The phages' agarose gels are represented in **Figure 18**. In quantification by nanodrop, DNA purity is also evaluated though absorbance ratios, which was accomplished in all phages DNA, considering that pure DNA nucleic acids yield a 260/280 (nm) ratio proximal to 1.8, indicating that there was not protein contamination in any DNA sample [223]. The DNA concentrations and absorbance ratios of the phages are presented in **Table 1**.

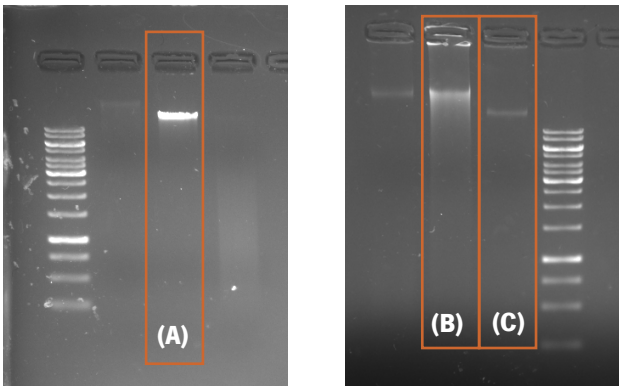

Figure 18 – Agarose (1% w/v) gels of phages' DNA after isolation. In **(A)** is represented the DNA of EcoSus42. In **(B)** is represented the DNA of EcoSus34. In **(C)** is represented the DNA of EcoSus65. At the ends of the gels is represented the DNA ladder (Thermo Scientific GeneRuler 1kb).

As expected for tailed phages, all the phages had a linear dsDNA genome, but with distinct lengths [224]. The obtained phages' genomic data is represented in **Table 2** and the respective sequences represented in **Figure 19**. In accordance with predictions, myovirus-like phages (EcoSus34 and EcoSus65) have bigger genome sequences than the podovirus-like phage (EcoSus42), confirming that, typically, myoviruses phages possess larger genomes compared to those of other phage families [225].

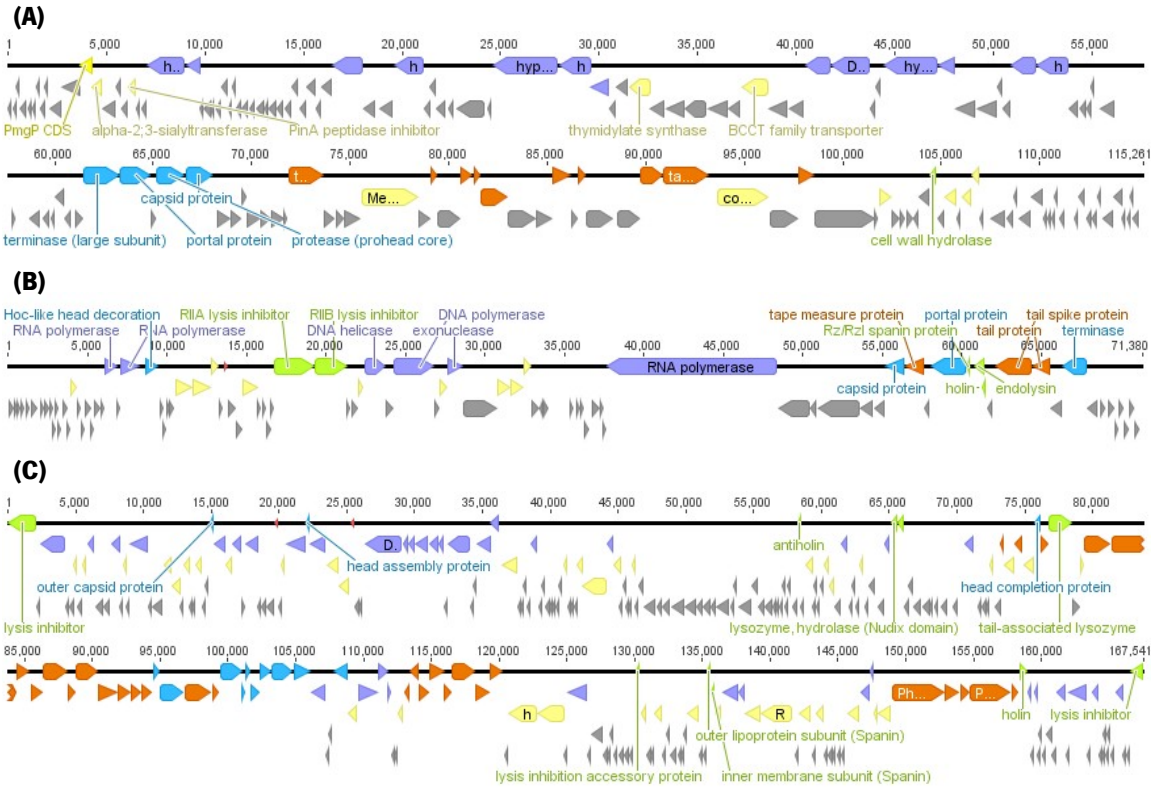

Figure 19 – Genome overview of the phages EcoSus34 **(A)**, EcoSus42 **(B)** and EcoSus65 **(C)**. Genome map with the predicted ORFs colored according with the respective predicted protein function. Gray represents hypothetical proteins, orange represents structural proteins, green represents lysis and lysis inhibition proteins, blue represents DNA packaging and morphogenesis proteins, purple represents DNA replication, transcription and repair proteins, red represents immunity related proteins and, lastly, yellow represents other type of function (regulatory proteins, metabolic enzymes, transmembrane proteins, etc.). The nucleotide position (in bp) is indicated above the genome sequence. The figure was generated using Geneious software (version 9.1.4).

The chosen bacterium for DNA extraction was *E. coli* strain EC43 based on its susceptibility to phage EcoSus65, being the only strain that exhibited viral infection traits induced by any of our isolated phages, apart from the hosts. The concentration and absorbance ratios of EC43 genome are present in **Table 1**.

Table 1 – Values of concentration in ng/µL for phages EcoSus34, EcoSus42, EcoSus65 and for ETEC bacterial strain EC43, with respective absorbance ratios at 260/280 nm, obtained through Nanodrop.

	Concentration (ng/µL)	Absorbance ratios (260/280 nm)
EcoSus34	373.8	1.73
EcoSus42	94.9	1.85
EcoSus65	166.1	1.55
ETEC EC43	72.2	1.90

Table 2 – Genomic data of phages EcoSus34, EcoSus42 and EcoSus65. Is represented each phages' sequence length in bp, the content of GC in percentage (%) and the number of ORFs.

	Sequence length (bp)	GC content (%)	Number of ORFs
EcoSus34	115 261	45	160
EcoSus42	71 380	34	92
EcoSus65	167 541	35	272

Taxonomical determination

The taxonomy of the isolated phages in study is the same as the closest homolog that represents a *Caudoviricetes* in the database [226]. According to that, the phages' taxonomies (classified as family, genus) are *Schitoviridae, Gamaleyavirus* (EcoSus42) and *Straboviridae, Tequatrovirus* (EcoSus65). Regarding EcoSus34, the lack of reliable and close homologs in BlastN, it was categorized as an unclassified *Caudoviricetes*. The discovery of genetically similar yet distinguishable phages within the same environment, indicates the likelihood of substantial molecular evolution taking place in their native surroundings [189].

Functional annotation of phages

For a good functional annotation, the identification of ORFs is especially important because it indicates the presence of a potential protein-coding region based on the presence of start and stop codons, however, does not necessarily mean that codes a functional protein [227]. The phages EcoSus34, EcoSus42 and EcoSus65 all had high coding density, implying that, in all genome length, 92.6%, 93.2% and 94.8% are protein-coding sequences, respectively.

The functional annotation of proteins in all phages was based on a comparison with a close homolog, from the *Caudoviricetes* class, considering a threshold E-value inferior to 1×10^{-5} and, when applicable, a coverage of the protein alignment greater than or equal to 80%, with similarity higher than 70% (**Tables C, D and E** - Annex 2) [228].

Starting with EcoSus34, 75% of the ORFs did not have an assigned function (described as hypothetical protein) and approximately 6% were typical structural proteins that constitute phages' virion particles (tail, assembly, baseplate, neck) (**Figure 19 - A**). EcoSus42 presented, around 68.5% of ORFs as hypothetical proteins and 4% were structural characteristic proteins of viruses (**Figure 19 - B**). Within the EcoSus65 ORFs, only 51% were functionally undefined, which was the better characterized phage among the three. This phage had roughly 14% of structure-related proteins (**Figure 19 - C**) [150,186]. Most of the anticipated proteins in all phages exhibited a small size (under 200 amino acids), EcoSus34 with 62% and, EcoSus42 and EcoSus65 both with 70%, while only 9%, 11% and 8% of proteins were formed with more than 500 amino acids, respectively.

Proteins related to the lysogenic life cycle, such as integrase, recombinase, repressors and excisionase, were not identified in the resulting annotations of EcoSus34 and EcoSus42, meaning that these phages might be exclusively virulent [188]. However, ORF 42 of EcoSus65 was categorized as a recombinase. Despite being potentially linked to phages' lysogeny, recombinases can have another purpose, such as, promoting phage packaging, aiding adaptation to diverse environments, or acting as a DNA repair pathway [229]. In this case, according to the homolog proteins found in *Escherichia* phages of the same genus as EcoSus65 (*Tequatrovirus*), this protein was described as recombinase A, playing a crucial role in the recombinational repair of damaged DNA in *E. coli* [230].

In all phages, the presence of ORFs coding endolysins or analogous (hydrolases) that can break down the bacterial cell wall to facilitate the release of mature progeny phages were clearly identified [231]. The endolysin activity is tightly regulated by holins (allows the passage of endolysins into the extracellular environment) and these types of proteins were found in EcoSus42 and EcoSus65 but not in EcoSus34 [100]. A *Punavirus* phage has been studied for holin-independent endolysin release, which still cause bacterial lysis, despite retarded. This fact may explain why EcoSus34 lysis of the host still occurs even with the absence of holin [232]. A similar occurrence takes place with Rz-like spanin, a lysis associated protein in Gram-negative bacteria (encodes an outer membrane lipoprotein), that is present in two phages, EcoSus42 and EcoSus65 and absent in EcoSus34. Although it is suggested that Rz-like spanin may not play a crucial role in the phage life cycle [233].

Another important protein for phages lytic cycle is terminase. Compose by a large and small subunit, it is responsible for the genome-packaging of phages into the capsid, so it is possible the transfer of viral information into the host genome [234]. This protein was categorized in all the three phages' genome, as well as proteins related with DNA transcription, including the RNA polymerase (allows the phages transcription of their own gene without depending on the host machinery) [191]. The three phages encoded DNA replication and repair associated proteins, such as, DNA helicase, DNA primase, DNA ligase, exo- and endo- nucleases, DNA topoisomerase and DNA clamp loader. Also, proteins linked to DNA packaging and morphogenesis, namely capsid proteins and portal proteins, were commonly described in all phages as well.

The absence of virulence-associated proteins, such as encoding toxins and other virulence factors that could impact eukaryotic cells, is a crucial criterion for eligibility of phages in therapeutic use [235]. EcoSus34, EcoSus42 and EcoSus65 did not exhibit any presence of toxins in ORFs functional annotation. Additionally, the absence of APDS within the phages' genome was confirmed using the PADLOC [236] and DefenseFinder [237] softwares. This analysis was done since the presence of CRISPR–Cas systems has been previously detected into virulent phages' genome (ICP1, a phage infecting *Vibrio cholerae*) [238]. The appearance of this defense systems in phages is a representation of unexpected genetic evolution in phage-host interaction, which benefit phage infection, since phages acquire new traits that overcome bacterial APDS [154].

Besides having their own CRISPR–Cas system to overcome bacterial APDS, phages can encode anti-CRISPR proteins (Acr or Aca protein families), proteins to prevent bacterial RM mechanisms (DarA, DarB and Ocr) or proteins that prevent Abi events though gene mutation or by interfering with the TA bacterial ratio (Dmd, a phage-encoded molecule that replaces antitoxins) [239–242]. Analyzing phages genome, only EcoSus65 showed the presence of a possible counter-defense for bacterial APDS, this being a Dmd molecule (ORF 39). The absence of both defense systems and counter defense proteins within EcoSus34 and EcoSus42 may be an explanation for the lower host range in comparison with EcoSus65.

However, EcoSus42 and EcoSus65 encoded a superinfection immunity protein (ORF 34 and ORF 45, respectively), that can be responsible for the prevention of a secondary infection of bacteria by an identical or closely related phage [243]. This immunity involves the production of repressor proteins that, despite inhibiting proximal phage infection, also hinders additional lysogenic infections [244]. For that reason, superinfection immunity proteins can be viewed as a possible counter defense system, since most bacterial APDS are encoded in mobile elements such as prophages (lysogenic

phage). Nevertheless, it cannot be a straightforward thought due to the constant competition between bacteria and phages that fuels the evolution of defense systems in both [245]. Additionally, EcoSus65 also encoded a periplasmic protein named Spackle (ORF 37), that has been described in T4 phage and associated with inhibition of activity a tail spike protein (gp5). Consequently, by preventing penetration by the tail tube of incoming phages, it confers immunity against secondary phage infection [246].

Besides the proteins present in phages' genome referred previously, a diverse array of biological proteins is also encoded by these phages. EcoSus34 included metabolic enzymes in its genome, that are involved in various biological processes (alpha-2,3-sialyltransferase, thymidylate synthase, protease, dCTP deaminase and dUTP diphosphatase). Additionally, it featured regulatory proteins (PinA peptidase inhibitor), transmembrane proteins that mediate the selective translocation (BCCT family transporter), polysaccharide lyases that target *E. coli* polysaccharides (colanic acid-degrading protein), and proteins associated with calcium signaling (RyR domain-containing protein). On the other hand, EcoSus42, besides also having metabolic enzymes (ATPase, metallopeptidase, deoxycytidine triphosphate deaminase, and thymidylate synthase), it encoded DNA-related proteins (DNA processing protein and DNA-binding protein) and lysis inhibitors (RIIA and RIIB proteins), this last category being in common with EcoSus65 proteins. However, in addition to that, EcoSus65 encoded a variety of metabolic enzymes (dextranase, ADP-ribosyltransferase, pyrophosphatase, glucosyltransferase, thymidylate synthase, reductase, peptidase, thymidine kinase, ribonuclease, lysozyme, dNMP kinase, RNA and DNA ligases, deoxycytidylate deaminase, polynucleotide kinase, alkaline phosphatase, ribonucleotide reductase and dihydrofolate reductase). It also featured tRNA-related proteins (cef modifier of suppressor tRNAs and tRNA ligase modifier), stress regulators (Mrh transcription modulator under heat shock), regulatory proteins (translation repressors, transcription modulator and inhibitor, glutaredoxins and transcription factors) and ADP-ribose metabolism represented by macrodomain proteins.

It should also be noted that, the narrow host spectrum demonstrated by the three phages, can be attributed to the remarkably specific tail spike protein, which facilitates the attachment of the phage to the polysaccharides on the bacterial cell surface [194]. All our phages encode this type of proteins, being necessary further specific research on the subject to acknowledge the limitation in the adsorption process of these phages.

For comparative genomic analysis, as mentioned in Section 3.2., the phages genomes were run though BLASTn in order to indicate the closest homologs. Regarding EcoSus42, it was clear the

homology between other phages from the same family and genus (*Schitoviridae, Gamaleyavirus*) studied priorly in *Escherichia* strains. Likewise, in phage EcoSus65, the homology among phages from corresponding family and genus (*Straboviridae, Tequatrovirus*) was achieved. Recognizing orthologous genes plays a crucial role in comparative genomic studies by facilitating the exploration of evolutionary connections among the genome of diverse species [247]. For that analysis, three of the closest homologs of both phages (with a coverage higher than 80% and an identity higher than 90%) and a representative of their genus (PhAPEC7 and T4) were selected to conduct a comparison to identify orthologous clusters, using OrthoVenn3 [247] software, employing an E-value equal to 1×10^{-5} and a default inflation value of 1.5 (**Figure 20**) [248,249].

The phage EcoSus42 underwent comparative evaluation with phages from *Gamaleyavirus* genus, that include phages E20, F22, PhAPEC7 and KKP 3715 (respective accession numbers on GenBank: OP745616.1; MN855733.1; NC_024790.1; OR067834.1). The analysis spotted a total of 81 clusters within EcoSus42 genome, in which 63 are common between all five phages from this analysis, giving a percentage of 77.8% of orthologous (**Figure 20 - A**). A comparative assessment was conducted as well on EcoSus65 with phages PhiZZ23, KAW1E185, JLBYU22 and T4 (respective accession numbers on GenBank: NC_054901.1; NC_054922.1; OK272484.1; NC_000866.4) from *Tequatrovirus* genus. Encoded in the EcoSus65 genome was identified a total of 265 clusters and, among these, 225 are shared between the five types of phages, a relatively high percentage of orthologous of 84.9%. (**Figure 20 - B**). This comparative analysis of phages EcoSus42 and EcoSus65 confirms the taxonomy previously given to the phages by the phylogenetic proximity amongst phages from the same genus and by the high percentage of orthologous proteins that, when equals or surpasses 40%, the phages are categorized within the same genus [250].

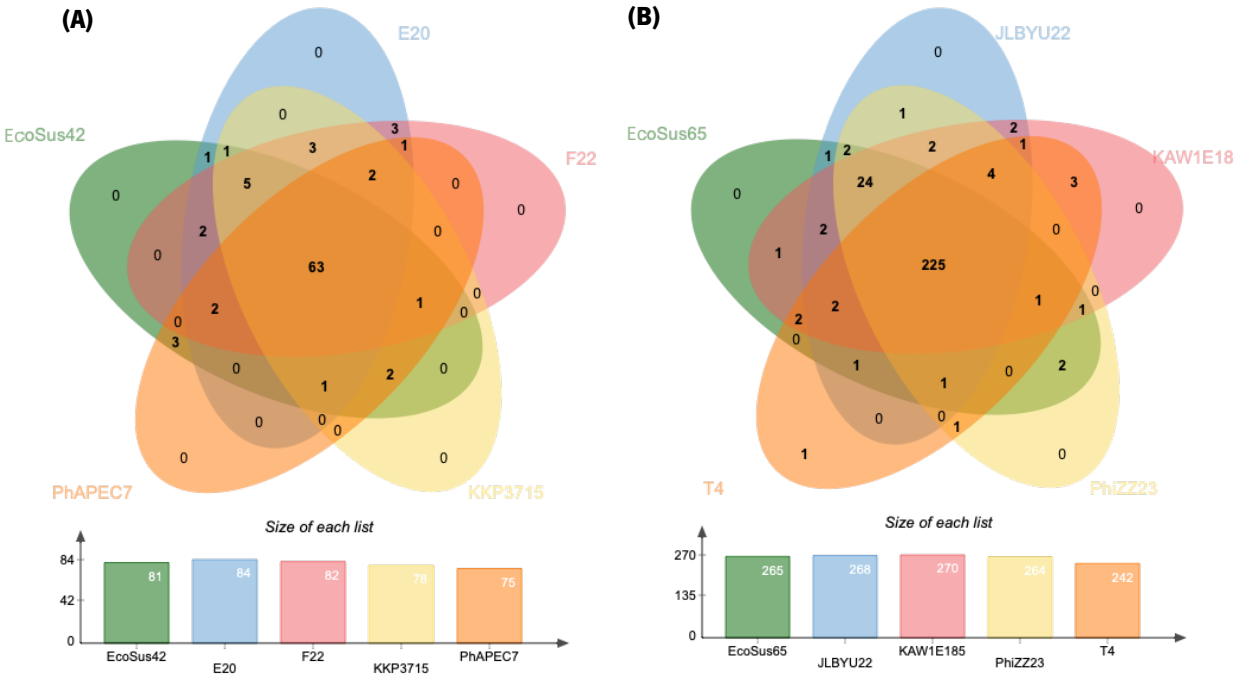

Figure 20 – Comparison of the gene products of EcoSus42 and EcoSus65 with the closest related phage genomes, available in the GenBank database, using OrthoVenn3 software. In **(A)** and **(B)** are represented the Venn diagrams of EcoSus42 and EcoSus65, respectively, illustrating the count of orthologous clusters that are mutually shared between their closest related phages.

As for EcoSus34, the situation varies since it is potentially a new type of phage that has not been described before. The BLASTn analysis was unclear, considering that this phage was only genetically similar to two *Escherichia* phages (H11 and IME392), which themselves were not extensively characterized. Therefore, EcoSus34 was subjected to a phylogenetical analysis, using Virus Classification and Tree Building Online Resource (VICTOR) [251] software, to predict its proximity to other well-defined phages and possibly anticipate its taxonomy also.

The analysis was conducted using 40 different phages from GenBank database, with a myovirus-like structure and *E. coli* as hosts [252–254]. These phages were selected in order to represent different genus (20) of *E. coli* phages, in accordance to ICTV (International Committee on Taxonomy of Viruses) and aligned with EcoSus34 known morphology (myovirus), namely: *Vequintavirus* (from *Vequintavirinae* subfamily); *Felixounavirus* and *Suspvirus* (from *Ounavirinae* subfamily); *Peduovirus* (from *Peduoviridae* family); *Tequatrovirus*, *Mosigvirus*, *Dhakavirus* and *Gaprivervirus* (from *Tevenvirinae* subfamily); *Muvirus*; *Kayfunavirus* (from *Studiervirinae* subfamily); *Phapecoctavirus* and *Justusliebigvirus* (from *Stephanstirmvirinae* subfamily); *Wifcevirus*; *Asteriusvirus*; *Krischvirus* and *Pseudotevenvirus* (from *Straboviridae* family); *Punavirus*; *Carltongylesvirus* (from *Cleopatravirinae* subfamily); *Taipeivirus* (from Cvivirinae subfamily) and *Goslarvirus*. The resulting phylogenetical tree from VICTOR software, with the recommended GBDP (Genome Blast Distance

Phylogeny) distance formula for amino acid-based analysis (d_6) is represented in **Figure 21** as well as the accession numbers of the 40 different phages used [251,255].

Examining the phylogenetical tree, it is possible to state that EcoSus34 shares a slight similarity with five other phages (P1, RCS47, Mu, Wphi and P2), representatives of *Punavirus*, *Muvirus* and *Peduovirus*. However, it is possible to understand its singularity since it does not share substantial resemblance with the types of phages analyzed, justifying the single branch separating EcoSus34 from the remaining phages. In account to that it is reasonable to propose the creation of a new genus including EcoSus34, still, it is needed a thorough analyses of this matter to define the proper taxonomy of this phage.

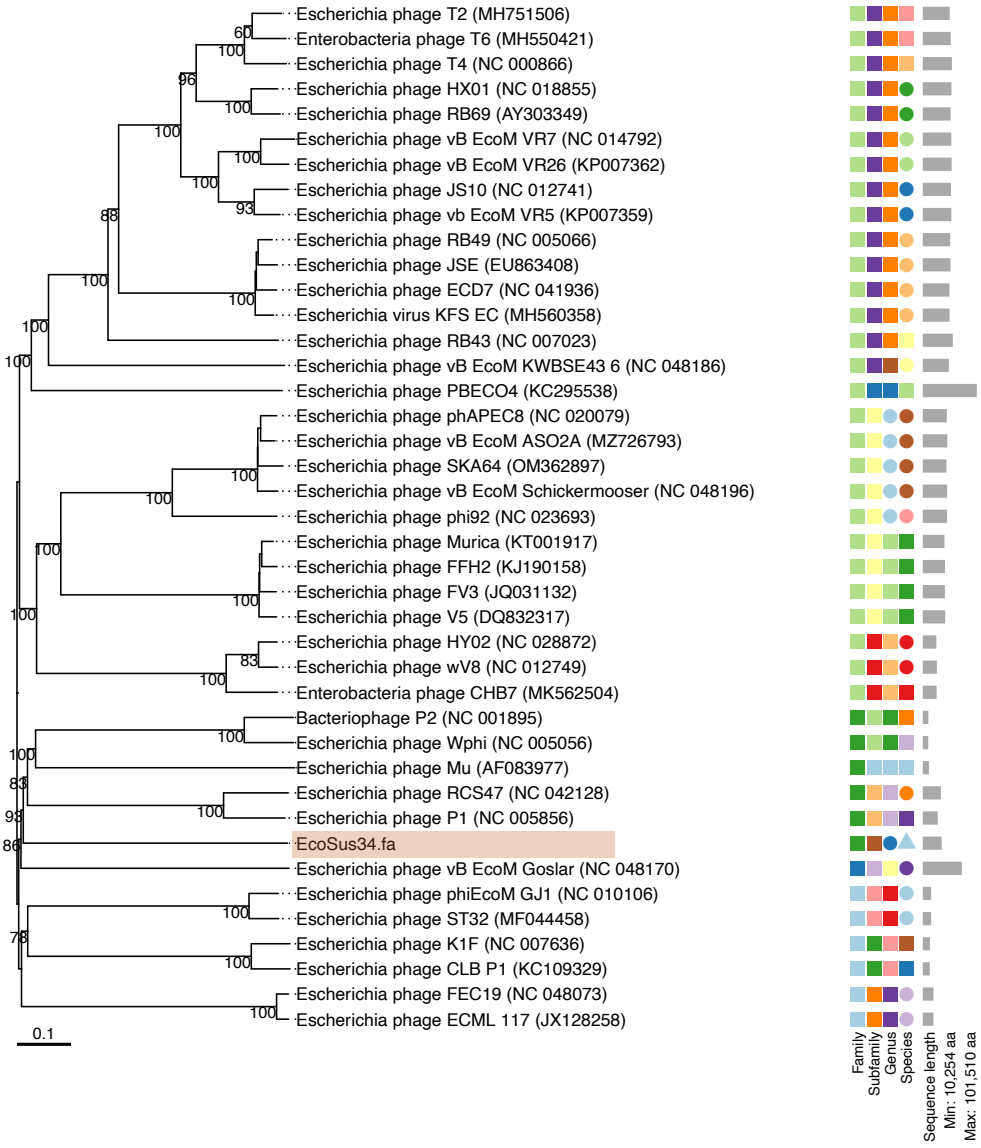

Figure 21 – Phylogenomic analysis of phage EcoSus34 (orange) with 40 *E. coli* phages from the GenBank database at the amino acid level using VICTOR software. The branch lengths are scaled in terms of the GBDP (Genome Blast Distance Phylogeny) distance formula d_6. Annotations on the right side of the tree include indicators of families, subfamilies, genera and species, according to ICTV, and the sequences lengths (aa).

Functional annotation of bacteria

The genomic information of the bacteria was examined to determine the presence of potential APDS. Firstly, 27 genomic sequences of ETEC strains from GenBank and, the genome of the sequenced *E. coli* strain EC43 were recovered for analysis. Regarding sequences lengths, the larger-sized strain was *E. coli* UMNK88, with a value of approximately 5.2 Mbp and the shorter sequence belonged to strain *E. coli* K-12 MG1655, with a size of approximately 4.6 Mbp. Considering all 28 sequences (EC43 included), the average size was 4.9 Mbp, which was in accordance with reported sizes of an *E. coli* sequence length (≈ 4.88 Mbp) [256]. In terms of percentage of GC, all strain had the same value of 51%.

The majority of APDS are closely associated with mobile components within the bacterial genome, such as prophages. Upon the integration of a prophage into its host genome, the repression of most phage genes becomes essential to cell viability, which prophages depend on. The activity of repressor proteins, expressed by prophages, results in resistance to superinfection by homologous phages, which is a distinctive feature of all lysogens [257,258].

The presence of prophages in each bacterial strain was first investigated. It was observed that, 100% of the strains tested had presence of prophages integrated into their genome, and in average, approximately 11.6 prophages (including intact, incomplete and questionable prophages) *per* genome were observed (**Table 3**). From the 28 bacterial strains analyzed, only two undergone functional annotation, one of those being the strain EC43, due to the outcome of that strain on our tests and previous reported analysis by Ferreira *et al.* [172]. The additional strain was *E. coli* ETEC H10407 (Accession numbers on GenBank: FN649414.1) because, besides the fact that it is a well-studied bacteria strain, also had the highest percentage of intact prophages, in comparison with the other strains, with no evidence of questionable prophages. Both strains had prophage occurrences and this data is represented in **Table F** on Annex 3. *E. coli* ETEC H10407 encoded 14 in total (10 intact, 4 incomplete and 0 questionable) and EC43 comprised 9 prophages (3 intact, 4 incomplete and 2 questionable).

Table 3 – Prophages' occurrence among 28 ETEC strains and respective frequency. In 27 strains, the genome was retrieved from GenBank and in 1 strain the genome was isolated and sequenced. The bacterial DNA was analyzed though Phaster and Phastest softwares.

	Number of prophages	Occurrences in strains	Frequency among strains (%)
Total	From 0 to 9	15	53.6
	From 10 to 20	13	46.4
Intact	From 0 to 4	17	60.7
	From 5 to 10	11	39.3
Incomplete	From 0 to 4	13	46.4
	From 5 to 10	15	53.6
Questionable	From 0 to 4	24	85.7
	From 5 to 10	4	14.3

Subsequently, to each prophages' incorporated coding sequences (CDS), that being a nucleotide sequence that is ultimately transformed into a protein through translation, it was done functional annotation, analyzing homology and conserved domains using BlastP and HHpred softwares, always considering a threshold E-value inferior to 1×10^{-5} [227]. This annotation mentioned is displayed in **Table G and H** on Annex 3 and the type and genomic data of each prophage, encoded by the two strains, are presented in **Table 4**.

Beforehand, it is important to mention that the functional annotation was done only on intact and incomplete prophages (total of 21 prophages between both strains) so, in subsequent discussion the questionable prophages were not considered. In the two strains, most proteins had a function associated, in ETEC H10407, from a total of 586 CDS, 58.7% were hypothetical proteins and, in EC43, from a total of 205 CDS, only 46.3% of the proteins were hypothetical as well (**Table G and H** - Annex 3).

Typical structural proteins encoded by phage-derived genome were found in prophages of both strains, however, only in intact prophages. Around 17.1% (100 proteins) and 21.9% (45 proteins) represented proteins associated with tail, assembly and baseplate in ETEC H10407 and EC43, respectively. As for proteins related with DNA packaging and morphogenesis, such as capsid, portal and terminase proteins, only around 6.8% (40 proteins) of ETEC H10407 and approximately 8.3% (17 proteins) of EC43 prophages encoded these types of proteins. Other categories of proteins distinctive to phages were present in all prophages of both strains, particularly related to cell lysis, DNA replication and modification, metabolic and DNA repair enzymes (reductase, nuclease, methyltransferase, exonuclease, peptidase, proteases, chitinase, ATPase, etc.), regulatory proteins, DNA transcription (polymerase), membrane proteins, stress regulators (cold shock protein), among

others. Also, there was evidence of proteins associated with lysogeny of phages (integrase, recombinase, repressors, and excisionase) in 20 of the total 21 prophages of both strains, which was expected since this type of phages follow a lysogenic life cycle. It has been reported that another type of protein, known as transposase, exhibits integrase-like action in certain phages. Curiously enough, that protein is encoded by almost all prophages in ETEC H10407 and by two prophages of EC43 [224].

Table 4 – Genomic data of the prophages encoded in ETEC H10407 and EC43 genomes. It is represented the type of prophages, its sequence length in Kbp, the content of GC in percentage (%) and the number of CDS.

	Type of prophage	Sequence length (Kbp)	GC content (%)	Number of CDS
ETEC H10407	Intact	57.3	48.9	62
	Intact	47.6	47.3	62
	Intact	49.3	52.2	67
	Intact	70.5	49.2	79
	Incomplete	7.4	52.1	11
	Intact	9.2	51.7	17
	Intact	13.2	51.6	13
	Intact	48.4	50.4	59
	Intact	39.6	50.6	57
	Incomplete	18.5	49.7	12
	Incomplete	5.5	47.3	8
	Intact	44.6	50.4	61
	Incomplete	8.2	54.2	13
	Intact	52.9	52.3	65
EC43	Intact	35.1	51.6	43
	Incomplete	39.6	48.8	32
	Intact	37.8	51.5	45
	Incomplete	5.5	47.5	9
	Questionable	24.6	52.7	31
	Incomplete	28.5	51.2	20
	Intact	58.4	51.5	48
	Questionable	29.7	48.7	14
	Incomplete	29.7	53.3	8

TA systems (Section 1.4.), can be contributors to APDS through an Abi mechanism. These systems are widespread in bacterial genomes among mobile genetic elements (prophages) [190]. Within the prophages' CDS in ETEC H10407, the functions of seven proteins were related to TA systems. Among the seven proteins mentioned, four of them were associated with CbeA/CbtA systems of TA, earlier described as a type IV (present in prophage 6 and 13). It was proven that a specific Abi system operates as unique Type IV TA systems and are prevalent across *E. coli* domains [259]. The remaining three proteins that possibly encoded a type of TA-associated function were not as well defined as CbeA/CbtA systems, since the homology-based prediction are inferior. Two of them characterized a RatA/RatB TA system, a ribosomal associated toxin and antitoxin (prophage 10). It has been demonstrated that RatA functions as a novel toxin in *E. coli*, effectively inhibiting the translation initiation stage. However, the role of RatB as the corresponding antitoxin was not experimentally confirmed, casting doubt on the notion that RatAB constitutes a true TA system [260,261]. The last protein mentioned encoded a toxin-antitoxin system and a toxic polypeptide from Hok (also denoted Gef) family (prophages 4). Hok/Gef is a Type I TA system that inhibits translation or initiates mRNA degradation, and it has been demonstrated to cause cell death in *E. coli* through cell lysis and creation of ghost cells, especially among pathogenic strains due to the higher likelihood of chromosomally encoded toxins [262,263]. Additionally, it was found a protein containing a YdaT domain described as toxin related (prophage 3). And, although *ydaT* genes were initially thought to form a toxin–antitoxin operon in E. coli, where YdaT was presumed to be the toxin, recent experimental evidence did not support this hypothesis. Instead, this protein encodes counterparts of the CII transcriptional regulators, which does not attribute APDS to the bacteria, rather contributes to a decision between lysis-lysogeny life cycles due to host physiological state [264–266].

Continuing the analysis on strain ETEC H10407, an anti-phage DSR was defined in prophage 14. DSR encode a sirtuin domain and provides bacterial defense against phage infections. In recent studies, it was established that, in *E. coli*, DSR proteins degrade NAD^+ during viral infection, leading to the decrease of this crucial molecule and ceasing the propagation of the phage (Section 1.4.). Also, it was shown that the inhibition of sirtuin enzymatic activity results in an enhanced production of virus progeny in infected human cells [165,267]. Evidence of another form of APDS, superinfection exclusion (Sie) system (Section 1.4.), was present in prophage 2 of ETEC H10407. The Sie systems do not have a global mode of action, instead, it can be accomplished through diverse mechanisms, with many involving modifications to the cell surface or other components of the cell envelope. It

has been escribed that, Sie proteins encoded by *E. coli* prophages inhibit the cell surface adsorption of infecting phages or can produce a small inner membrane protein that inhibits superinfection [258,268].

Concerning the functional annotation of EC43, it was notable the lesser quantity of proteins possibly related to APDS compared to the nine proteins encoded into ETEC H10407 prophages. In total, there were four proteins, within the prophages of EC43, that may cause a bacterial defensive reaction upon phage infection. Similar with ETEC H10407, a prophage of EC43 (prophage 4) also encoded the Type I TA system from the Hok/Gef family. As cited earlier, the overexpression of these types of proteins in Gram-negative bacteria proves toxic to cells. This toxicity results in the demise of cells from within, as these proteins disrupt a crucial function in the cell [269].

Another type of protein, related to APDS, was encoded by prophage 9 of EC43, namely a methyltransferase subunit (modification subunit) of a type I RM system (Section 1.4.). However, additionally to the modification subunit, this type of RM systems also needs a restriction (endonuclease) and specify subunit to form the resulting complex, both of which are not encoded by this prophage. Although it is possible a modification without the restriction subunit, it is also required a protein called S-adenosylmethionine (AdoMet) for cleavage, which is also not encoded by the prophage, limiting the functionality of the RM system in this cell [148,270]. The remaining two proteins, that can be related with a form of bacterial defense system, are both encoded as protein Kil (prophage 2 and 6). This protein is associated with a *kil* gene that, when expressed by induction of a lambda (λ) prophage, is responsible for host cell filamentation, loss of viability, and ultimate cell death. However, the reported function of the Kil protein is to inhibit the proper assembly of the essential bacterial protein (FtsZ) into the necessary structure for cell division and it is also suggested that this protein may delay bacterial lysis [271,272].

Aside from APDS related proteins, it was thought to be relevant to note that EC43 encodes an enzyme with antibiotic resistance activity. Chloramphenicol acetyltransferase (prophage 9) catalyzes the transfer of an acetyl group from acetyl CoA to the primary hydroxyl of chloramphenicol antibiotic, that, upon acetylation, is inactivated [273].

Furthermore, it is worth mentioning that the average number of prophages that the analyzed ETEC strains harbored was already considered a high number, elevating the probability of experiencing APDS [177]. Given that ETEC H10407 and EC43 had a total of 14 and 7 prophages (excluding questionable prophages), respectively, it is reasonable to expect a higher likelihood of APDS occurrence in ETEC H10407, a hypothesis that was confirmed. Additionally, the few counts of ADPS within EC43 prophages, coupled with an ineffective RM system, along with the presence of

counter defenses in EcoSus65 genome could justify the infection of the respective phage against this strain. These results can confirm the evolution between bacteria and phages to overcome barriers in their interaction, however, if prophage occurrences continue to raise, ETEC could be enhancing its resistance to phages, posing a challenge to the utilization of this virus for therapeutic purposes.

To achieve comparative analysis of the prophages, the same methodology of the phages' phylogenetic analysis was implied. Using VICTOR software, the genomic sequence of all prophages from both bacterial strains (ETEC H10407 and EC43), were compared with seven representatives from families of defined temperate phages of *E. coli*, such as, *Lambdavirus* (phages DN1 and HK629), *Muvirus* (phage MU), *Punavirus* (phages P1 and RCS47) and *Peduovirus* (phages P2 and Wphi) [197,274]. The resulting phylogenetic tree, with the recommended GBDP distance formula for nucleotide-based analysis (d_0) is represented in **Figure 22** as well as the accession numbers of the seven different phages used [251].

Based on this phylogenetic analysis it is possible to affirm that only around 52% of all 23 prophages shared similarities with the phages from the GenBank database. Prophages 1 and 8 of EC43 and prophage 4 of ETEC H10407 can possibly be related to a potential family including *Peduovirus, Muvirus* and *Punavirus* phages, however, only prophage 1 from EC43 may belong to the genus *Peduovirus*, while the other two phages are represented as individual genus within the same family. Regarding prophages 3, 5 and 7 of EC43 and prophages 1,2, 3, 9, 12 and 14 of ETEC H10407, they assumed similarities with the *Lambdavirus* phages and could all potentially belong to the same family as the phages previously mentioned. Although, prophages 3 and 7 of EC43 and prophages 2, 3 and 14 of ETEC H10407 were described as a part of *Lambdavirus* genus, while prophages 1 and 12 of ETEC H10407 and prophages 5 and 9 from EC43 and ETEC H10407, respectively, formed separate branches within this anticipated family, belonging to two different genera.

The remaining prophages of both strains, namely prophages 2, 4, 6 and 9 of EC43 and prophages 5, 6, 7, 8, 10, 11 and 13 of ETEC H10407 were represented into four distinct families. Firstly, prophage 9 of EC43 and prophage 10 of ETEC H10407 were described as unique prophages each, without any similarity with any of the phages represented in the phylogenetical tree, thus forming two separate families and genera. Regarding prophage 2 and 6 of EC43, both were associated into a different family. Curiously enough, these prophages encoded the same unique defense-associated protein (protein Kil) that is signature of the Lambda phages, however, prophages

2 and 6 did not share similarities with any phages from *Lambdavirus* genus nor belonged to the same potential family of Lambda phages. Lastly, prophages 4 of EC43 and prophages 5, 6, 7, 8, 11 and 13 of ETEC H10407 are branched together into a single "new" family and represented in four different genera. From ETEC H10407, prophages 6, 7 and 13 were forming one equal genus, despite being categorized into distinct species, and prophages 5 and 8 were associated with two divergent genera. Concerning prophages 4 (EC43) and 11 (ETEC H10407), in spite of belonging to separate bacterial strains, both were being associated with the same family, genus and specie, consequently meaning that they are equal prophages, which can be sustained by the analysis of the protein functional annotation, since both encode the same proteins including the same TA system-associated protein (**Table G and H** - Annex 3).

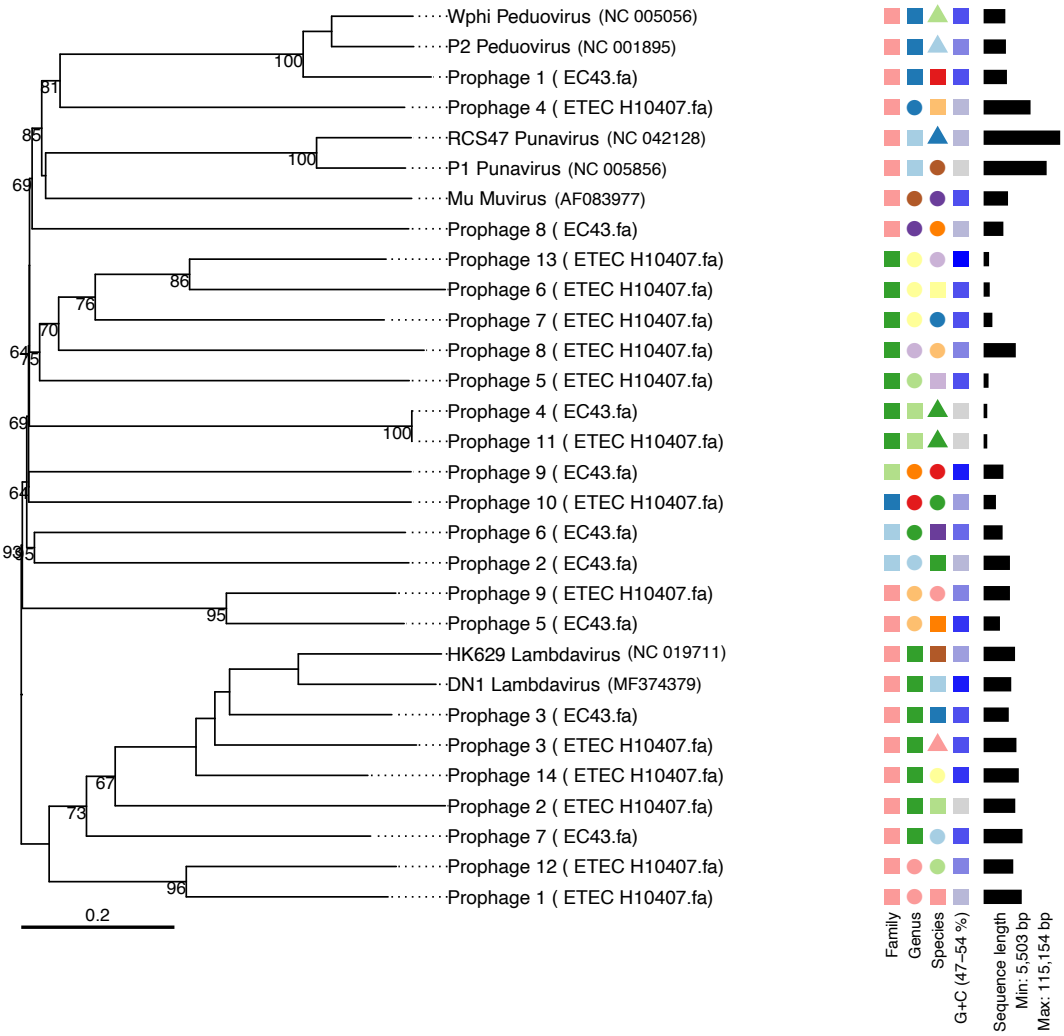

Figure 22 – Phylogenomic analysis of the prophages of strains ETEC H10407 and EC43 with seven *E. coli* temperate phages from the GenBank database, at the nucleotide level using VICTOR software. The branch lengths are scaled in terms of the GBDP (Genome Blast Distance Phylogeny) distance formula d_0. Annotations on the right side of the tree include indicators of families, subfamilies, genera and species, according to ICTV, the sequences lengths (bp) and the GC content in percentage (%).

4. Conclusion and future perspectives

Overall, the primary aim of this project was to isolate and characterize phages capable of infecting ETEC strains, comprehend the potential effect of the bacterial APDS in their lytic activity and identify counter-defense proteins whitin the phages genome. In order to achieve that goal, three novel phages targeting ETEC strains were isolated successfully and were subsequently characterized through lytic spectra, morphology, and genome analysis. Agreeing with the probability, all three phages, namely EcoSus34, EcoSus42 and EcoSus65 belong to *Caudoviricetes* (dsDNA tailed phages) since it is the predominant class.

The lytic spectra and efficiency of all isolated phages were substantially low, which is not the ideal scenario for phages intended for therapy. Nevertheless, based on previous works, this was predictable of ETEC-targeting phages. These results, and the evidence of phage adsorption prior to a possible abortive infection event, confirmed the importance of APDS occurrence in bacterial strains and the urgence of finding phages that can overcome defense mechanisms. Based on that, the analysis of both phages and bacterial genomes was determinant for a deeper understanding of the barriers that may hinder the potential use of these phages for therapeutic purposes.

The genomic analysis of the three phages indicated that possibly all exhibit typical features of exclusively virulent life cycles, a crucial measure for phages' implementation in therapy. However, phage EcoSus65 stood out as the only phage among the three in which were identified proteins capable of countering the bacterial defense mechanisms. Even so, it is worth noting that the presence of such proteins cannot be ruled out in the remaining phages, since less than half of their ORFs have an assigned function. Regardless, it is evident the prevalence of EcoSus65 over the other phages, being the most suitable candidate for potential therapeutic applications if its spectrum of lytic activity improves.

The *in silico* analysis of the ETEC strains suggested that the incidence of prophages within the bacterial genome can be correlated with the quantity of defense-associated proteins encoded by CDS. In other words, an increased count of prophages may correspond to a higher occurrence of APDS. The fact that, from a sample of 28 strains, all had between five and 20 prophages incorporated in their genome, raises a cause for concern in the future of ETEC infections control with phages. This hypothesis is supported by the analysis of prophages' functional annotation

analysis, in which both selected strains encode APDS-related proteins, such as TA systems, DSR, Sie systems and RM systems. Among these four identified defense mechanisms, it was evident that TA systems were the most prevalent, being encoded in seven out of a total of 21 prophages analyzed across both strains. The presence of such a mechanism, capable of causing cell death upon phage adsorption, was supported *in vitro*.

It is also important to note that, the presence of the same prophage specie in distinct bacterial strains, which encode the same defense system, and the manifestation of specific proteins from a particular phage genus in another type of temperate phage, may evidence the event of HGT between bacteria, suggesting that this phenomenon could be a possible cause for the increasing number of APDS occurrence.

Although this study has been completed and the main goals accomplished, there is still further analysis that can be done to improve the understanding on this matter and to take a step closer to the implementation of phage therapy for ETEC-related diseases.

Regarding the three isolated phages, further investigation into the suitability of EcoSus65 for phage therapy is needed, including *ex vivo* and *in vivo* studies along with physiological characterization, such as one-step growth curve and resistance to environmental factors (pH and temperature stability). Given that TA systems might be the most common type of APDS among ETEC strains, it is essential to conduct experiments to isolate phages that encode proteins capable of counteracting these systems, for instance, by inactivation of toxins/antitoxins, by prevention of system activation or by inhibiting the TA components expression. Employing genetic modification assays, such types of proteins can be further used to enhance infectivity rates and responses to bacterial APDS in other phages, including EcoSus65. Additionally, it is crucial continuing the *in silico* analysis done in this work on additional ETEC strains, especially in EC40 and EC70 due to the results in Abi assay, to gather more data on APDS, thereby strengthening the validity of these conclusions.

5. References

1. Rodrigues JF, Piazza RMF, Ferreira LCS, Martinez MB. Diarrheagenic Escherichia coli. *Brazilian J Microbiol*. 2016;47:3-30. doi:10.1016/j.bjm.2016.10.015
2. Basavaraju, M., & Gunashree BS. Escherichia coli: An Overview of Main Characteristics. 2023;IntechOpen. doi: 10.5772/intechopen.105508.
3. Evans DJ Jr. ED. *Escherichia Coli in Diarrheal Disease*. Medical Mi. (Baron S, ed.). Galveston (TX): University of Texas Medical Branch at Galveston; 1996.
4. Nagy Ã, Fekete Z. Enterotoxigenic Escherichia coli in veterinary medicine. *Int J Med Microbiol*. 2005;295:443-454. doi:10.1016/j.ijmm.2005.07.003
5. Bouckenooghe AR, Jiang ZD, Cabada FJ De, Ericsson CD, Dupont HL. Enterotoxigenic Escherichia coli as Cause of Diarrhea among Mexican Adults and US Travelers in Mexico. 1997:137-140.
6. DALTON4 CB, MINTZ ED, J.G.WELLS, BOPP5 CA, TAUXE R V. Outbreaks of enterotoxigenic Escherichia coli infection in American adults : a clinical and epidemiologic profile. *Epidemiol Infect*. 1999;123:9-16.
7. Iv JDA, Bagamian KH, Muhib F, et al. Articles Burden of enterotoxigenic Escherichia coli and shigella non-fatal diarrhoeal infections in 79 low-income and lower middle-income countries : a modelling analysis. *Lancet Glob Heal*. 7(3):e321-e330. doi:10.1016/S2214-109X(18)30483-2
8. Kim K, Song M, Liu Y, Ji P. Enterotoxigenic Escherichia coli infection of weaned pigs : Intestinal challenges and nutritional intervention to enhance disease resistance. 2022;(August):1-15. doi:10.3389/fimmu.2022.885253
9. Rhouma M, Fairbrother JM, Beaudry F, Letellier A. Post weaning diarrhea in pigs : risk factors and non - colistin - based control strategies. *Acta Vet Scand*. 2017:1-19. doi:10.1186/s13028-017-0299-7
10. García A, Fox JG. A One Health Perspective for Defining and Deciphering Escherichia coli Pathogenic Potential in Multiple Hosts. 2021;71(1):3-45. doi:10.30802/AALAS-CM-20-000054
11. Kaas RS, Friis C, Ussery DW, Aarestrup FM. Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse Escherichia coli genomes. *BMC Genomics*. 2012;13(577). doi: 10.1186/1471-2164-13-577.
12. Baeriswyl S, Bidet P, Touchon M, et al. Organised Genome Dynamics in the Escherichia coli Species Results in Highly Diverse Adaptive Paths. 2009;5(1). doi:10.1371/journal.pgen.1000344
13. Geurtsen J, Been M De, Weerdenburg E, Zomer A, Mcnally A. Genomics and pathotypes of the many faces of Escherichia coli. 2022;(June):1-30. doi:10.1093/femsre/fuac031

14. Nataro JP, Kaper JB. Diarrheagenic Escherichia coli. 1998;11(1):142-201. doi: 10.1128/CMR.11.1.142.
15. Gyles CL, Fairbrother JM. Escherichia coli in postweaning diarrhea in pigs : an update on bacterial types , pathogenesis , and prevention strategies. 2005;6(1):17-40. doi:10.1079/AHR2005105
16. Fleckenstein JM, Hardwidge PR, Munson GP, Rasko DA, Sommerfelt H, Steinsland H. Molecular mechanisms of enterotoxigenic Escherichia coli infection. *Microbes Infect*. 2010;12(2):89-98. doi:10.1016/j.micinf.2009.10.002
17. Kaper JB, Nataro JP, Mobley HLT. Pathogenic Escherichia coli. *Nat Rev / Microbiol*. 2004;2(February). doi:10.1038/nrmicro818
18. Croxen MA, Law RJ, Scholz R, Keeney KM, Wlodarska M, Finlay BB. Recent Advances in Understanding Enteric Pathogenic Escherichia coli. 2013;26(4):822-880. doi:10.1128/CMR.00022-13
19. Weiglmeier PR, Rösch P, Berkner H. Cure and Curse: E. coli Heat-Stable Enterotoxin and Its Receptor Guanylyl Cyclase C. *Toxins (Basel)*. 2010;2:2213-2229. doi:10.3390/toxins2092213
20. Nagy B, Fekete PZ, Nagy B, Zs P, Enterotoxigenic F. Enterotoxigenic Escherichia coli (ETEC) in farm animals. *Vet Res*. 1999;30(2-3):259-284.
21. Zhang Y. Enterotoxigenic Escherichia coli : intestinal pathogenesis mechanisms and colonization resistance by gut microbiota. 2022;14(1):1-22. doi:10.1080/19490976.2022.2055943
22. Butt S, Saleh M. Impact of the Escherichia coli Heat-Stable Enterotoxin b (STb) on Gut Health and Function. *Toxins 2020*. 2020;12(760). doi:10.3390/toxins12120760.
23. Qadri F, Svennerholm A, Faruque ASG, Sack RB. Enterotoxigenic Escherichia coli in Developing Countries : Epidemiology , Microbiology , Clinical Features , Treatment , and Prevention. 2005;18(3):465-483. doi:10.1128/CMR.18.3.465
24. Alam NH, Ashraf H. Treatment of Infectious Diarrhea in Children. *Pediatr Drugs*. 2003;5(3):151-165. doi: 10.2165/00128072-200305030-00002.
25. Luppi A. Swine enteric colibacillosis : diagnosis , therapy and antimicrobial resistance. *Porc Heal Manag*. 2017;3(16). doi:10.1186/s40813-017-0063-4
26. Seo H, Zhang W. Development of effective vaccines for enterotoxigenic Escherichia coli. *Lancet Infect Dis*. 2020;20:150-152. doi:10.1016/S1473-3099(19)30631-0
27. Zhang W, Sack DA. Progress and hurdles in the development of vaccines against enterotoxigenic Escherichia coli in humans. *Expert Rev Vaccines 11(6),*. 2012;11(6):677-694. doi: 10.1586/erv.12.37. PMID: 22873126.
28. Riddle MS, Chen WH, Kirkwood CD, Maclennan CA, Edward FH. Update on vaccines for enteric

pathogens. *Clin Microbiol Infect*. 2018;24(10):1039-1045. doi:10.1016/j.cmi.2018.06.023

29. Leach S, Lundgren A, Carlin N, Löfstrand M, Svennerholm A. Cross-reactivity and avidity of antibody responses induced in humans by the oral inactivated multivalent enterotoxigenic Escherichia coli (ETEC) vaccine ETVAX. *Vaccine*. 2017. doi:10.1016/j.vaccine.2017.06.006
30. Darsley MJ, Chakraborty S, Denearing B, et al. The Oral , Live Attenuated Enterotoxigenic Escherichia coli Vaccine ACE527 Reduces the Incidence and Severity of Diarrhea in a Human Challenge Model of Diarrheal Disease. 2012;19(12):1921-1931. doi:10.1128/CVI.00364-12
31. Seo H, Garcia C, Ruan X, Duan Q, Sack DA, Zhang W. Preclinical Characterization of Immunogenicity and Efficacy against Diarrhea from MecVax, a Multivalent Enterotoxigenic E. coli Vaccine Candidate. *Infect Immun*. 2021;89(7). https://doi.org/10.1128/ IAI.00106-21.
32. Sun Y, Woo S. Intestinal challenge with enterotoxigenic Escherichia coli in pigs , and nutritional intervention to prevent postweaning diarrhea. *Anim Nutr*. 2017;3(4):322-330. doi:10.1016/j.aninu.2017.10.001
33. Seo H, Duan Q, Zhang W. Vaccines against gastroenteritis , current progress and challenges. *Gut Microbes*. 2020;11(6):1486-1517. doi:10.1080/19490976.2020.1770666
34. Zegeye ED, Govasli ML, Sommerfelt H. Development of an enterotoxigenic Escherichia coli vaccine based on the heat-stable toxin. *Hum Vaccin Immunother*. 2019;15(6):1379-1388. doi:10.1080/21645515.2018.1496768
35. Talaat KR, Porter CK, Bourgeois AL, et al. Oral delivery of Hyperimmune bovine serum antibodies against CS6-expressing enterotoxigenic Escherichia coli as a prophylactic against diarrhea. *Gut Microbes*. 2020;12(1). doi:10.1080/19490976.2020.1732852
36. CO T, G L, H L, et al. Protection by milk immunoglobulin concentrate against oral challenge with enterotoxigenic Escherichia coli. *N Engl J Med*. 1988;12(318(19)):1240-1243. doi: 10.1056/NEJM198805123181904.
37. Rathe M, Müller K, Sangild PT, Husby S. Clinical applications of bovine colostrum therapy : a systematic review. *Nutr Rev*. 2014;72(4):237-254. doi:10.1111/nure.12089
38. Stelwagen K, Carpenter E, Haigh B, et al. Immune components of bovine colostrum and milk. *J Anim Sci*. 2009;87(1):3-9. doi:10.2527/jas.2008-1377
39. Stoppato M, Gaspar C, Regeimbal J, et al. Oral administration of an anti-CfaE secretory IgA antibody protects against Enterotoxigenic Escherichia coli diarrheal disease in a nonhuman primate model. *Vaccine*. 2020;(xxxx):1-7. doi:10.1016/j.vaccine.2020.01.064
40. Otto W, Najnigier B, Stelmasiak T, Robins- RM. Randomized control trials using a tablet formulation of

hyperimmune bovine colostrum to prevent diarrhea caused by enterotoxigenic Escherichia coli in volunteers. 2011;5521. doi:10.3109/00365521.2011.574726

41. Marquardt RR, Jin LZI, P JK, Fang L, Frohlich AA, Baidoo SK. Passive protective e ¡ ect of egg-yolk antibodies against enterotoxigenic Escherichia coli K88 + infection in neonatal and early-weaned piglets. 1999;23:283-288.
42. Baidoo SK, Nyachoti CM, Marquardt RR. Response of early-weaned pigs to spray-dried porcine or animal plasma-based diets supplemented with egg-yolk antibodies against enterotoxigenic Escherichia coli 1. 2018;(May):2895-2903.
43. Liu Y, Espinosa CD, Abelilla JJ, et al. Non-antibiotic feed additives in diets for pigs : A review. *Anim Nutr*. 2018. doi:10.1016/j.aninu.2018.01.007
44. Højberg O, Canibe N, Poulsen HD, Hedemann MS, Jensen BB. Influence of Dietary Zinc Oxide and Copper Sulfate on the Gastrointestinal Ecosystem in Newly Weaned Piglets. 2005;71(5):2267-2277. doi:10.1128/AEM.71.5.2267
45. Pejsak Z, Kaźmierczak P, Butkiewicz AF, Wojciechowski J. Alternatives to zinc oxide in pig production. *Pol J Vet Sci*. 2023;26(2):319-330. doi:10.24425/pjvs.2023.145033
46. Tsiloyiannis VK, Kyriakis SC, Vlemmas J, Sarris K. The effect of organic acids on the control of post-weaning oedema disease of piglets. *Res Vet Sci*. 2001;70:281-285. doi:10.1053/rvsc.2001.0475
47. Osellia MR, Inamorea AF, Rittia MSB, Osib PB, Swaldc IO, Engheria EM. Alternatives to in-feed antibiotics in pigs : Evaluation of probiotics , zinc or organic acids as protective agents for the intestinal mucosa . A comparison of in vitro and in vivo results. 2005;54:203-218. doi:10.1051/animres
48. Windisch W, Schedle K, Plitzner C, et al. Use of phytogenic products as feed additives for swine and poultry. *J Anim Sci 2008*. 2008;86:140-148. doi:10.2527/jas.2007-0459
49. Black RE, Levine MM, Clements M Lou, Cisneros L. Treatment of Experimentally Induced Enterotoxigenic Escherichia coli Diarrhea with Trimethoprim , Trimethoprim-Sulfamethoxazole , or Placebo. 1982;4(2):540-545.
50. Fern D, Aguilar MR, Rojo L. Folic Acid Antagonists : Antimicrobial and Immunomodulating Mechanisms and Applications. *Int J Mol Sci*. 2019;20(4996). doi:10.3390/ijms20204996.
51. Sack DA, Kaminsky DC, Sack RB, Itotia JN, Arthur RR, Kapikian AZ, Orskov F OI. Prophylactic doxycycline for travelers' diarrhea. Results of a prospective double-blind study of Peace Corps volunteers in Kenya. *N Engl J Med*. 1978;6(298):758-763. doi: 10.1056/NEJM197804062981402.
52. Wilson DN. Ribosome-targeting antibiotics and mechanisms of bacterial resistance. *Nat Publ Gr*. 2014;12(1):35-48. doi:10.1038/nrmicro3155

53. Heck JE, Staneck JL, Cohen M, et al. Prevention of Travelers ' Diarrhea : Ciprofloxacin versus TrimethopridSulfamethoxazole in Adult Volunteers Working in Latin America and the Caribbean. *J Travel Med*. 1994;1:36-142.

54. Pham TDM, Ziora ZM, Blaskovich MAT. Quinolone antibiotics. *Med Chem Commun*. 2019;10:1719-1739. doi:10.1039/c9md00120d

55. Cho H, Uehara T, Bernhardt TG. Beta-Lactam Antibiotics Induce a Lethal Malfunctioning of the Bacterial Cell Wall Synthesis Machinery. *Cell*. 2014;159(6):1300-1311. doi:10.1016/j.cell.2014.11.017

56. Macheboeuf P, Contreras-martel C, Job V, Dideberg O. Penicillin Binding Proteins : key players in bacterial cell cycle and drug resistance processes. *FEMSMicrobiol*. 2006;30:673-691. doi:10.1111/j.1574-6976.2006.00024.x

57. Rhouma M, Beaudry F, Thériault W, et al. In vivo therapeutic efficacy and pharmacokinetics of colistin sulfate in an experimental model of enterotoxigenic Escherichia coli infection in weaned pigs. *Vet Res*. 2016:1-11. doi:10.1186/s13567-016-0344-y

58. Li J, Nation RL, Milne RW, Turnidge JD, Coulthard K. Evaluation of colistin as an agent against multi-resistant Gram-negative bacteria. *Int J Antimicrob Agents*. 2005;25:11-25. doi:10.1016/j.ijantimicag.2004.10.001

59. Rhouma M, Beaudry F, Letellier A. Resistance to colistin : what is the fate for this antibiotic in pig production ? *Int J Antimicrob Agents*. 2016;48(2):119-126. doi:10.1016/j.ijantimicag.2016.04.008

60. Umair M, Hassan B, Farzana R, et al. Articles International manufacturing and trade in colistin , its implications in colistin resistance and One Health global policies : a microbiological , economic , and anthropological study. *The Lancet Microbe*. 2023;4(4):e264-e276. doi:10.1016/S2666-5247(22)00387-1

61. Biswas S, Brunel J, Dubus J, Rolain J. Colistin : an update on the antibiotic of the 21st century. *Expert Rev Anti Infect Ther*. 2012;10(8):917-934. doi:10.1586/eri.12.78

62. Olaitan AO, Dandachi I, Baron SA, Daoud Z, Morand S, Rolain J. Banning colistin in feed additives : a small step in the right direction. *Lancet Infect Dis*. 21(1):29-30. doi:10.1016/S1473-3099(20)30915-4

63. Jansen W, Hout J Van, Wiegel J, Iatridou D, Chantziaras I, Briyne N De. Colistin Use in European Livestock : Veterinary Field Data on Trends and Perspectives for Further Reduction. *Vet Sci*. 2022;9(650):1-14. https://doi.org/10.3390/ vetsci9110650.

64. Arancibia EM, Pitart C, Ruiz J, et al. Evolution of antimicrobial resistance in enteroaggregative Escherichia coli and enterotoxigenic Escherichia coli causing traveller ' s diarrhoea. *J Antimicrob*

Chemother. 2009;64:343-347. doi:10.1093/jac/dkp178

65. (AGISAR) WAG on IS of AR. Critically Important Antimicrobials for Human Medicine. *World Heal Organ*. 2007;6th.
66. Levy SB. Factors impacting on the problem of antibiotic resistance. *J Antimicrob Chemother*. 2002;49:25-30.
67. Martinez JL. The role of natural environments in the evolution of resistance traits in pathogenic bacteria. *Proc R Soc B*. 2009;276:2521-2530. doi:10.1098/rspb.2009.0320
68. Demain AL, Sanchez S. Microbial drug discovery : 80 years of progress. *J Antibiot (Tokyo)*. 2009;62:5-16. doi:10.1038/ja.2008.16
69. MARTINEZ JL, BAQUERO F. Mutation Frequencies and Antibiotic Resistance. *Antimicrob Agents Chemother*. 2000;44(7):1771-1777.
70. Vikesland PJ, Pruden A, Alvarez PJJ, et al. Toward a Comprehensive Strategy to Mitigate Dissemination of Environmental Sources of Antibiotic Resistance. *Environ Sci Technol*. 2017;51:13061-13069. doi:10.1021/acs.est.7b03623
71. Dodd MC. Potential impacts of disinfection processes on elimination and deactivation of antibiotic resistance genes during water and wastewater treatment. *J Environ Monit*. 2012;14:1754-1771. doi:10.1039/c2em00006g
72. Davies J. Inactivation of Antibiotics and the Dissemination of Resistance Genes. *Science (80-)*. 1994;264:375-382.
73. Normark BH, Normark S. Evolution and spread of antibiotic resistance. *J Intern Med*. 2002;252:91-106.
74. Davies J, Davies D. Origins and Evolution of Antibiotic Resistance. *Microbiol Mol Biol Rev*. 2010;74(3):417-433. doi:10.1128/MMBR.00016-10
75. Ben Y, Fu C, Hu M, Liu L, Hung M, Zheng C. Human health risk assessment of antibiotic resistance associated with antibiotic residues in the environment : A review. *Environ Res*. 2019;169(July 2018):483-493. doi:10.1016/j.envres.2018.11.040
76. Wang N, Guo X, Yan Z, et al. A Comprehensive Analysis on Spread and Distribution Characteristic of Antibiotic Resistance Genes in Livestock Farms of Southeastern China. *PLoS One*. 2016:1-14. doi:10.1371/journal.pone.0156889
77. Wei R, Ge F, Huang S, Chen M, Wang R. Occurrence of veterinary antibiotics in animal wastewater and surface water around farms in Jiangsu Province , China. *Chemosphere*. 2011;82(10):1408-1414. doi:10.1016/j.chemosphere.2010.11.067

78. Blaser MJ. Antibiotic use and its consequences for the normal microbiome. *MICROBIOME*. 2016;352(6285).
79. Ashbolt NJ, Amézquita A, Backhaus T, Borriello P, Brandt KK. Human Health Risk Assessment (HHRA) for Environmental Development and Transfer of Antibiotic Resistance. 2013;121(9):993-1001. http://dx.doi.org/10.1289/ehp.1206316.
80. Debarbieux L, Pirnay J, Verbeken G, et al. A bacteriophage journey at the European Medicines Agency. *FEMS Microbiol Lett*. 2016;363. doi:10.1093/femsle/fnv225
81. Schmerold I, Geijlswijk I Van, Gehring R. European regulations on the use of antibiotics in veterinary medicine. *Eur J Pharm Sci*. 2023;189(January):106473. doi:10.1016/j.ejps.2023.106473
82. Turner JL, Dritz SS, Minton JE. Alternatives to R Conventional Antimicrobials in Swine Diets. *Prof Anim Sci*. 2001;17:217-226. doi:10.15232/S1080-7446(15)31633-8
83. Vandenheuvel D, Lavigne R. Bacteriophage Therapy : Advances in Formulation Strategies and Human Clinical Trials. *Annu Rev Virol*. 2015;2:599-618. doi:10.1146/annurev-virology-100114-054915
84. Summers WC. Bacteriophage herapy. *Annu Rev Microbiol*. 2001;55(437). doi: 10.1146/annurev.micro.55.1.437.
85. Zhang J, Li Z, Cao Z, et al. Bacteriophages as antimicrobial agents against major pathogens in swine : a review. *J Anim Sci Biotechnol*. 2015:1-7. doi:10.1186/s40104-015-0039-7
86. Rohde C, Wittmann J, Kutter E. Bacteriophages: A Therapy Concept against Multi-Drug–Resistant Bacteria. *Surg Infect (Larchmt)*. 2020;19(8):737-744. doi:10.1089/sur.2018.184
87. Altamirano FLG, Barr JJ. Phage Therapy in the Postantibiotic Era. *Clin Microbiol Rev*. 2019;32(2):1-25. https://doi.org/10 .1128/CMR.00066-18.
88. Aslam S, Lampley E, Wooten D, et al. Lessons Learned From the First 10 Consecutive Cases of Intravenous Bacteriophage Therapy to Treat Multidrug- Resistant Bacterial Infections at a Single Center in the United States. *Open Forum Infect Dis*. 2020;7(9). doi:10.1093/ofid/ofaa389
89. Keen EC. A century of phage research: Bacteriophages and the shaping of modern biology. *Bioessays*. 2014;37:6-9. doi:10.1002/bies.201400152
90. Orlova E V. Bacteriophages and Their Structural Organisation. *InTech*. 2012. doi: 10.5772/34642.
91. Dion MB, Oechslin F, Moineau S. Phage diversity, genomics and phylogeny. *Nat Rev Microbiol*. 2020;18:125-138. doi:10.1038/s41579-019-0311-5
92. Lokareddy RK, Sankhala RS, Roy A, et al. Portal protein functions akin to a DNA-sensor that couples genome-packaging to icosahedral capsid maturation. *Nat Commun*. 2017:1-11. doi:10.1038/ncomms14310

93. Valle M, Agirrezabala X, Martı J, Gonza M, Valencia A. Structure of the Connector of Bacteriophage T7 at 8 A Resolution : Structural Homologies of a Basic Component of a DNA Translocating Machinery. 2005:895-902. doi:10.1016/j.jmb.2005.02.005

94. Toit A Du. The language of phages. *Nat Publ Gr*. 2017:2017. doi:10.1038/nrmicro.2017.8

95. Clokie MRJ, Millard AD, Letarov A V, Heaphy S. Phages in nature. *Landes Biosci*. 2011;7081(1):31-45. doi:10.4161/bact.1.1.14942

96. Rakhuba D V, Kolomiets EI, Dey ES, Novik GI. Bacteriophage Receptors , Mechanisms of Phage Adsorption and Penetration into Host Cell. *Polish J Microbiol*. 2010;59(3):145-155.

97. Mijbas M, Alomari M, Dec M. Bacteriophages as an Alternative Method for Control of Zoonotic and Foodborne Pathogens. *Viruses*. 2021;13(2348). https://doi.org/10.3390/v13122348.

98. Lin DM, Koskella B, Lin HC, Lin DM, Lin HC, Gastroenterology S. Phage therapy : An alternative to antibiotics in the age of multi-drug resistance. *World J Gastrointest Pharmacol Ther*. 2017;8(3):162-173. doi:10.4292/wjgpt.v8.i3.162

99. Tarahovsky YS, Ivanitsky GR, Institute AAK. Lysis of Escherichia coli cells induced by bacteriophage T4. *FEMS Microbiol Lett*. 1994;122:195-199.

100. Loessner MJ. Bacteriophage endolysins — current state of research and applications. *Curr Opin Microbiol*. 2005;8:480-487. doi:10.1016/j.mib.2005.06.002

101. Hanlon GW. Bacteriophages : an appraisal of their role in the treatment of bacterial infections. *Int J Antimicrob Agents*. 2007;30:118-128. doi:10.1016/j.ijantimicag.2007.04.006

102. García P, Rodríguez L, Rodríguez A, Martínez B. Food biopreservation : promising strategies using bacteriocins , bacteriophages and endolysins. *Trends Food Sci Technol*. 2010;21:373-382. doi:10.1016/j.tifs.2010.04.010

103. Little JW. *Lysogeny, Prophage Induction, and Lysogenic Conversion*. Their Role. (K MW, Friedman DI, Adhya SL, eds.). Washington, D.C.; 2005.

104. Howard-varona C, Hargreaves KR, Abedon ST, Sullivan MB. Lysogeny in nature : mechanisms , impact and ecology of temperate phages. *ISME J*. 2017:1-10. doi:10.1038/ismej.2017.16

105. Venturini C, Fabijan AP, Iredell J, Lubian AF, Barbirz S. Biological foundations of successful bacteriophage therapy. *EMBO Mol Med*. 2022;14:1-20. doi:10.15252/emmm.202012435

106. Knowles B, Silveira CB, Bailey BA, et al. Lytic to temperate switching of viral communities. *Nature*. 2016. doi:10.1038/nature17193

107. Suh GA, Lodise TP, Tamma PD, et al. Considerations for the Use of Phage Therapy in Clinical Practice. *Antimicrob Agents Chemother*. 2022;66(3):1-36.

108. Bru H, Canchaya C, Hardt W. Phages and the Evolution of Bacterial Pathogens : from Genomic Rearrangements to Lysogenic Conversion. *Microbiol Mol Biol Rev*. 2004;68(3):560-602. doi:10.1128/MMBR.68.3.560

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

110. Viertel TM, Ritter K, Horz H. Viruses versus bacteria—novel approaches to phage therapy as a tool against multidrug-resistant pathogens. *J Antimicrob Chemother*. 2014;69:2326-2336. doi:10.1093/jac/dku173

111. Hyman P. Phages for Phage Therapy : Isolation , Characterization , and Host Range Breadth. *Pharmaceuticals*. 2019;12(35). doi:10.3390/ph12010035

112. SMITH HW, HUGGINS MB. Effectiveness of Phages in Treating Experimental Escherichia coli Diarrhoea in Calves , Piglets and Lambs. *J Gen Microbiol*. 1983;129:2659-2675.

113. Bach SJ, Johnson RP, Stanford K, Mcallister TA. Bacteriophages reduce Escherichia coli O157 : H7 levels in experimentally inoculated sheep. *Can J Anim Sci*. 2009;89(2). https://doi.org/10.4141/CJAS08083.

114. Cha S Bin, Yoo AN, Lee WJ, Shin MK, Jung MH, Shin SW. Effect of Bacteriophage in Enterotoxigenic Escherichia coli (ETEC) Infected Pigs. *J Vet Med Sci*. 2012;74(8):1037-1039. doi:10.1292/jvms.11-0556

115. Alomari MMM, Dec M, Nowaczek A, et al. Therapeutic and Prophylactic E ff ect of the Experimental Bacteriophage Treatment to Control Diarrhea Caused by E. coli in Newborn Calves. 2021. doi:10.1021/acsinfecdis.1c00010

116. Imklin N, Sriprasong P, Phuttapatimok S. In vivo assessment of bacteriophages specific to multidrug resistant Escherichia coli on fecal bacterial counts and microbiome in nursery pigs. *Res Vet Sci*. 2022;151(February):138-148. doi:10.1016/j.rvsc.2022.07.012

117. Jamalludeen N, Johnson RP, Shewen PE, Gyles CL. Evaluation of bacteriophages for prevention and treatment of diarrhea due to experimental enterotoxigenic Escherichia coli O149 infection of pigs. 2009;136:135-141. doi:10.1016/j.vetmic.2008.10.021

118. Alam S, Mccallin S, Barretto C, et al. Oral T4-like phage cocktail application to healthy adult volunteers from Bangladesh. 2012;434:222-232. doi:10.1016/j.virol.2012.09.002

119. Febvre HP, Rao S, Gindin M, et al. PHAGE Study : Effects of Supplemental Bacteriophage Intake on Inflammation and Gut Microbiota in Healthy Adults. *Nutrients*. 2019;11(666). doi:10.3390/nu11030666

120. Bolocan AS, Callanan J, Forde A, Ross P, Hill C. Phage therapy targeting Escherichia coli — a story with no end ? *FEMS Microbiol Lett*. 2016;363. doi:10.1093/femsle/fnw256
121. Zalewska-Piatek B, Piatek R. Phage Therapy as a Novel Strategy in the Treatment of Urinary Tract Infections Caused by E . Coli. *Antibiotics*. 2020;9(304). doi:10.3390/antibiotics9060304.
122. Liu D, Belleghem JD Van, Vries CR De, et al. The Safety and Toxicity of Phage Therapy : A Review of Animal and Clinical Studies. *Viruses*. 2021;13(1268). https://doi.org/10.3390/v13071268.
123. Loc-carrillo C, Abedon ST. Pros and cons of phage therapy. *Bacteriophage*. 2011;1:111-114. doi:10.4161/bact.1.2.14590
124. Principi N, Silvestri E, Esposito S. Advantages and Limitations of Bacteriophages for the Treatment of Bacterial Infections. 2019;10(May):1-9. doi:10.3389/fphar.2019.00513
125. Adesanya O, Oduselu T, Akin-ajani O, Olubusuyi M. An exegesis of bacteriophage therapy : An emerging player in the fight against anti-microbial resistance. 2020;6(May):204-230. doi:10.3934/microbiol.2020014
126. Henein A. What are the limitations on the wider therapeutic use of phage? *Bacteriophage*. 2013;3(2). doi:10.4161/bact.24872
127. Yang Q, Le S, Zhu T, Wu N. Regulations of phage therapy across the world. *Front Microbiol*. 2023;14(1250848). doi:10.3389/fmicb.2023.1250848
128. Oliveira A, Dias C, Oliveira R, et al. Paving the way forward : Escherichia coli bacteriophages in a One Health approach. *Crit Rev Microbiol*. 2022;0(0):1-18. doi:10.1080/1040841X.2022.2161869
129. Hitchcock NM, Devequi D, Nunes G, et al. Current Clinical Landscape and Global Potential of Bacteriophage Therapy. *Viruses*. 2023;15(1020). https://doi.org/10.3390/v15041020.
130. Hyman P, Abedon ST. *Bacteriophage Host Range and Bacterial Resistance*. Vol 70. 1st ed. Elsevier Inc.; 2010. doi:10.1016/S0065-2164(10)70007-1
131. Kutateladze M, Adamia R. Bacteriophages as potential new therapeutics to replace or supplement antibiotics. *Trends Biotechnol*. 2010;28(12):591-595. doi:10.1016/j.tibtech.2010.08.001
132. Nilsson AS, Nilsson AS. Pharmacological limitations of phage therapy. *Ups J Med Sci*. 2019;124(4):218-227. doi:10.1080/03009734.2019.1688433
133. Lin J, Du F, Long M. Limitations of Phage Therapy and Corresponding Optimization Strategies : A Review. *Molecules*. 2022;27(1857). https://doi.org/10.3390/ molecules27061857.
134. Stanley SY, Maxwell KL. Phage-Encoded Anti-CRISPR Defenses. *Annu Rev ofGenetics*. 2018;52:445-464. https://doi.org/10.1146/annurev-genet-120417- 031321.
135. Millman A, Melamed S, Leavitt A, et al. An expanding arsenal of immune systems that protect bacteria

from phages. *Cell Host Microbe*. 2022;30(11):1556-1569. doi: 10.1016/j.chom.2022.09.017.

136. Bernheim A, Sorek R. The pan-immune system of bacteria: antiviral defence as a community resource. *Nat Rev Microbiol*. 2020;18(2):113-119. doi:10.1038/s41579-019-0278-2
137. Yuan X, Huang Z, Zhu Z, Zhang J, Wu Q. Recent advances in phage defense systems and potential overcoming strategies. *Biotechnol Adv*. 2023;65(December 2022):108152. doi:10.1016/j.biotechadv.2023.108152
138. Labrie SJ, Samson JE, Moineau S. Bacteriophage resistance mechanisms. *Nat Publ Gr*. 2010;8(5):317-327. doi:10.1038/nrmicro2315
139. Riede I, Eschbach M. Evidence that TraT interacts with OmpA of Escherichia coli. *Elsevier Sci Publ B*. 1986;205(2). doi: 10.1016/0014-5793(86)80905-x.
140. Pedruzzi I, Rosenbusch JP, Locher KP. Inactivation in vitro of the Escherichia coli outer membrane protein FhuA by a phage T5-encoded lipoprotein. 1998;168:119-125.
141. Scholl D, Adhya S, Merril C. Escherichia coli K1 ' s Capsule Is a Barrier to Bacteriophage T7. 2005;71(8):4872-4874. doi:10.1128/AEM.71.8.4872
142. Forde A, Fitzgerald GF. Molecular organization of exopolysaccharide (EPS) encoding genes on the lactococcal bacteriophage adsorption blocking plasmid , pCI658. 2003;49:130-142. doi:10.1016/S0147-619X(02)00156-7
143. DESTOUMIEUX-GARZÓ D, DUQUESNE S, PEDUZZI J, GOULARD C, DESMADRIL M. The iron – siderophore transporter FhuA is the receptor for the antimicrobial peptide microcin J25 : role of the microcin Val 11 – Pro 16 β -hairpin region in the recognition mechanism. *Biochem J*. 2005;876:869-876.
144. Seed KD. Battling Phages : How Bacteria Defend against Viral Attack. *PLOS Pathog*. 2015;11(6). doi:10.1371/journal.ppat.1004847
145. Lu M, Henning U. Superinfection exclusion by T-even-type coliphages. *Trends Microbiol*. 1994;2(4):4-6. doi: 10.1016/0966-842x(94)90601-7.
146. Dy RL, Richter C, Salmond GPC, Fineran PC. Remarkable Mechanisms in Microbes to Resist Phage Infections. 2014;(June):1-25. doi:10.1146/annurev-virology-031413-085500
147. Tock MR, Dryden DTF. The biology of restriction and anti-restriction. *Curr Opin Microbiol*. 2005;8:466-472. doi:10.1016/j.mib.2005.06.003
148. Yuan R, Hamilton DL. Type I and Type III Restriction-Modification Enzymes. *Springer Ser Mol Biol*. 1984. https://doi.org/10.1007/978-1-4613-8519-6_2.
149. Blumenthal RM, Cheng X. *Restriction-Modification Systems*. Vol 0.; 2002.

150. Burstein D, Sun CL, Brown CT, et al. Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. *Nat Commun*. 2016;7:1-8. doi:10.1038/ncomms10613

151. Nidhi S, Anand U, Oleksak P, et al. Novel CRISPR – Cas Systems : An Updated Review of the Current Achievements , Applications , and Future Research Perspectives. 2021:1-42.

152. Garneau JE, Fremaux C, Horvath P, Magada AH. The CRISPR / Cas bacterial immune system cleaves bacteriophage and plasmid DNA. 2010:10-15. doi:10.1038/nature09523

153. Barrangou R, Fremaux C, Deveau H, et al. CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science (80-).* 2007;315(March):1709-1713.

154. Samson JE, Magadán AH, Sabri M, Moineau S. Revenge of the phages : defeating bacterial defences. *Nat Publ Gr*. 2013;11(10):675-687. doi:10.1038/nrmicro3096

155. Lopatina A, Tal N, Sorek R. Abortive Infection: Bacterial Suicide as an Antiviral Immune Strategy. *Annu Rev ofVirology*. 2020;12(40). doi: 10.1146/annurev-virology-011620-040628.

156. Bouchard JD, Dion E, Moineau S. Characterization of the Two-Component Abortive Phage Infection Mechanism AbiT from Lactococcus lactis. 2002;184(22):6325-6332. doi:10.1128/JB.184.22.6325

157. Caflisch KM, Suh GA, Patel R. Biological challenges of phage therapy and proposed solutions: a literature review. *Expert Rev Anti Infect Ther*. 2019;0(0):1. doi:10.1080/14787210.2019.1694905

158. Harms A, Brodersen DE, Mitarai N, Gerdes K. Toxins , Targets , and Triggers : An Overview of Toxin-Antitoxin Biology. *Mol Cell*. 2018;70(5):768-784. doi:10.1016/j.molcel.2018.01.003

159. Goeders N, Melderen L Van. Toxin-Antitoxin Systems as Multilevel Interaction Systems. 2014:304-324. doi:10.3390/toxins6010304

160. Owen VS, Wenner N, Dulberger CL, et al. Prophages encode phage-defense systems with cognate self-immunity ll ll Prophages encode phage-defense systems with cognate self-immunity. *Cell Host Microbe*. 2021;29:1620-1633. doi:10.1016/j.chom.2021.09.002

161. Millman A, Bernheim A, Stokar-avihail A, Leavitt A. Bacterial Retrons Function In Anti-Phage Defense Bacterial Retrons Function In Anti-Phage Defense. *Cell*. 2020;183:1-11. doi:10.1016/j.cell.2020.09.065

162. Cohen D, Melamed S, Millman A, et al. Cyclic GMP – AMP signalling protects bacteria against viral infection. *Nature*. 2019;(June). doi:10.1038/s41586-019-1605-5

163. Patel DJ, Yu Y, Jia N. Review Bacterial origins of cyclic nucleotide-activated antiviral immune signaling. *Mol Cell*. 2022;82(24):4591-4610. doi:10.1016/j.molcel.2022.11.006

164. Chakravarti A, Patel DJ. Atypical bacterial Argonautes regulate antiphage defense. 2023;(August). doi:10.1038/s41422-023-00862-8

165. Garb J, Lopatina A, Bernheim A, et al. Multiple phage resistance systems inhibit infection via SIR2-dependent NAD+ depletion. *bioRxiv*. 2021:1-20. doi: https://doi.org/10.1101/2021.12.14.472415.

166. Ka D, Oh H, Park E, Kim J, Bae E. Structural and functional evidence of bacterial antiphage protection by Thoeris defense system via NAD+ degradation. *Nat Commun*. 2020;11(2816):1-8. doi:10.1038/s41467-020-16703-w

167. Meyer JR, Dobias DT, Weitz JS, Barrick JE, Ryan T, Lenski RE. Repeatability and Contingency in the Evolution of a Key Innovation in Phage Lambda. 2012;335(6067):428-432. doi:10.1126/science.1214449.Repeatability

168. Sutherland IW. Polysaccharide lyases. *FEMS Microbiol Rev 16*. 1995;16:323-347.

169. Glonti T, Chanishvili N, Taylor PW. Bacteriophage-derived enzyme that depolymerizes the alginic acid capsule associated with cystic fibrosis isolates of Pseudomonas aeruginosa. 2010;108:695-702. doi:10.1111/j.1365-2672.2009.04469.x

170. Baker JR, Dong S, Pritchard DG. The hyaluronan lyase of Streptococcus pyogenes bacteriophage H4489A. *Biochem J*. 2002;365:317-322.

171. Laanto E, Hoikkala V, Ravantti J, Sundberg L-R. Long-term genomic coevolution of host-parasite interaction in the natural environment. *Nat Commun*. 2017;8(111). doi:10.1038/s41467-017-00158-7

172. Ferreira A, Silva D, Almeida C, et al. Effect of phage vB _ EcoM _ FJ1 on the reduction of ETEC O9 : H9 infection in a neonatal pig cell line. *Vet Res*. 2023;54(26):1-13. doi:10.1186/s13567-023-01157-x

173. Buchanan RL, Klawitter LA. The effect of incubation temperature , initial pH , and sodium chloride on the growth kinetics of Escherichia coli 0157 : H7. *Food Microbiol*. 1992;9:185-196.

174. Twest R Van, Kropinski AM. Bacteriophage Enrichment from Water and Soil. *Methods Protoc Vol 1 Isol Charact Interact*. 2009;501:15-21. doi:10.1007/978-1-60327-164-6

175. Akindolire MA, Ateba CN. Isolation and characterisation of bacteriophages with lytic activity against virulent Escherichia coli O157 : H7 : potential bio-control agents. 2019;(January). doi:10.20944/preprints201901.0132.v1

176. Azeredo J, Sillankorva S, Pires DP. Pseudomonas Bacteriophage Isolation and Production. 2014;1149:23-32. doi:10.1007/978-1-4939-0473-0

177. Ferreira A, Almeida C, Silva D, García-meniño I, Mora A. Studies on the interaction of three lytic bacteriophages with a wide collection of Escherichia coli strains implicated in swine enteric colibacillosis. *bioRxiv*. 2022. doi: 10.1101/2022.07.04.498734.

178. Mazzocco A, Waddell TE, Lingohr E, Johnson RP. Enumeration of Bacteriophages Using the Small Drop

Plaque Assay System. *Bacteriophages Methods Protoc Vol 1 Isol Charact Interact*. 2009;501:81-85. doi:10.1007/978-1-60327-164-6

179. Gencay YE, Birk T, Camilla M, Sørensen H. Methods for Isolation , Purifi cation , and Propagation of Bacteriophages of Campylobacter jejuni. *Methods Mol Biol*. 2017;1512:19-28. doi:10.1007/978-1-4939-6536-6
180. Yu P, Mathieu J, Li M, Dai Z, Alvarez PJJ. Isolation of Polyvalent Bacteriophages by Sequential Multiple-Host Approaches. 2016;82(3):808-815. doi:10.1128/AEM.02382-15.Editor
181. Kutter E. Phage Host Range and Efficiency of Plating. *Bacteriophages Methods Protoc Vol 1 Isol Charact Interact*. 2009;501:141-149. doi:10.1007/978-1-60327-164-6
182. Vera-Mansilla J, Sánchez P, Silva-Valenzuela CA, Molina-Quiroz RC. Isolation and Characterization of Novel Lytic Phages Infecting Multidrug-Resistant Escherichia coli. *Microbiol Spectr*. 2022;10(1). doi: 10.1128/spectrum.01678-21.
183. Mirzaei MK, Nilsson AS. Isolation of Phages for Phage Therapy : A Comparison of Spot Tests and Efficiency of Plating Analyses for Determination of Host Range and Efficacy. 2015:1-13. doi:10.1371/journal.pone.0118557
184. Islam S, Zhou Y, Liang L, et al. Application of a Phage Cocktail for Control of Salmonella in Foods and Reducing Biofilms. *Viruses*. 2019;11(841):1-19. doi:10.3390/v11090841.
185. Ackermann H. *Bacteriophage Electron Microscopy*. Vol 82. 1st ed. Elsevier Inc.; 2012. doi:10.1016/B978-0-12-394621-8.00017-0
186. Dewanggana MN, Evangeline C, Ketty MD, Waturangi DE, Magdalena S. Isolation, characterization, molecular analysis and application of bacteriophage DW - EC to control Enterotoxigenic Escherichia coli on various foods. *Sci Rep*. 2022;(0123456789):1-10. doi:10.1038/s41598-021-04534-8
187. Golec P, Kamil D, Hejnowicz MS, et al. A reliable method for storage of tailed phages. 2011;84:486-489. doi:10.1016/j.mimet.2011.01.007
188. Necel A, Bloch S, Nejman-Faleńczyk B, et al. Characterization of a bacteriophage , vB _ Eco4M-7 , that effectively infects many Escherichia coli O157 strains. *Sci Rep*. 2020;10(3743):1-14. doi:10.1038/s41598-020-60568-4
189. Kulikov E, Kropinski AM, Golomidova A, et al. Isolation and characterization of a novel indigenous intestinal N4-related coliphage vB _ EcoP _ G7C. *Virology*. 2012;426(2):93-99. doi:10.1016/j.virol.2012.01.027
190. Vassallo CN, Doering CR, Littlehale ML, Teodoro GIC, Laub MT. A functional selection reveals previously undetected anti-phage defence systems in the E. coli pangenome. 2022;7(October).

doi:10.1038/s41564-022-01219-4

191. Melo LDR, Sillankorva S, Ackermann H-W, Kropinski AM, Azeredo J, Cerca N. Isolation and characterization of a new Staphylococcus epidermidis broad-spectrum bacteriophage. *J Gen Virol*. 2014;95(Pt2):506-515. doi:10.1099/vir.0.060590-0

192. Manohar P, Tamhankar AJ, Lundborg CS, Id NR. Isolation , characterization and in vivo efficacy of Escherichia phage myPSH1131. 2018:1-17.

193. Mentzer A Von, Blackwell GA, Pickard D, et al. Long - read - sequenced reference genomes of the seven major lineages of enterotoxigenic Escherichia coli (ETEC) circulating in modern time. *Sci Rep*. 2021:1-16. doi:10.1038/s41598-021-88316-2

194. Wei B, Cong C, Zheng L, Chen L, Yan X. Isolation , characterization and whole genome analysis of the novel genus Lederbergvirus , phage vB _ EcoP _ E212 infecting enterotoxigenic Escherichia coli K88. *Virus Res*. 2023;331(December 2022):199125. doi:10.1016/j.virusres.2023.199125

195. Aziz RK, Bartels D, Best AA, et al. The RAST Server : Rapid Annotations using Subsystems Technology. *BMC Genomics*. 2008;9(75). doi:10.1186/1471-2164-9-75

196. Schattner P, Brooks AN, Lowe TM. The tRNAscan-SE , snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res*. 2005;33:686-689. doi:10.1093/nar/gki366

197. Maffei E, Shaidullina A, Burkolter M, et al. Systematic exploration of Escherichia coli phage – host interactions with the BASEL phage collection. *PLoS Biol*. 2021;19(11):1-52. doi:10.1371/journal.pbio.3001424

198. Martinez-vaz BM, Mickelson MM. In silico Phage Hunting : Bioinformatics Exercises to Identify and Explore Bacteriophage Genomes. 2020;11(September):1-7. doi:10.3389/fmicb.2020.577634

199. Altschup SF, Gish W, Pennsylvania T, Park U. Basic Local Alignment Search Tool 2Department of Computer Science. 1990:403-410.

200. Turner D, Adriaenssens EM. Phage Annotation Guide : Guidelines for Assembly and High-Quality Annotation. 2022;2(4):170-182. doi:10.1089/phage.2021.0013

201. Soding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res*. 2005;33:244-248. doi:10.1093/nar/gki408

202. Bao H, Shahin K, Zhang Q, et al. Morphologic and genomic characterization of a broad host range Salmonella enterica serovar Pullorum lytic phage vB _ SPuM _ SP116. *Microb Pthogenes*. 2019;136(50):103659. doi:10.1016/j.micpath.2019.103659

203. Arndt D, Grant JR, Marcu A, et al. PHASTER : a better , faster version of the PHAST phage search tool. 2016;44(May):16-21. doi:10.1093/nar/gkw387

204. Arndt D, Marcu A, Liang Y, Wishart DS. PHAST , PHASTER and PHASTEST : Tools for finding prophage in bacterial genomes. 2019;20(July 2017):1560-1567. doi:10.1093/bib/bbx121

205. Wang M, Zeng Z, Jiang F, et al. Role of enterotoxigenic Escherichia coli (ETEC) prophage in spreading antibiotic resistance in a porcine-derived environment. *Env Microbiol*. 2020;22(12):4974-4984. doi:10.1111/1462-2920.15084

206. Abedon ST. Phage therapy dosing : The problem (s) with multiplicity of infection (MOI) Phage therapy dosing : The problem (s) with multiplicity of infection (MOI). *Bacteriophage*. 2016;6(3):1-6. doi:10.1080/21597081.2016.1220348

207. Zinke M, Schröder GF, Lange A. Major tail proteins of bacteriophages of the order Caudovirales. *J Biol Chem*. 2022;298(1):101472. doi:10.1016/j.jbc.2021.101472

208. Fokine A, Rossmann MG. Molecular architecture of tailed double-stranded DNA phages. 2014;4:1-22.

209. Prangishvili HAD. Prokaryote viruses studied by electron microscopy. 2012:1843-1849. doi:10.1007/s00705-012-1383-y

210. Maal KB, Delfan AS, Salmanizadeh S. Isolation and Identification of Two Novel Escherichia coli Bacteriophages and Their Application in Wastewater Treatment and Coliform ' s Phage Therapy. 2015;8(3). doi:10.5812/jjm.14945

211. Abedon ST. Lysis from without. 2011;1(1):46-49. doi:10.4161/bact.1.1.13980

212. Hagens S, Loessner MJ. Bacteriophage for Biocontrol of Foodborne Pathogens : Calculations and Considerations. 2010:58-68.

213. Abedon ST. Detection of Bacteriophages: Phage Plaques. *n Harper, DR, Abedon, ST, Burrowes, BH, McConville, ML Bacteriophages Springer, Cham*. 2021. https://doi.org/10.1007/978-3-319-41986-2_16.

214. Shah S, Das R, Chavan B, Urmi Bajpai, Hanif S, Ahmed S. Beyond antibiotics : phage-encoded lysins against Gram-negative pathogens. *Front Microbiol*. 2023;2:1-16. doi:10.3389/fmicb.2023.1170418

215. Lai M, Lin N, Hu A. Antibacterial activity of Acinetobacter baumannii phage 6 AB2 endolysin (LysAB2) against both Gram-positive and Gram-negative bacteria. 2011:529-539. doi:10.1007/s00253-011-3104-y

216. Dennehy JJ, T.Abedon S. Adsorption: Phage Acquisition of Bacteria. *Harper, DR, Abedon, ST, Burrowes, BH, McConville, ML Bacteriophages Springer, Cham*. 2021. https://doi.org/10.1007/978-3-319-41986-2_2.

217. Abedon ST, Katsaounis TI. Basic Phage Mathematics. *Methods Mol Biol*. 2018;1681. https://doi.org/10.1007/978-1-4939-7343-9_1,.

218. Sezonov G, Joseleau-Petit D, Ari RD. Escherichia coli Physiology in Luria-Bertani Broth. *J Bacteriol*. 2007;189(23):8746-8749. doi:10.1128/JB.01368-07

219. Koskella B, Meaden S. Understanding Bacteriophage Specificity in Natural Microbial Communities. 2013:806-823. doi:10.3390/v5030806

220. Jakociune D, Moodley A. A Rapid Bacteriophage DNA Extraction Method. *Methods and Protoc*. 2018;1(27):5-9. doi:10.3390/mps1030027

221. Soleimani-delfan A, Bouzari M, Wang R. A rapid competitive method for bacteriophage genomic DNA extraction. *J Virol Methods*. 2021;293(April):114148. doi:10.1016/j.jviromet.2021.114148

222. Dairawan M, Shetty PJ. The Evolution of DNA Extraction Methods. 2020:39-45. doi:10.34297/AJBSR.2020.08.001234.Received

223. Templar A, Schofield DM, Nesbeth DN. Measuring E. coli and bacteriophage DNA in cell sonicates to evaluate the CAL1 reaction as a synthetic biology standard for qPCR. *Biomol Detect Quantif*. 2016:4-13. doi:10.1016/j.bdq.2016.12.001

224. Ackermann HW. TAILED BACTERIOPHAGES: THE ORDER CAUDOVIRALES. *Adv Virus Res*. 1998;51:135-201. doi: 10.1016/s0065-3527(08)60785-x.

225. Comeau AM, Tremblay D, Moineau S, et al. Phage Morphology Recapitulates Phylogeny : The Comparative Genomics of a New Group of Myoviruses. *PLoS One*. 2012;7(7):1-11. doi:10.1371/journal.pone.0040102

226. Turner D, Kropinski AM. A Roadmap for Genome-Based Phage Taxonomy. 2021:1-10.

227. Sieber P, Platzer M, Schuster S. The Definition of Open Reading Frame Revisited. *Trends Genet*. 2018;34(3):167-170. doi:10.1016/j.tig.2017.12.009

228. Zeng J, Pan Y, Yang J, Hou M, Zeng Z, Xiong W. Metagenomic insights into the distribution of antibiotic resistome between the gut-associated environments and the pristine environments. *Environ Int*. 2019;126(December 2018):346-354. doi:10.1016/j.envint.2019.02.052

229. Murphy KC. *Phage Recombinases and Their Applications*. Vol 83. 1st ed. Elsevier Inc.; 2012. doi:10.1016/B978-0-12-394438-2.00008-6

230. Nafissi N, Slavcev R. Bacteriophage recombination systems and biotechnical applications. 2014:2841-2851. doi:10.1007/s00253-014-5512-2

231. Yuan Y, Gao M. Proteomic Analysis of a Novel Bacillus Jumbo Phage Revealing Glycoside Hydrolase As Structural Component. 2016;7(May):1-11. doi:10.3389/fmicb.2016.00745

232. Cahill J, Young R. *Phage Lysis : Multiple Genes for Multiple Barriers*. 1st ed. Elsevier Inc.; 2019. doi:10.1016/bs.aivir.2018.09.003

233. Summer EJ, Berry J, Tran TAT, Niu L, Struck DK, Young R. Rz / Rz1 Lysis Gene Equivalents in Phages of Gram-negative Hosts. 2007:1098-1112. doi:10.1016/j.jmb.2007.08.045

234. Yap ML, Rossmann MG. Structure and function of bacteriophage T4. *Futur Microbiol*. 2014;9:1319-1327. doi:10.2217/fmb.14.91.Structure

235. Boyd EF. *Bacteriophage-Encoded Bacterial Virulence Factors and Phage-Pathogenicity Island Interactions*. Vol 82. 1st ed. Elsevier Inc.; 2012. doi:10.1016/B978-0-12-394621-8.00014-5

236. Payne LJ, Meaden S, Mestre MR, Palmer C, Fineran PC, Jackson SA. PADLOC : a web server for the identification of antiviral defence systems in microbial genomes. *Nucleic Acids Res*. 2022;50:541-550. https://doi.org/10.1093/nar/gkac400.

237. Tesson F, Georjon H, Vaysset H, et al. A Comprehensive Resource for Exploring Antiphage Defense: DefenseFinder Webservice, Wiki and Databases. *bioRxiv*. 2024. https://doi.org/10.1101/2024.01.25.577194.

238. Hara BJO, Barth ZK, Mckitterick AC, Seed KD. A highly specific phage defense system is a conserved feature of the Vibrio cholerae mobilome. 2017:1-17.

239. Stanley SY, Maxwell KL. Phage-Encoded Anti-CRISPR Defenses. 2018.

240. Iida S, Streiff MB, Bickle TA, Arber W. Two DNA Antirestriction Systems Characterization of Bacteriophage Pi , darA , and d & 3 : of darA- Phages Bacteriophage Pl is resistant to DNA restriction. *Virology*. 1987;157:156-166.

241. Walkinshaw MD, Taylor P, Sturrock SS, et al. Structure of Ocr from Bacteriophage T7 , a Protein that Mimics B-Form DNA. 2002;9:187-194.

242. Otsuka Y, Yonesaki T. Dmd of bacteriophage T4 functions as an antitoxin against Escherichia coli LsoA and RnlA toxins. 2012;83(January):669-681. doi:10.1111/j.1365-2958.2012.07975.x

243. Biggs KRH, Bailes CL, Scott L, Wichman HA, Schwartz EJ. Ecological Approach to Understanding Superinfection Inhibition in Bacteriophage. 2021:1-17.

244. Berngruber TW, Weissing FJ, Gandon S. Inhibition of Superinfection and the Evolution of Viral Latency. 2010;84(19):10200-10208. doi:10.1128/JVI.00865-10

245. Chin C, Zih Y, Wang Y, Ming C. Insights into Superinfection Immunity Regulation of Xanthomonas axonopodis Filamentous Bacteriophage cf. *Curr Microbiol*. 2024. doi:10.1007/s00284-023-03539-y

246. Shi K, Oakland JT, Kurniawan F, Moeller NH, Banerjee S, Aihara H. Structural basis of superinfection exclusion by bacteriophage T4 Spackle. *Commun Biol*. 2020;3(691). doi:10.1038/s42003-020-01412-3

247. Sun J, Lu F, Luo Y, Bie L, Xu L, Wang Y. OrthoVenn3 : an integrated platform for exploring and visualizing

orthologous data across genomes. 2023;51(April):397-403.

248. Ferreira R, Amado R, Padr J, et al. The first sequenced Sphaerotilus natans bacteriophage – characterization and potential to control its filamentous bacterium host. 2021;(June 2020):1-13. doi:10.1093/femsec/fiab029

249. Ferreira R, Gonçalves RFS, Pinheiro AC, et al. Characterization and Genomic Analysis of a New Phage Infecting Helicobacter pylori. *Int J Mol Sci*. 2022;23(7885). https://doi.org/ 10.3390/ijms23147885.

250. Lavigne R, Seto D, Mahadevan P. Unifying classical and molecular taxonomic classification : analysis of the Podoviridae using BLASTP-based tools. 2008;159:406-414. doi:10.1016/j.resmic.2008.03.005

251. Meier-Kolthoff JP, Goker M. VICTOR : genome-based phylogeny and classification of prokaryotic viruses. *Bioinformatics*. 2017;33(21):3396-3404. doi:10.1093/bioinformatics/btx440

252. Korf IHE, Meier-koltho JP, Adriaenssens EM, et al. Still Something to Discover : Novel Insights into Escherichia coli Phage Diversity and Taxonomy. *Viruses*. 2019;11(454):1-29. doi:10.3390/v11050454.

253. Lavigne R, Darius P, Summer EJ, et al. Classification of Myoviridae bacteriophages using protein sequence similarity. *BMC Microbiol*. 2009;9(224):1-16. doi:10.1186/1471-2180-9-224

254. Sørensen PE, Broeck W Van Den, Kiil K, et al. New insights into the biodiversity of coliphages in the intestine of poultry. *Sci Rep*. 2020:1-12. doi:10.1038/s41598-020-72177-2

255. Sprotte S, Brinks E, Wagner N, Kropinski AM, Neve H, Franz CMAP. Characterization of the first Pseudomonas grimontii bacteriophage ,. *Arch Virol*. 2021;166(10):2887-2894. doi:10.1007/s00705-021-05173-0

256. Pattabiraman V, Bopp CA. Draft Whole-Genome Sequences of 10 Serogroup O6 Enterotoxigenic Escherichia coli Strains. 2014;2(6):2013-2014. doi:10.1128/genomeA.01274-14.Copyright

257. LeGault KN, Hays SG, Angermeyer A, et al. Temporal shifts in antibiotic resistance elements govern phage-pathogen conflicts. *Science (80-).* 2021;373(534). doi:10.1126/science.abg2166

258. Bondy-Denomy J, Davidson AR. When a Virus is not a Parasite : The Beneficial Effects of Prophages on Bacterial Fitness. *J Microbiol*. 2014;52(3):235-242. doi:10.1007/s12275-014-4083-3

259. Dy RL, Przybilski R, Semeijn K, Salmond GPC, Fineran C. A widespread bacteriophage abortive infection system functions through a Type IV toxin – antitoxin mechanism. 2014;42(7):4590-4605. doi:10.1093/nar/gkt1419

260. Zhang Y, Inouye M. RatA (YfjG), An Escherichia Coli Toxin, Inhibits 70S Ribosome Association to Block Translation Initiation. *Mol Microbiol*. 2011;79(6):1418-1429. doi:10.1111/j.1365-2958.2010.07506.x.RatA

261. Fino C, Vestergaard M, Ingmer H, Pierrel F, Gerdes K, Harms A. PasT of Escherichia coli sustains antibiotic tolerance and aerobic respiration as a bacterial homolog of mitochondrial Coq10. 2020;(May):1-36. doi:10.1002/mbo3.1064

262. Fozo EM, Makarova KS, Shabalina SA, Yutin N, Koonin E V, Storz G. Abundance of type I toxin – antitoxin systems in bacteria : searches for new candidates and discovery of novel families. 2010;38(11):3743-3759. doi:10.1093/nar/gkq054

263. Wilmaerts D, Loose P De, Vercauteren S, De S, Verstraeten N, Michiels J. Functional analysis of cysteine residues of the Hok / Gef type I toxins in Escherichia coli. 2021;(October 2020):1-9.

264. Prolic-Kalinsek M, Volkov AN, Hadzi S, et al. Structural basis of DNA binding by YdaT , a functional equivalent of the CII repressor in the cryptic prophage CP-933P from Escherichia coli O157:H7. *Acta Cryst*. 2023;79:245-258. doi:10.1107/S2059798323001249

265. Jurėnas D, Fraikin N, Goormaghtigh F, et al. Bistable Expression of a Toxin-Antitoxin System Located in a Cryptic Prophage of Escherichia coli O157:H7. 2021;12(6). doi: 10.1128/mBio.02947-21.

266. Benler S, Koonin E V. Phage lysis-lysogeny switches and programmed cell death : Danse macabre. *BioEssays*. 2020;42. doi:10.1002/bies.202000114

267. Koyuncu E, Budayeva HG, Miteva Y V., et al. Sirtuins Are Evolutionarily Conserved Viral Restriction Factors. 2014;5(6). doi:10.1128/mBio.02249-14.Editor

268. Cumby N, Edwards AM, Davidson AR, Maxwell KL. The Bacteriophage HK97 gp15 Moron Element Encodes a Novel. 2009;2009(33). doi:10.1128/JB.00843-12

269. Kim K, Yang HJ, Choi E, et al. The multi-target antibiotic efficacy of Angelica dahurica Bentham et Hooker extract exposed to the Escherichia coli O157 : H7. 2011;5:333-342. doi:10.1007/s13206-011-5407-6

270. Wilson GG, Murray NE. RESTRICTION AND MODIFICATION SYSTEMS. *Annu Rev Genet*. 1991;25:585-627. https://doi.org/10.1146/annurev.ge.25.120191.003101.

271. Sergueev K, Yu D, Austin S, Court D. Cell toxicity caused by products of the p L operon of bacteriophage lambda. 2001;272:227-235.

272. Haeusser DP, Hoashi M, Weaver A, et al. The Kil Peptide of Bacteriophage l Blocks Escherichia coli Cytokinesis via ZipA-Dependent Inhibition of FtsZ Assembly. 2014;10(3). doi:10.1371/journal.pgen.1004217

273. Shaw W V., Leslie GW. Chloramphenicol acetyltransferase. *Annu Rev Biophys Biophys Chern*. 1991;20:363-386.

274. Harper DR, Abedon ST, Burrowes BH, McConville ML. *Bacteriophages: Biology, Technology, Therapy*.

1st ed. Springer Cham; 2021. https://doi.org/10.1007/978-3-319-41986-2%0A©.

275. Amabilis-sosa LE, Monserrat V, Garc JL. Efficient Bacteria Inactivation by Ultrasound in Municipal Wastewater. 2018:1-13. doi:10.3390/environments5040047

6. Complementary information

Annex 1

Table A – Strains of ETEC that were selected to do the propagation, isolation, and characterization of phages as hosts. Highlighted in orange are the strain used in both isolation and characterization stages. Highlighted in blue are the strain used only in isolation stage. The remaining strains were only used during characterization stage. The "EC" strains were obtained in Spanish pig farms and the "SP" were obtained in Portuguese pig farms.

Bacterium	Strains			
Enterotoxigenic *Escherichia coli*	EC1	EC27	EC55	SP100
	EC2	EC28	EC56	SP101
	EC3	EC29	EC57	SP102
	EC4	EC31	EC58	SP103
	EC5	EC33	EC59	SP106
	EC6	EC34	EC60	SP107
	EC7	EC35	EC61	SP123
	EC8	EC36	EC62	SP125
	EC9	EC37	EC63	SP126
	EC10	EC38	EC64	SP127
	EC11	EC39	EC65	SP130
	EC12	EC40	EC66	SP131
	EC13	EC41	EC67	SP137
	EC14	EC42	EC68	SP140
	EC15	EC43	EC69	SP143
	EC16	EC44	EC70	SP145
	EC17	EC45	SP5	SP146
	EC18	EC46	SP10	SP162
	EC19	EC47	SP11	SP167
	EC20	EC48	SP18	SP168
	EC21	EC49	SP23	
	EC22	EC50	SP28	
	EC23	EC51	SP87	
	EC24	EC52	SP94	
	EC25	EC53	SP98	
	EC26	EC54	SP99	

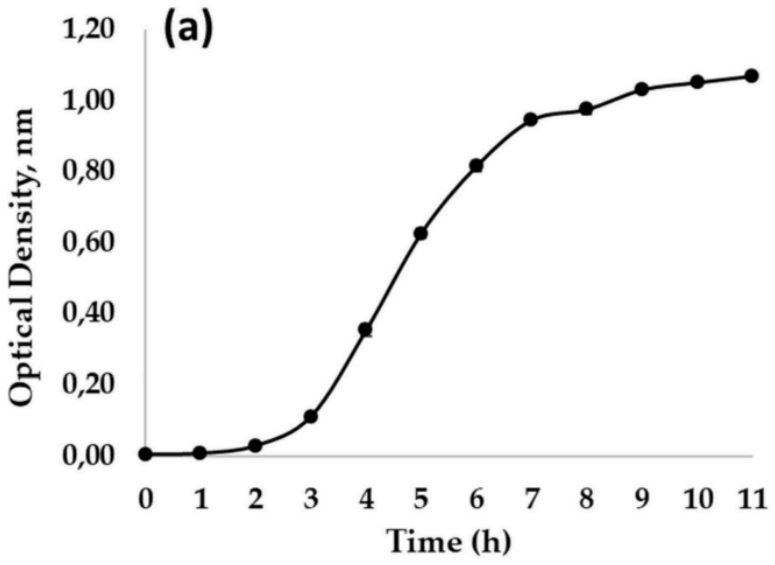

Figure A – *E. coli* growth curve, cultivated at 37°C and 120 rpm for 11 hours. The curve represents the evolution of OD_{600} with time (hours). This data is adapted from Amabilis-Sosa *et al.* [275].

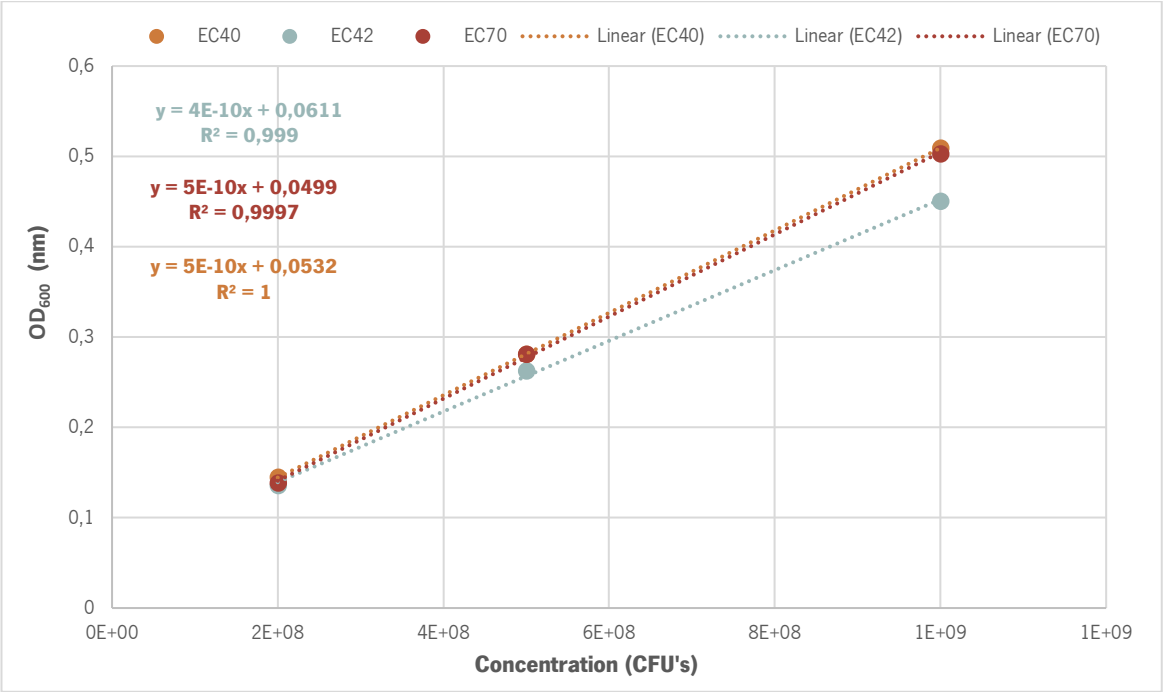

Figure B – Calibration curve of ETEC strains EC40, EC42 and EC70. It represents the relation between the bacterial concentration (CFU's) and its respective OD_{600}. From the calibration curve we can retrieve the equation to calculate the bacterial concentration at any value of OD.

Table B – Competitive assay of 11 strains of ETEC that were selected to do the propagation and isolation of phages as hosts. The columns represent the strains used as bacterial lawns and the rows represent the strains used as spots. The blue represents clear plaques, without formation of phage plaques. The orange represents formation of phage plaques.

Lawn / Spot	EC 16	EC 22	EC 34	EC 36	EC 42	EC 52	EC 53	EC 65	SP 127	SP 130	SP 143
EC 16											
EC 22											
EC 34											
EC 36											
EC 42											
EC 52											
EC 53											
EC 65											
SP 127											
SP 130											
SP 143											

Annex 2

Table C – Functional annotation of phage EcoSus34. The ORFs sequences were ran through BlastP and HHpred softwares. The functional attribution of ORFs encoded proteins were based on a comparison with a close homolog, considering the E-value, where blank spaces represent irrelevant homolog (E-value ≥ 1×10^{-5}). In the table is also represented the start and stop site (bp), the direction and the size (aa) of the sequences.

ORF	Start (bp)	End (bp)	Direction	Size (aa)	Function	Closest homolog	E-value
gp_0001	61	240	reverse	59	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	1×10^{-32}
gp_0002	249	467	reverse	72	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	5×10^{-42}
gp_0003	418	627	reverse	69	Hypothetical protein		
gp_0004	630	962	reverse	110	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	4×10^{-69}
gp_0005	1024	1269	reverse	81	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	7×10^{-40}
gp_0006	1271	1459	reverse	62	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	4×10^{-38}
gp_0007	1459	1644	reverse	61	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	3×10^{-36}
gp_0008	1641	1898	reverse	85	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	2×10^{-52}
gp_0009	1891	2061	reverse	56	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	3×10^{-30}
gp_0010	2116	2745	reverse	209	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	9×10^{-151}
gp_0011	2736	3704	reverse	322	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	0
gp_0012	3670	4305	reverse	211	Alpha-2;3-sialyltransferase	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	8×10^{-151}
gp_0013	4269	4790	reverse	173	Alpha-2;3-sialyltransferase	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	5×10^{-122}
gp_0014	4802	5509	reverse	235	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	4×10^{-165}
gp_0015	5509	5739	reverse	76	Hypothetical protein	Hypothetical protein FF15_gp156 [*Salmonella* phage vB-SalM-SJ3]	2×10^{-14}
gp_0016	5788	6105	reverse	105	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	8×10^{-72}

gp_0017	6102	6521	reverse	139	PinA peptidase inhibitor	Hypothetical protein [*Escherichia* phage EP_H11]	9×10^{-91}
gp_0018	6508	6741	reverse	77	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	6×10^{-46}
gp_0019	6796	7050	reverse	84	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	2×10^{-45}
gp_0020	7050	8993	reverse	647	DNA ligase	Putative DNA ligase [*Escherichia* phage vB_EcoM_IME392]	0
gp_0021	9058	9816	reverse	252	DNA polymerase	Hypothetical protein [*Escherichia* phage EP_H11]	0
gp_0022	9813	9980	reverse	55	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	3×10^{-30}
gp_0023	9982	10353	reverse	123	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	5×10^{-87}
gp_0024	10365	10484	reverse	39	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	2×10^{-12}
gp_0025	10588	10830	reverse	80	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	2×10^{-52}
gp_0026	10830	11207	reverse	125	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	3×10^{-86}
gp_0027	11197	11358	reverse	53	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	5×10^{-30}
gp_0028	11358	11564	reverse	68	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	2×10^{-41}
gp_0029	11572	11868	reverse	98	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	1×10^{-61}
gp_0030	11871	12245	reverse	124	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	5×10^{-74}
gp_0031	12254	12565	reverse	103	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	4×10^{-69}
gp_0032	12597	13010	reverse	137	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	8×10^{-90}
gp_0033	13023	13241	reverse	72	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	5×10^{-44}
gp_0034	13252	13626	reverse	124	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	2×10^{-87}
gp_0035	13686	13970	reverse	94	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	2×10^{-60}
gp_0036	14083	14433	reverse	116	Hypothetical protein	Hypothetical protein [*Colwellia* sp.]	9×10^{-6}
gp_0037	14414	14719	reverse	101	Hypothetical protein		
gp_0038	14795	15139	reverse	114	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	1×10^{-77}
gp_0039	15132	15638	reverse	168	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	4×10^{-117}

gp_0040	15644	15943	reverse	99	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	2×10^{-64}
gp_0041	15918	16448	reverse	176	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	2×10^{-127}
gp_0042	16448	18016	reverse	522	DNA-directed RNA polymerase	Putative DNA-directed RNA polymerase beta subunit [*Escherichia* phage vB_EcoM_IME392]	0
gp_0043	18032	18688	reverse	218	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	1×10^{-161}
gp_0044	18895	19587	reverse	230	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	4×10^{-167}
gp_0045	19596	21110	reverse	504	DNA-directed RNA polymerase	Hypothetical protein [*Escherichia* phage EP_H11]	0
gp_0046	21130	21444	reverse	104	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	2×10^{-70}
gp_0047	21431	21775	reverse	114	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	4×10^{-77}
gp_0048	21772	22239	reverse	155	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	5×10^{-110}
gp_0049	22372	22611	reverse	79	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	4×10^{-50}
gp_0050	22766	24226	reverse	486	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	0
gp_0051	24367	24537	reverse	56	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	1×10^{-33}
gp_0052	24617	27904	reverse	1095	DNA polymerase	Hypothetical protein [*Escherichia* phage EP_H11]	0
gp_0053	27974	29611	reverse	545	DNA helicase	Putative DNA helicase [*Escherichia* phage vB_EcoM_IME392]	0
gp_0054	29575	30504	reverse	309	DNA primase	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	0
gp_0055	30557	30877	reverse	106	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	3×10^{-71}
gp_0056	30867	31466	reverse	199	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	3×10^{-143}
gp_0057	31551	32624	reverse	357	Thymidylate synthase	Putative thymidylate synthase [*Escherichia* phage vB_EcoM_IME392]	0
gp_0058	32621	33343	reverse	240	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	5×10^{-179}
gp_0059	33433	34362	reverse	309	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	0
gp_0060	34404	35483	reverse	359	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	0
gp_0061	35531	36565	reverse	344	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	0
gp_0062	36572	37183	reverse	203	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	3×10^{-149}

gp_0063	37270	38652	reverse	460	BCCT family transporter	High-affinity choline uptake protein [*Escherichia* phage EP_H11]	0
gp_0064	38656	39564	reverse	302	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	0
gp_0065	39666	40406	reverse	246	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	5×10^{-179}
gp_0066	40511	41758	reverse	415	DNA topoisomerase	Putative DNA topoisomerase subunit A [*Escherichia* phage vB_EcoM_IME392]	0
gp_0067	41761	43746	reverse	661	DNA topoisomerase	DNA gyrase subunit B [*Escherichia* phage EP_H11]	0
gp_0068	43801	44064	reverse	87	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	2×10^{-56}
gp_0069	44061	44300	reverse	79	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	2×10^{-47}
gp_0070	44448	47207	reverse	919	DNA repair exonuclease	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	0
gp_0071	47210	48067	reverse	285	DNA polymerase	Putative DNA polymerase I [*Escherichia* phage vB_EcoM_IME392]	0
gp_0072	48120	49118	reverse	332	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	0
gp_0073	49215	50255	reverse	346	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	0
gp_0074	50252	50479	reverse	75	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	1×10^{-45}
gp_0075	50483	50932	reverse	149	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	5×10^{-105}
gp_0076	50932	52191	reverse	419	Replicative helicase	Putative DNA helicase [*Escherichia* phage vB_EcoM_IME392]	0
gp_0077	52201	53841	reverse	546	Helicase	Putative DNA helicase [*Escherichia* phage vB_EcoM_IME392]	0
gp_0078	53844	54362	reverse	172	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	3×10^{-121}
gp_0079	54408	54719	reverse	103	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	2×10^{-38}
gp_0080	54763	55008	reverse	81	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	8×10^{-52}
gp_0081	54998	55153	reverse	51	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	3×10^{-27}
gp_0082	55412	56164	reverse	250	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	0
gp_0083	57872	58009	forward	45	Hypothetical protein		
gp_0084	58756	59298	reverse	180	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	5×10^{-125}
gp_0085	59449	59763	reverse	104	Hypothetical protein		

gp_0086	59862	60023	reverse	53	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	2×10^{-27}
gp_0087	60010	60492	reverse	160	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	7×10^{-113}
gp_0088	60492	60833	reverse	113	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	4×10^{-76}
gp_0089	61030	61503	forward	157	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	2×10^{-104}
gp_0090	61507	63333	forward	608	Terminase (large subunit)	Hypothetical protein [*Escherichia* phage EP_H11]	0
gp_0091	63334	64911	forward	525	Portal protein	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	0
gp_0092	64904	65158	forward	84	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	9×10^{-50}
gp_0093	65171	66607	forward	478	Protease (prohead core)	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	0
gp_0094	66663	68090	forward	475	Capsid protein	Hypothetical protein [*Escherichia* phage EP_H11]	0
gp_0095	68241	68951	forward	236	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	2×10^{-162}
gp_0096	68985	69521	forward	178	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	8×10^{-128}
gp_0097	69521	69808	forward	95	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	5×10^{-58}
gp_0098	69805	70434	forward	209	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	4×10^{-154}
gp_0099	70444	71007	forward	187	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	9×10^{-131}
gp_0100	71009	71674	forward	221	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	5×10^{-160}
gp_0101	71676	71867	forward	63	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	1×10^{-34}
gp_0102	71878	73677	forward	599	Tail sheath protein	Putative tail sheath protein [*Escherichia* phage vB_EcoM_IME392]	0
gp_0103	73688	74212	forward	174	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	2×10^{-124}
gp_0104	74275	74703	forward	142	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	4×10^{-102}
gp_0105	74713	75540	forward	275	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	0
gp_0106	75619	78459	forward	946	Glycoside hydrolase (chitinase)	Membrane-bound lytic murein transglycosylase d precursor [*Escherichia* phage EP_H11]	0
gp_0107	78459	79094	forward	211	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	4×10^{-152}
gp_0108	79102	79479	forward	125	Baseplate wedge subunit	Hypothetical protein [*Escherichia* phage EP_H11]	7×10^{-83}

gp_0109	79476	80600	forward	374	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	0
gp_0110	80603	81262	forward	219	Baseplate assembly protein	Putative tail protein [*Escherichia* phage vB_EcoM_IME392]	2×10^{-158}
gp_0111	81270	81638	forward	122	Baseplate wedge subunit	Hypothetical protein [*Escherichia* phage EP_H11]	4×10^{-85}
gp_0112	81635	83008	forward	457	Baseplate wedge protein	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	0
gp_0113	82986	84479	forward	497	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	0
gp_0114	84488	85255	forward	255	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	4×10^{-168}
gp_0115	85268	86218	forward	316	Tail protein	Putative tail fibers protein [*Escherichia* phage vB_EcoM_IME392]	0
gp_0116	86228	86554	forward	108	Hypothetical protein	Hypothetical protein ufovp237_12 [uncultured caudovirales phage]	2×10^{-11}
gp_0117	86567	86989	forward	140	Tail assembly chaperone	Hypothetical protein hos14_gp120 [bordetella phage vb_bbrm_phb04]	8×10^{-14}
gp_0118	87000	88364	forward	454	Hypothetical protein	Deduced tail fiber protein [xanthomonas phage op1h]	3×10^{-9}
gp_0119	88599	89744	forward	381	Hypothetical protein	Tail fiber protein [synechococcus phage s-sbp1]	2×10^{-6}
gp_0120	89763	90878	forward	371	Tail fiber protein	Putative tail fiber protein [*Escherichia* phage vB_EcoM-ro157lw]	1×10^{-34}
gp_0121	90888	93251	forward	787	Tailspike protein	Hypothetical protein ljijohlm_00121 [*Escherichia* phage kkp 3954]	0
gp_0122	93660	96281	forward	873	Colanic acid-degrading protein	Putative colanic acid-degrading protein [*Escherichia* phage vB_EcoM_IME392]	0
gp_0123	96306	97757	forward	483	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	0
gp_0124	97774	98595	forward	273	Baseplate wedge protein	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	0
gp_0125	98607	101651	forward	1014	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	0
gp_0126	101661	101837	forward	58	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	2×10^{-31}
gp_0127	101862	102509	forward	215	Glycoside hydrolase (chitinase)	Putative lysis protein [*Escherichia* phage vB_EcoM_IME392]	1×10^{-156}
gp_0128	102524	102940	forward	138	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	2×10^{-88}
gp_0129	102944	103174	forward	76	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	1×10^{-31}
gp_0130	103243	103575	forward	110	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	1×10^{-64}

gp_0131	103628	103870	reverse	80	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	8×10^{-44}
gp_0132	103870	104424	reverse	184	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	2×10^{-119}
gp_0133	104421	104783	reverse	120	Cell wall hydrolase	Hypothetical protein [*Escherichia* phage EP_H11]	4×10^{-84}
gp_0134	104815	105192	reverse	125	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	3×10^{-83}
gp_0135	105205	105828	reverse	207	Deaminase (dCTP)	Hypothetical protein [*Escherichia* phage EP_H11]	3×10^{-148}
gp_0136	105828	106013	reverse	61	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	8×10^{-30}
gp_0137	106099	106566	reverse	155	Diphosphatase (dUTP)	Deoxyuridine 5'-triphosphate nucleotidohydrolase [*Escherichia* phage EP_H11]	6×10^{-104}
gp_0138	106566	106946	reverse	126	Ryr domain-containing protein	Hypothetical protein [*Escherichia* phage EP_H11]	6×10^{-81}
gp_0139	106946	107170	reverse	74	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	3×10^{-29}
gp_0140	107522	108235	reverse	237	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	1×10^{-172}
gp_0141	108300	108653	reverse	117	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	1×10^{-73}
gp_0142	108646	108885	reverse	79	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	4×10^{-45}
gp_0143	108940	109596	reverse	218	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	1×10^{-152}
gp_0144	109596	110165	reverse	189	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM_IME392]	2×10^{-137}
gp_0145	110155	110409	reverse	84	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	3×10^{-41}
gp_0146	110419	110583	reverse	54	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	3×10^{-27}
gp_0147	110624	110776	reverse	50	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	2×10^{-27}
gp_0148	110763	110957	reverse	64	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	4×10^{-36}
gp_0149	111018	111251	reverse	77	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	1×10^{-48}
gp_0150	111783	111953	reverse	56	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	1×10^{-31}
gp_0151	111946	112305	reverse	119	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	2×10^{-78}
gp_0152	112408	112641	reverse	77	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	1×10^{-34}
gp_0153	112652	112963	reverse	103	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	1×10^{-47}

gp_0154	112963	113106	reverse	47	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	6×10^{-22}
gp_0155	113187	113726	reverse	179	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	7×10^{-94}
gp_0156	113723	114205	reverse	160	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	6×10^{-29}
gp_0157	114198	114383	reverse	61	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	3×10^{-31}
gp_0158	114443	114640	reverse	65	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP_H11]	1×10^{-38}
gp_0159	114634	114747	reverse	37	Hypothetical protein		
gp_0160	114740	115030	reverse	96	Hypothetical protein	Hypothetical protein va1_031 [*Vibrio* phage va1]	1×10^{-20}

Table D – Functional annotation of phage EcoSus42. The ORFs sequences were ran through BlastP and HHpred softwares. The functional attribution of ORFs encoded proteins were based on a comparison with a close homolog, considering the E-value, where blank spaces represent irrelevant homolog (E-value ≥ 1×10^{-5}). In the table is also represented the start and stop site (bp), the direction and the size (aa) of the sequences.

ORF	Start (bp)	End (bp)	Direction	Size (aa)	Product	Closest homolog	E-value
gp_0001	50	373	forward	107	Hypothetical protein	Hypothetical protein LD35_gp01 [*Escherichia* phage vB_EcoP_PhAPEC7]	1×10^{-69}
gp_0002	460	606	forward	48	Hypothetical protein	Hypothetical protein DPIBCGCG_00085 [*Escherichia* phage KKP 3715]	9×10^{-18}
gp_0003	608	787	forward	59	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM_PD205]	2×10^{-11}
gp_0004	800	1096	forward	98	Hypothetical protein	Hypothetical protein PP764_gp66 [*Escherichia* phage phi G17]	2×10^{-47}
gp_0005	1177	1383	forward	68	Hypothetical protein	MAG: hypothetical protein [*Caudoviricetes* sp.]	5×10^{-42}
gp_0006	1516	1899	forward	127	Hypothetical protein	RNA polymerase subunit A [*Escherichia* phage vB_Eco_F22]	6×10^{-87}
gp_0007	2045	2344	forward	99	Hypothetical protein	Hypothetical protein p4b_00043 [Klebsiella phage vlcpip4b]	2×10^{-55}
gp_0008	2350	2646	forward	98	Hypothetical protein	Hypothetical protein E20_05 [*Escherichia* phage E20]	9×10^{-56}
gp_0009	2651	2869	forward	72	Hypothetical protein	Hypothetical protein DPIBCGCG_00080 [*Escherichia* phage KKP 3715]	9×10^{-42}
gp_0010	2866	3039	forward	57	Hypothetical protein	Hypothetical protein gp2.4 [*Escherichia* phage vB_EcoP_G7C]	3×10^{-34}
gp_0011	3032	3166	forward	44	Hypothetical protein		
gp_0012	3166	3378	forward	70	Hypothetical protein	Hypothetical protein IME11_70 [*Escherichia* phage IME11]	1×10^{-31}
gp_0013	3460	3768	forward	102	Hypothetical protein	MAG: hypothetical protein [*Caudoviricetes* sp.]	8×10^{-67}
gp_0014	3765	3986	forward	73	Hypothetical protein	Hypothetical protein PP766_gp11 [*Escherichia* phage U1G]	2×10^{-46}
gp_0015	3983	4339	forward	118	DNA processing protein	Hypothetical protein DPIBCGCG_00072 [*Escherichia* phage KKP 3715]	1×10^{-67}
gp_0016	4340	4597	forward	85	Hypothetical protein	Hypothetical protein gp9.2 [*Escherichia* phage vB_EcoP_G7C]	4×10^{-53}
gp_0017	4594	4908	forward	104	Hypothetical protein	Hypothetical protein E20_14 [*Escherichia* phage E20]	1×10^{-68}
gp_0018	4905	5285	forward	126	Hypothetical protein	Ribonucleotide reductase NrdA-like [*Escherichia* phage vB_EcoP_SP5M]	1×10^{-84}

gp_0019	5278	5463	forward	61	Hypothetical protein	MAG TPA: hypothetical protein [*Caudoviricetes* sp.]	1×10^{-32}
gp_0020	5460	5840	forward	126	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoS_Uz-1]	8×10^{-86}
gp_0021	5874	6098	forward	74	Hypothetical protein	MAG: hypothetical protein [*Caudoviricetes* sp.]	4×10^{-36}
gp_0022	6121	6939	forward	272	RNA polymerase	RNA polymerase subunit 1 [*Escherichia* phage vB_EcoP_SP5M]	0
gp_0023	6929	7090	forward	53	Hypothetical protein	Hypothetical protein LD35_gp17 [*Escherichia* phage vB_EcoP_PhAPEC7]	7×10^{-28}
gp_0024	7145	8362	forward	405	RNA polymerase	MAG: hypothetical protein [*Caudoviricetes* sp.]	0
gp_0025	8679	9512	forward	277	Hoc-like head decoration	Hoc-like head decoration [*Escherichia* phage vB_EcoP_SP5M]	5×10^{-180}
gp_0026	9618	9806	forward	62	Hypothetical protein	Hypothetical protein DPIBCGCG_00060 [*Escherichia* phage KKP 3715]	1×10^{-34}
gp_0027	9803	10012	forward	69	Hypothetical protein	MAG TPA: hypothetical protein [*Caudoviricetes* sp.]	5×10^{-42}
gp_0028	10045	10194	forward	49	Hypothetical protein	Hypothetical protein E20_26 [*Escherichia* phage E20]	2×10^{-24}
gp_0029	10195	10521	forward	108	Hypothetical protein	Hypothetical protein PP763_gp27 [*Escherichia* phage vB_EcoP_SP5M]	1×10^{-68}
gp_0030	10556	11608	forward	350	ATPase	ATPase [*Escherichia* phage vB_EcoP_PhAPEC7]	0
gp_0031	11616	12791	forward	391	Metallopeptidase	Hypothetical protein DPIBCGCG_00056 [*Escherichia* phage KKP 3715]	0
gp_0032	12791	13297	forward	168	Deoxycytidine triphosphate deaminase	DCTP deaminase [*Escherichia* phage phi G17]	2×10^{-117}
gp_0033	13328	13528	forward	66	Hypothetical protein	Hypothetical protein psb1_0069 [*Shigella* phage psb-1]	1×10^{-34}
gp_0034	13602	13862	forward	86	Superinfection immunity protein	Putative membrane immunity protein [*Escherichia* phage ST4]	3×10^{-50}
gp_0035	13878	14366	forward	162	Hypothetical protein	Hypothetical protein [*Escherichia* phage ST4]	3×10^{-88}
gp_0036	14366	14809	forward	147	Hypothetical protein	Hypothetical protein DPIBCGCG_00051 [*Escherichia* phage KKP 3715]	1×10^{-99}
gp_0037	14809	15753	forward	314	Thymidylate synthase	Thymidylate synthase [*Escherichia* phage E20]	0
gp_0038	15750	15986	forward	78	Hypothetical protein	Hypothetical protein [*Escherichia* phage ST4]	4×10^{-44}
gp_0039	16047	16259	forward	70	Hypothetical protein	Hypothetical protein PP763_gp37 [*Escherichia* phage vB_EcoP_SP5M]	7×10^{-45}
gp_0040	16252	16584	forward	110	Hypothetical protein	MAG: hypothetical protein [*Caudoviricetes* sp.]	5×10^{-72}
gp_0041	16577	16768	forward	63	Hypothetical protein	MAG: hypothetical protein [*Caudoviricetes* sp.]	1×10^{-38}

gp_0042	16796	19360	forward	854	RIIA lysis inhibitor	RiiA lysis inhibitor [*Escherichia* phage vB_Eco_F22]	0
gp_0043	19365	21416	forward	683	RIIB lysis inhibitor	RIIB lysis inhibitor [*Escherichia* phage vB_EcoP_PhAPEC5]	0
gp_0044	21416	21583	forward	55	Hypothetical protein	MAG TPA: hypothetical protein [Bacteriophage sp.]	1×10^{-30}
gp_0045	21643	22038	forward	131	Hypothetical protein	MAG: hypothetical protein [*Caudoviricetes* sp.]	3×10^{-91}
gp_0046	22067	22414	forward	115	Nucleoside triphosphate pyrophosphohydrolase	Nucleoside triphosphate pyrophosphohydrolase [*Escherichia* phage vB_Eco_F22]	2×10^{-72}
gp_0047	22448	23758	forward	436	DNA helicase	DNA helicase [*Escherichia* phage E20]	0
gp_0048	23769	24299	forward	176	Hypothetical protein	Hypothetical protein gp38 [*Escherichia* phage vB_EcoP_G7C]	1×10^{-126}
gp_0049	24309	26888	forward	859	DNA polymerase	DNA polymerase [*Escherichia* phage vB_EcoP_G7C]	0
gp_0050	26885	27190	forward	101	Hypothetical protein	Hypothetical protein DPIBCGCG_00039 [*Escherichia* phage KKP 3715]	4×10^{-68}
gp_0051	27190	27663	forward	157	Nucleotide kinase	Hypothetical protein PP766_gp52 [*Escherichia* phage U1G]	1×10^{-106}
gp_0052	27663	28643	forward	326	Exonuclease	MAG: hypothetical protein [*Caudoviricetes* sp.]	0
gp_0053	28697	30793	forward	698	Hypothetical protein	MAG: hypothetical protein [*Caudoviricetes* sp.]	0
gp_0054	30850	31602	forward	250	ATPase	AAA family ATPase [*Escherichia* phage vB_EcoP_G7C]	0
gp_0055	31643	32446	forward	267	DNA-binding protein	ssDNA-binding protein [*Escherichia* phage E20]	0
gp_0056	32446	33000	forward	184	RuvC-like Holliday junction resolvase	MAG: hypothetical protein [*Caudoviricetes* sp.]	3×10^{-131}
gp_0057	33002	33451	forward	149	Hypothetical protein	Hypothetical protein E20_54 [*Escherichia* phage E20]	8×10^{-95}
gp_0058	33511	33672	reverse	53	Hypothetical protein		
gp_0059	33687	34064	forward	125	Hypothetical protein	MAG: hypothetical protein [*Caudoviricetes* sp.]	2×10^{-69}
gp_0060	35462	35644	forward	60	Hypothetical protein		
gp_0061	35852	36028	forward	58	Hypothetical protein	MAG TPA: hypothetical protein [*Caudoviricetes* sp.]	5×10^{-5}
gp_0062	36098	36256	forward	52	Hypothetical protein	Hypothetical protein PP764_gp18 [*Escherichia* phage phi G17]	7×10^{-29}
gp_0063	36240	36620	forward	126	Hypothetical protein	Hypothetical protein LD33_gp83 [*Escherichia* phage vB_EcoP_PhAPEC5]	1×10^{-85}
gp_0064	36805	37227	forward	140	Hypothetical protein	Hypothetical protein F22_0059 [*Escherichia* phage vB_Eco_F22]	3×10^{-91}

gp_0065	37237	37470	forward	77	Hypothetical protein	Hypothetical protein IME11_18 [*Escherichia* phage IME11]	3×10^{-44}
gp_0066	37463	37669	forward	68	Hypothetical protein	Hypothetical protein BRM13314_00004 [*Salmonella* phage BRM 13314]	6×10^{-12}
gp_0067	37704	48416	reverse	3570	RNA polymerase	Virion RNA polymerase [*Escherichia* phage vB_EcoP_PhAPEC5]	0
gp_0068	48511	50475	reverse	654	Hypothetical protein	Virion structural protein [*Escherichia* phage vB_EcoP_SP5M]	0
gp_0069	50488	50931	reverse	147	Hypothetical protein	Structural protein [*Escherichia* phage ST4]	5×10^{-100}
gp_0070	50945	53599	reverse	884	Hypothetical protein	Hypothetical protein E20_66 [*Escherichia* phage E20]	0
gp_0071	53601	54437	reverse	278	Hypothetical protein	Virion structural protein [*Escherichia* phage vB_EcoP_PhAPEC7]	0
gp_0072	54514	55149	reverse	211	Hypothetical protein	Hypothetical protein [*Escherichia* phage ST4]	2×10^{-151}
gp_0073	55223	56425	reverse	400	Capsid protein	Major head protein [*Escherichia* phage vB_EcoP_PhAPEC7]	0
gp_0074	56442	57662	reverse	406	Tape measure protein	Tape measure protein [*Escherichia* phage ST4]	0
gp_0075	57682	58026	reverse	114	Hypothetical protein	Hypothetical protein LD35_gp73 [*Escherichia* phage vB_EcoP_PhAPEC7]	1×10^{-75}
gp_0076	58040	60310	reverse	756	Portal protein	Portal [*Escherichia* phage vB_Eco_F22]	0
gp_0077	60319	60537	reverse	72	Rz/Rzl spanin protein	Putative Rz/Rzl spanin protein [*Escherichia* phage vB_EcoS_Uz-1]	1×10^{-41}
gp_0078	60800	61435	reverse	211	Endolysin	Lysozyme [*Escherichia* phage vB_Eco_F22]	8×10^{-153}
gp_0079	61425	61559	reverse	44	Holin	Holin [*Escherichia* phage vB_EcoP_PhAPEC7]	9×10^{-23}
gp_0080	61670	62002	reverse	110	Hypothetical protein	Hypothetical protein PP763_gp72 [*Escherichia* phage vB_EcoP_SP5M]	8×10^{-74}
gp_0081	62111	64414	reverse	767	Tail spike protein	Hypothetical protein HERCULESSET_42 [*Salmonella* phage vB_Hercules_SET]	0
gp_0082	64469	65602	reverse	377	Tail protein	Tail protein [*Escherichia* phage phi G17]	0
gp_0083	65599	66309	reverse	236	Hypothetical protein	Hypothetical protein F22_0077 [*Escherichia* phage vB_Eco_F22]	2×10^{-175}
gp_0084	66316	67905	reverse	529	Terminase	MAG: hypothetical protein [*Caudoviricetes* sp.]	0
gp_0085	67898	68587	reverse	229	Hypothetical protein	Hypothetical protein PP763_gp66 [*Escherichia* phage vB_EcoP_SP5M]	4×10^{-168}
gp_0086	68759	69049	forward	96	Hypothetical protein	Hypothetical protein PP764_gp75 [*Escherichia* phage phi G17]	6×10^{-50}
gp_0087	69251	69649	forward	132	Hypothetical protein	MAG: hypothetical protein [*Caudoviricetes* sp.]	1×10^{-73}

gp_0088	69615	69929	forward	104	Hypothetical protein	Hypothetical protein DPIBCGCG_00004 [*Escherichia* phage KKP 3715]	4×10^{-62}
gp_0089	69933	70256	forward	107	Hypothetical protein	Hypothetical protein E20_84 [*Escherichia* phage E20]	4×10^{-38}
gp_0090	70253	70594	forward	113	Hypothetical protein	Hypothetical protein [*Escherichia* phage ST4]	1×10^{-74}
gp_0091	70591	70905	forward	104	Hypothetical protein	Hypothetical protein ECBP1_0082 [*Escherichia* phage ECBP1]	1×10^{-64}
gp_0092	70902	71228	forward	108	Hypothetical protein	MAG TPA: hypothetical protein [Bacteriophage sp.]	2×10^{-65}

Table E – Functional annotation of phage EcoSus65. The ORFs sequences were ran through BlastP and HHpred softwares. The functional attribution of ORFs encoded proteins were based on a comparison with a close homolog, considering the E-value, where blank spaces represent irrelevant homolog (E-value ≥ 1×10^{-5}). In the table is also represented the start and stop site (bp), the direction and the size (aa) of the sequences.

ORF	Start (bp)	End (bp)	Direction	Size (aa)	Product	Closest homolog	E-value
gp_0001	1	2178	reverse	725	Lysis inhibitor	RIIA lysis inhibitor [*Escherichia* phage slur02]	0
gp_0002	2189	2392	reverse	67	Hypothetical protein	RIIA.1 hypothetical protein [*Escherichia* phage RB14]	3×10^{-42}
gp_0003	2447	4264	reverse	605	DNA topoisomerase	DNA topoisomerase II large subunit [*Escherichia* phage vB_EcoM_IME537]	0
gp_0004	4334	4594	reverse	86	Hypothetical protein	Gp175 [*Shigella* phage Sf22]	9×10^{-59}
gp_0005	4600	4971	reverse	123	Hypothetical protein	Hypothetical protein KMC37_gp089 [*Yersinia* phage vB_YepM_ZN18]	1×10^{-82}
gp_0006	4974	5150	reverse	58	FmdB-like transcriptional regulator	FmdB-like transcriptional regulator [*Escherichia* phage RB14]	6×10^{-38}
gp_0007	5153	5563	reverse	136	Hypothetical protein	mRNA metabolism modulator [*Escherichia* phage vB_Eco_F31]	3×10^{-91}
gp_0008	5563	5778	reverse	71	Cef modifier of supressor tRNAs	Cef modifier of supressor tRNAs [*Escherichia* phage YUEEL01]	2×10^{-44}
gp_0009	5951	6442	reverse	163	Transcriptional regulator	Transcriptional regulator [*Escherichia* phage vB_EcoM_FJ1]	4×10^{-115}
gp_0010	6519	7061	reverse	180	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM_FJ1]	3×10^{-128}
gp_0011	7064	7564	reverse	166	Hypothetical protein	Hypothetical protein FDH36_gp239 [*Escherichia* phage HP3]	4×10^{-120}
gp_0012	7628	8311	reverse	227	Exonuclease	Exonuclease [*Escherichia* phage HY01]	1×10^{-167}
gp_0013	8311	8553	reverse	80	Hypothetical protein	Exonuclease [Enterobacteria phage Kha5h]	3×10^{-49}
gp_0014	8546	8791	reverse	81	Dextranase	Dextranase [Enterobacteria phage Aplg8]	2×10^{-49}
gp_0015	8778	9038	reverse	86	Hypothetical protein	Hypothetical protein F412_gp258 [*Escherichia* phage wv7]	1×10^{-53}
gp_0016	9045	10364	reverse	439	Helicase	Dda-like helicase [*Escherichia* phage vB_EcoM_112]	0
gp_0017	10361	10672	reverse	103	Hypothetical protein	Gp189 [*Shigella* phage Sf22]	1×10^{-68}
gp_0018	10674	11420	reverse	248	Hypothetical protein	Srd anti-sigma factor [*Escherichia* phage slur07]	3×10^{-177}
gp_0019	11543	12145	reverse	200	NAD–protein ADP-ribosyltransferase	RNA polymerase ADP-ribosylase [Enterobacteria phage RB27]	4×10^{-147}

gp_0020	12142	12765	reverse	207	NAD–protein ADP-ribosyltransferase	RNA polymerase ADP-ribosylase [*Escherichia* phage vB_EcoM_G4498]	3×10^{-150}
gp_0021	12833	13015	reverse	60	Hypothetical protein	Hypothetical protein RB51ORF023 [Enterobacteria phage RB51]	6×10^{-36}
gp_0022	13024	13494	reverse	156	Mrh transcription modulator under heat shock	Molybdenum ABC transporter, periplasmic molybdenum-binding protein [*Shigella* phage ESh17]	8×10^{-109}
gp_0023	13487	13639	reverse	50	Hypothetical protein	Hypothetical protein AS348_gp115 [*Escherichia* phage slur14]	2×10^{-26}
gp_0024	13648	13851	reverse	67	Hypothetical protein	Putative transcription modulator [*Escherichia* phage Killian]	1×10^{-38}
gp_0025	13826	14311	reverse	161	Modulating protein	Transcription modulator [*Escherichia* phage vB_Eco_F26]	3×10^{-111}
gp_0026	14320	14661	reverse	113	Hypothetical protein	Putative 12.6 kda protein [*Escherichia* phage W143]	2×10^{-75}
gp_0027	14661	14873	reverse	70	Hypothetical protein	Hypothetical protein [*Escherichia* phage EP01]	1×10^{-43}
gp_0028	14972	15208	reverse	78	Outer capsid protein	Virion structural protein [*Escherichia* phage RB14]	2×10^{-49}
gp_0029	15243	16079	reverse	278	Endonuclease	HNH endonuclease [*Citrobacter* phage PhiZZ23]	0
gp_0030	16069	16587	reverse	172	Pyrophosphatase	Putative dCTP pyrophosphatase [*Escherichia* phage vB_EcoM_G2540-3]	1×10^{-123}
gp_0031	16589	17239	reverse	216	Endonuclease	Homing endonuclease [*Salmonella* phage GRNsp7]	1×10^{-153}
gp_0032	17310	17510	forward	66	Hypothetical protein	Hypothetical protein [*Shigella* phage ESh18]	5×10^{-36}
gp_0033	17507	18535	reverse	342	DNA primase	DNA primase [*Escherichia* phage slur07]	0
gp_0034	18538	18702	reverse	54	Hypothetical protein	Hypothetical protein Shfl2p037 [*Shigella* phage Shfl2]	4×10^{-28}
gp_0035	18704	19060	reverse	118	Hypothetical protein	RB32ORF033c hypothetical protein [*Escherichia* phage RB14]	2×10^{-79}
gp_0036	19062	19694	reverse	210	Hypothetical protein	Hypothetical protein KMC16_gp036 [*Escherichia* phage vB_EcoM_Lutter]	4×10^{-153}
gp_0037	19696	19989	reverse	97	Spackle periplasmic	Spackle periplasmic [*Escherichia* phage REP2]	2×10^{-65}
gp_0038	20050	20307	reverse	85	Hypothetical protein	Hypothetical protein ECML134_037 [*Escherichia* phage ECML-134]	1×10^{-54}
gp_0039	20309	20491	reverse	60	Discriminator of mRNA degradation	Dmd discriminator of mRNA degradation [*Escherichia* phage vB_EcoM_G50]	1×10^{-34}
gp_0040	20550	21977	reverse	475	Helicase	DNA primase/helicase [*Escherichia* phage vB_EcoM-101117BS1]	0
gp_0041	21987	22331	reverse	114	Head assembly protein	Head formation protein [*Escherichia* phage ECO07P2]	2×10^{-75}
gp_0042	22324	23505	reverse	393	Recombinase	Recombinase [Enterobacteria phage RB18]	0

gp_0043	23583	24425	reverse	280	Glucosyltransferase	Glucosyl transferase [*Escherichia* phage ES19]	0
gp_0044	24422	25162	reverse	246	Thymidylate synthase	Gp42 dCMP hydroxymethylase [*Escherichia* phage RB14]	0
gp_0045	25316	25567	reverse	83	Superinfection immunity protein	Immunity to superinfection [*Escherichia* phage vB_EcoM-UFV13]	2×10^{-49}
gp_0046	25575	25955	reverse	126	Hypothetical protein	Hypothetical protein [*Shigella* phage CT01]	3×10^{-88}
gp_0047	25966	26202	reverse	78	Hypothetical protein	Hypothetical protein QOTSP_274 [*Escherichia* phage vB_Eco_QOTSP]	5×10^{-47}
gp_0048	26383	29079	reverse	898	DNA polymerase	DNA polymerase [*Escherichia* phage slur02]	0
gp_0049	29158	29526	reverse	122	Translation repressor	Translation repressor [*Escherichia* phage RB3]	2×10^{-84}
gp_0050	29528	30091	reverse	187	DNA clamp loader	Putative clamp loader small subunit [*Escherichia* phage vB_EcoM_G2540-3]	3×10^{-134}
gp_0051	30093	31052	reverse	319	DNA polymerase (clamp loader)	Putative replication factor C small subunit [*Escherichia* phage JLBYU24]	0
gp_0052	31104	31790	reverse	228	DNA polymerase (sliding clamp)	Sliding clamp DNA polymerase accessory protein [*Shigella* phage ESh16]	1×10^{-164}
gp_0053	31846	32235	reverse	129	RNA polymerase	RNA polymerase binding [*Escherichia* phage slur02]	8×10^{-91}
gp_0054	32245	32433	reverse	62	Hypothetical protein	Protein GP45.2 [*Escherichia* phage vB_EcoM_ACG-C40]	1×10^{-37}
gp_0055	32488	34128	reverse	546	Endonuclease	SbcC-like subunit of palindrome specific endonuclease [*Escherichia* phage slur02]	0
gp_0056	34167	34373	reverse	68	Hypothetical protein	Hypothetical protein bas37_0168 [*Escherichia* phage karlgjung]	1×10^{-40}
gp_0057	34354	34617	reverse	87	Hypothetical protein	Hypothetical protein KMC09_gp057 [*Escherichia* phage vB_EcoM_G50]	1×10^{-55}
gp_0058	34614	35633	reverse	339	Nuclease	SbcD-like subunit of palindrome specific endonuclease [*Citrobacter* phage vB_CroM_CrRp10]	0
gp_0059	35611	36234	reverse	207	Endonuclease	HNH endonuclease [*Escherichia* phage ECO4]	3×10^{-151}
gp_0060	36411	37613	reverse	400	Glucosyltransferase	Alpha-glucosyltransferase [*Escherichia* phage RB14]	0
gp_0061	37680	37853	reverse	57	Hypothetical protein	Hypothetical protein KMB92_gp062 [*Citrobacter* phage PhiZZ6]	3×10^{-33}
gp_0062	37857	38060	reverse	67	Hypothetical protein	Hypothetical protein KMC13_gp173 [*Escherichia* phage vB_EcoM_IME537]	4×10^{-42}
gp_0063	38029	38346	reverse	105	Hypothetical protein	Hypothetical protein [*Shigella* phage ESh26]	2×10^{-67}
gp_0064	38348	38491	reverse	47	Hypothetical protein	Gp75 [*Shigella* phage pss-1]	7×10^{-25}
gp_0065	38550	39107	reverse	185	RNA polymerase	RNA polymerase sigma factor [*Escherichia* phage vB_EcoM_112]	2×10^{-135}

gp_0066	39186	39455	reverse	89	Hypothetical protein	Hypothetical protein [*Shigella* phage ESH35]	2×10^{-58}
gp_0067	39452	39667	reverse	71	Hypothetical protein	Hypothetical protein AS348_gp070 [*Escherichia* phage slur14]	1×10^{-42}
gp_0068	39670	39996	reverse	108	Hypothetical protein	Hypothetical protein KMC33_gp190 [*Shigella* phage Sf23]	3×10^{-71}
gp_0069	40049	40249	reverse	66	Hypothetical protein	Hypothetical protein ecmi02_0068 [*Escherichia* phage Ec_MI-02]	6×10^{-41}
gp_0070	40250	40381	reverse	43	Hypothetical protein	Hypothetical protein JLBYU31_84 [*Escherichia* phage JLBYU31]	6×10^{-23}
gp_0071	40389	40682	reverse	97	Hypothetical protein	Hypothetical protein PHAGINATOR_72 [*Shigella* phage vB_SboM_Phaginator]	1×10^{-62}
gp_0072	40675	40851	reverse	58	Hypothetical protein	Hypothetical protein AVU04_gp177 [*Escherichia* phage slur02]	5×10^{-33}
gp_0073	41010	41333	reverse	107	Glutaredoxin	Glutaredoxin [*Escherichia* phage vB_EcoM_ASO78A]	1×10^{-71}
gp_0074	41305	41568	reverse	87	Hypothetical protein	RB69ORF082c hypothetical protein [*Escherichia* phage RB69]	5×10^{-57}
gp_0075	41619	41834	reverse	71	Hypothetical protein	Hypothetical protein e112_078 [*Escherichia* phage vB_EcoM_112]	4×10^{-41}
gp_0076	41844	41957	reverse	37	Hypothetical protein	Hypothetical protein bas44_0189 [*Escherichia* phage AdolfPortmann]	2×10^{-17}
gp_0077	41950	42420	reverse	156	Reductase (small subunit)	Anaerobic ribonucleotide reductase small subunit [*Escherichia* phage teqhad]	1×10^{-111}
gp_0078	42417	44234	reverse	605	Reductase (large subunit)	Anaerobic ribonucleoside reductase large subunit [*Escherichia* phage wv7]	0
gp_0079	44231	44704	reverse	157	Endonuclease	Endonuclease VII [*Escherichia* phage vB_EcoM_G9062]	1×10^{-111}
gp_0080	44746	45231	reverse	161	Peptidase (inhibitor)	Protease inhibitor [*Escherichia* phage ES19]	4×10^{-113}
gp_0081	45215	45370	reverse	51	Hypothetical protein	Gp49.1 conserved protein of unknown function [*Escherichia* phage T4]	3×10^{-28}
gp_0082	45355	45675	reverse	106	Hypothetical protein	Ribonucleotide reductase [*Escherichia* phage REP2]	2×10^{-72}
gp_0083	45687	45821	reverse	44	Hypothetical protein	Hypothetical protein F26_0087 [*Escherichia* phage vB_Eco_F26]	3×10^{-21}
gp_0084	45860	46075	reverse	71	Hypothetical protein	Hypothetical protein PLBFAGBN_00105 [*Escherichia* phage Killian]	3×10^{-43}
gp_0085	46072	46335	reverse	87	Glutaredoxin	Putative glutaredoxin [*Escherichia* phage vB_EcoM_G10400]	1×10^{-57}
gp_0086	46337	46579	reverse	80	Hypothetical protein	Hypothetical protein bas45_0199 [*Escherichia* phage paulhmueller]	3×10^{-49}
gp_0087	46566	46883	reverse	105	Hypothetical protein	Hypothetical protein Shfl2p088 [*Shigella* phage Shfl2]	4×10^{-72}
gp_0088	46880	47809	reverse	309	Hypothetical protein	Hypothetical protein Shfl2p089 [*Shigella* phage Shfl2]	0

gp_0089	47862	48239	reverse	125	Hypothetical protein	Hypothetical protein ECTP7_01235 [*Escherichia* coli O157 typing phage 7]	8×10^{-83}
gp_0090	48269	48862	reverse	197	Hypothetical protein	Hypothetical protein AVU04_gp192 [*Escherichia* phage slur02]	8×10^{-126}
gp_0091	48920	49945	reverse	341	Hypothetical protein	Hypothetical protein AVU04_gp193 [*Escherichia* phage slur02]	0
gp_0092	49954	50844	reverse	296	Hypothetical protein	Thioredoxin [*Escherichia* phage vB_EcoM_CE1]	0
gp_0093	50852	51259	reverse	135	Hypothetical protein	Thioredoxin [*Citrobacter* phage PhiZZ6]	5×10^{-91}
gp_0094	51315	51842	reverse	175	Hypothetical protein	Thioredoxin [*Escherichia* phage vB_EcoM-UFV09]	2×10^{-120}
gp_0095	51903	52217	reverse	104	Hypothetical protein	Putative thioredoxin [*Escherichia* phage JLBYU24]	7×10^{-71}
gp_0096	52308	53276	reverse	322	Hypothetical protein	Hypothetical protein D862_gp183 [*Escherichia* phage vB_EcoM_ACG-C40]	0
gp_0097	53346	53609	reverse	87	Hypothetical protein	DUF4031 domain-containing protein [*Escherichia* phage ECML-134]	4×10^{-57}
gp_0098	53606	53755	reverse	49	Hypothetical protein	Hypothetical protein KMC13_gp210 [*Escherichia* phage vB_EcoM_IME537]	1×10^{-25}
gp_0099	53872	54882	reverse	336	Hypothetical protein	Thioredoxin [*Yersinia* phage pyps55t]	0
gp_0100	54882	55343	reverse	153	Hypothetical protein	Hypothetical protein D862_gp178 [*Escherichia* phage vB_EcoM_ACG-C40]	7×10^{-105}
gp_0101	55346	55867	reverse	173	Hypothetical protein	Hypothetical protein G10400_00098 [*Escherichia* phage vB_EcoM_G10400]	2×10^{-122}
gp_0102	55874	56407	reverse	177	Hypothetical protein	Hypothetical protein SP1_0175 [*Escherichia* phage vB_EcoM_SP1]	9×10^{-121}
gp_0103	56409	56675	reverse	88	Hypothetical protein	Hypothetical protein BH804_gp006 [*Shigella* phage SHFML-11]	2×10^{-53}
gp_0104	56677	56859	reverse	60	Hypothetical protein	Hypothetical protein bas46_0198 [*Escherichia* phage ChristianSchoenbein]	3×10^{-30}
gp_0105	57024	57197	reverse	57	Hypothetical protein	Hypothetical protein bas41_0219 [*Escherichia* phage FriedrichZschokke]	4×10^{-31}
gp_0106	57187	57381	reverse	64	Hypothetical protein	Hypothetical protein RB27_105 [Enterobacteria phage RB27]	2×10^{-37}
gp_0107	57384	57587	reverse	67	Hypothetical protein	Molybdopterin-guanine dinucleotide biosynthesis protein MobD [*Escherichia* phage teqskov]	1×10^{-38}
gp_0108	57587	57775	reverse	62	Hypothetical protein	Hypothetical protein KMC13_gp220 [*Escherichia* phage vB_EcoM_IME537]	2×10^{-36}
gp_0109	57871	58257	reverse	128	Hypothetical protein	Hypothetical protein D5505_00103 [*Escherichia* phage D5505]	1×10^{-86}
gp_0110	58254	58547	reverse	97	Antiholin	Lysis inhibition [*Shigella* phage Shfl2]	1×10^{-65}
gp_0111	58560	58772	reverse	70	Hypothetical protein	Hypothetical protein F25_0105 [*Escherichia* phage vB_Eco_F25]	2×10^{-42}

gp_0112	58815	59396	reverse	193	Thymidine kinase	Thymidine kinase [*Escherichia* phage wv7]	6×10^{-141}
gp_0113	59406	59591	reverse	61	Hypothetical protein	Hypothetical protein bas40_0219 [*Escherichia* phage FelixPlatter]	2×10^{-35}
gp_0114	59578	59760	reverse	60	Hypothetical protein	Hypothetical protein [*Shigella* phage ESh29]	3×10^{-31}
gp_0115	59757	59963	reverse	68	Hypothetical protein	Hypothetical protein KNU25_gp138 [*Escherichia* phage MLF4]	2×10^{-43}
gp_0116	59960	60172	reverse	70	Hypothetical protein	Hypothetical protein KMC13_gp228 [*Escherichia* phage vB_EcoM_IME537]	5×10^{-45}
gp_0117	60169	60636	reverse	155	Macrodomain	Macro domain protein [*Shigella* phage ESh31]	9×10^{-111}
gp_0118	60633	60974	reverse	113	TRNA ligase modifier	Valyl tRNA synthetase modifier [*Escherichia* phage vB_EcoM_G8]	3×10^{-75}
gp_0119	60967	61512	reverse	181	Hypothetical protein	Endoribonuclease [*Shigella* phage JK23]	1×10^{-128}
gp_0120	61520	61981	reverse	153	Endoribonuclease	Endoribonuclease [Enterobacteria phage Kha5h]	2×10^{-109}
gp_0121	62041	62319	reverse	92	Hypothetical protein	Hypothetical protein F412_gp154 [*Escherichia* phage wv7]	1×10^{-58}
gp_0122	62319	62585	reverse	88	Hypothetical protein	Hypothetical protein PHAGINATOR_117 [*Shigella* phage vB_SboM_Phaginator]	4×10^{-57}
gp_0123	62578	62799	reverse	73	Hypothetical protein	Hypothetical protein BI057_gp215 [*Shigella* phage SHFML-26]	1×10^{-45}
gp_0124	62799	63161	reverse	120	Autonomous glycyl radical cofactor	Autonomous glycyl radical cofactor [*Escherichia* phage ph0021]	1×10^{-81}
gp_0125	63169	63549	reverse	126	Hypothetical protein	Hypothetical protein KMC13_gp237 [*Escherichia* phage vB_EcoM_IME537]	1×10^{-87}
gp_0126	63495	64034	reverse	179	Hypothetical protein	Hypothetical protein UAB60_gp123 [*Salmonella* phage UAB_60]	6×10^{-130}
gp_0127	64187	64660	reverse	157	Hypothetical protein	Internal virion protein [*Escherichia* phage HY01]	9×10^{-110}
gp_0128	64670	65086	reverse	138	Endonuclease	Putative pyrimidine dimer DNA glycosylase [*Escherichia* phage vB_EcoM_G10400]	9×10^{-97}
gp_0129	65146	65640	reverse	164	Lysozyme	MAG: lysozyme [Bacteriophage sp.]	1×10^{-117}
gp_0130	65677	66117	reverse	146	Hydrolase	Nudix hydrolase [*Escherichia* phage F2]	2×10^{-106}
gp_0131	66114	66476	reverse	120	Hypothetical protein	Phage protein [*Escherichia* phage T4_ev240]	3×10^{-82}
gp_0132	66458	66850	reverse	130	Hypothetical protein	Hypothetical protein KMC31_gp122 [*Shigella* phage CM8]	2×10^{-87}
gp_0133	66819	67421	reverse	200	Hypothetical protein	Phage protein [*Escherichia* phage T4_ev240]	1×10^{-143}
gp_0134	67469	68062	reverse	197	Hypothetical protein	Hypothetical protein RB3_133 [*Escherichia* phage RB3]	5×10^{-134}

gp_0135	68106	68282	reverse	58	Hypothetical protein	Hypothetical protein BN81_138 [*Yersinia* phage phiD1]	4×10^{-33}
gp_0136	68351	68614	reverse	87	Hypothetical protein	E.8 conserved hypothetical protein [*Escherichia* phage T4]	2×10^{-57}
gp_0137	68856	69329	reverse	157	Hypothetical protein	Hypothetical protein e112_137 [*Escherichia* phage vB_EcoM_112]	8×10^{-109}
gp_0138	69717	70067	reverse	116	Hypothetical protein	Hypothetical protein FDH36_gp115 [*Escherichia* phage HP3]	5×10^{-79}
gp_0139	70598	71248	reverse	216	Endonuclease	Homing endonuclease [*Escherichia* phage AR1]	9×10^{-155}
gp_0140	71590	71877	reverse	95	Hypothetical protein	Hypothetical protein BI016_gp003 [*Escherichia* phage HY03]	3×10^{-62}
gp_0141	71880	72260	reverse	126	Hypothetical protein	Hypothetical protein BI016_gp004 [*Escherichia* phage HY03]	7×10^{-88}
gp_0142	72262	72447	reverse	61	Hypothetical protein	Hypothetical protein ACQ54_gp131 [*Escherichia* phage HY01]	4×10^{-33}
gp_0143	72506	72739	reverse	77	Internal protein	Putative internal head protein [*Escherichia* phage JLBYU31]	2×10^{-48}
gp_0144	72812	73270	reverse	152	Hypothetical protein	RNA ligase [*Shigella* phage CM8]	7×10^{-108}
gp_0145	73267	73509	reverse	80	Tail fiber assembly protein	Tail fiber chaperone [*Escherichia* phage T4]	1×10^{-44}
gp_0146	73509	74234	reverse	241	Deoxynucleoside monophosphate kinase	Putative deoxynucleotide monophosphate kinase [*Escherichia* phage 132]	5×10^{-178}
gp_0147	74284	74814	reverse	176	Tail completion protein	Glycoprotein 3 [*Escherichia* phage T4]	2×10^{-126}
gp_0148	74921	75745	reverse	274	DNA end protector protein	DNA end protector [*Escherichia* phage HY01]	0
gp_0149	75745	76197	reverse	150	Head completion protein	Head closure [*Salmonella* phage pse_SNUABM_01]	8×10^{-106}
gp_0150	76245	76835	forward	196	Baseplate wedge subunit	Baseplate wedge subunit [*Citrobacter* phage PhiZZ6]	2×10^{-141}
gp_0151	76819	78546	forward	575	Tail-associated lysozyme	Baseplate hub subunit and tail lysozyme [*Escherichia* phage slur07]	0
gp_0152	78539	79075	forward	178	Hypothetical protein	Hypothetical protein KMC04_gp155 [*Escherichia* phage vB_EcoM_F1]	9×10^{-125}
gp_0153	79076	79369	forward	97	Proline-alanine-alanine-arginine (PAAR) domain	PAAR motif of membran proteins [*Escherichia* phage slur07]	6×10^{-64}
gp_0154	79378	81360	forward	660	Baseplate wedge subunit	Baseplate wedge subunit [*Escherichia* phage vB_EcoM_KAW1E185]	0
gp_0155	81417	84455	forward	1012	Baseplate wedge subunit	Baseplate wedge initiator [*Escherichia* phage vB_EcoM_SA21RB]	0
gp_0156	84448	85452	forward	334	Baseplate wedge subunit	Baseplate wedge subunit [*Escherichia* phage vB_EcoM_G8]	0
gp_0157	85516	86382	forward	288	Baseplate wedge tail fiber connector	Baseplate wedge tail fiber connector [*Shigella* phage ESh28]	0

gp_0158	86382	88190	forward	602	Baseplate wedge subunit	Baseplate wedge subunit [*Escherichia* phage slur14]	0
gp_0159	88190	88849	forward	219	Baseplate wedge subunit	Baseplate wedge subunit [*Yersinia* phage PST]	1×10^{-158}
gp_0160	88846	90429	forward	527	Tail fiber protein	Tail collar fiber protein [*Escherichia* phage vB_EcoM_G8]	0
gp_0161	90426	91889	forward	487	Neck protein fibritin	Fibritin neck whisker [Enterobacteria phage GiZh]	0
gp_0162	91921	92850	forward	309	Neck protein	Neck protein [*Escherichia* phage vB_EcoM-S1P5QW]	0
gp_0163	92852	93622	forward	256	Neck protein	Head closure Hc2 [*Escherichia* phage vB_EcoM_112]	0
gp_0164	93664	94482	forward	272	Tail sheath stabilizer	Tail sheath stabilizer [*Escherichia* phage HY03]	0
gp_0165	94496	94990	forward	164	Terminase (small subunit)	Terminase small subunit [*Shigella* phage Sf24]	1×10^{-116}
gp_0166	94974	96806	forward	610	Terminase (large subunit)	Terminase large subunit [*Yersinia* phage phiD1]	0
gp_0167	96838	98817	forward	659	Tail sheath protein	Putative tail sheath protein [*Escherichia* phage vB_EcoM_G2133]	0
gp_0168	98934	99425	forward	163	Tail tube protein	Tail protein [*Escherichia* phage vB_VpM_PD112]	8×10^{-117}
gp_0169	99509	101083	forward	524	Portal vertex (capsid assembly)	Portal protein [*Escherichia* phage FelixPlatter]	0
gp_0170	101083	101322	forward	79	Prohead core (scaffold) protein	Prohead [*Citrobacter* phage PhiZZ23]	1×10^{-41}
gp_0171	101322	101747	forward	141	Prohead core (scaffold) protein	Head scaffolding protein [*Escherichia* phage UFV-AREG1]	2×10^{-95}
gp_0172	101747	102385	forward	212	Prohead core protein protease	Head maturation protease [*Escherichia* phage T4]	1×10^{-151}
gp_0173	102416	103225	forward	269	Prohead core (scaffold) protein	Prohead assembly (scaffolding) protein [*Escherichia* phage vec20]	0
gp_0174	103244	104794	forward	516	Capsid protein (major)	Major head protein [*Citrobacter* phage PhiZZ6]	0
gp_0175	104878	106161	forward	427	Capsid protein (vertex)	Capsid vertex protein [*Escherichia* phage slur07]	0
gp_0176	106191	107195	reverse	334	RNA ligase	RNA ligase [*Yersinia* phage PST]	0
gp_0177	107205	107483	reverse	92	Hypothetical protein	Hypothetical protein KMB92_gp233 [*Citrobacter* phage PhiZZ6]	9×10^{-60}
gp_0178	107470	107673	reverse	67	Hypothetical protein	DUF2774 domain-containing protein [*Escherichia* phage vB_EcoM_112]	4×10^{-39}
gp_0179	107776	108900	reverse	374	Immunogenic outer capsid protein	Capsid decoration protein [*Shigella* phage JK38]	0
gp_0180	108910	109590	reverse	226	Peptidase inhibitor activity	Inhibitor of prohead protease [*Shigella* phage JK38]	2×10^{-161}

gp_0181	109641	111152	forward	503	Helicase	DNA helicase [*Yersinia* phage vB_YepM_ZN18]	0
gp_0182	111149	111841	forward	230	Endonuclease	Homing endonuclease [*Yersinia* phage PYPS2T]	5×10^{-170}
gp_0183	111867	112097	forward	76	Helicase	DNA helicase [*Escherichia* phage vB_EcoM_112]	1×10^{-44}
gp_0184	112153	112320	reverse	55	Hypothetical protein	DUF2685 domain-containing protein [*Escherichia* phage T4]	1×10^{-30}
gp_0185	112349	112573	reverse	74	Hypothetical protein	Hypothetical protein ASO78A_180 [*Escherichia* phage vB_EcoM_ASO78A]	1×10^{-46}
gp_0186	112573	112986	reverse	137	Recombination, repair and ssDNA binding protein	UvsY-like recombination mediator [*Escherichia* phage slur14]	3×10^{-93}
gp_0187	113053	113451	reverse	132	Baseplate wedge subunit (lysozyme activity)	Baseplate wedge subunit [*Shigella* phage ESh36]	1×10^{-89}
gp_0188	113451	114077	reverse	208	Baseplate hub protein	Hypothetical protein [*Escherichia* phage ZCEC14]	4×10^{-150}
gp_0189	114128	114877	forward	249	Baseplate hub assembly protein	Baseplate hub assembly protein [Serratia phage PhiZZ30]	1×10^{-180}
gp_0190	114877	116052	forward	391	Baseplate hub subunit	Baseplate hub protein [*Escherichia* phage vB_EcoM_SCS4]	0
gp_0191	115997	116530	forward	177	Baseplate hub distal subunit	Baseplate hub distal subunit [Enterobacteria phage RB27]	4×10^{-124}
gp_0192	116527	118299	forward	590	Tape measure protein	Hypothetical protein R5505_00209 [*Escherichia* phage vB_EcoM_R5505]	0
gp_0193	118308	119402	forward	364	Baseplate subunit (Tail-tube)	Baseplate tail tube cap [*Shigella* phage SHFML-11]	0
gp_0194	119402	120367	forward	321	Baseplate subunit (Tail-tube)	Tail tube [Enterobacteria phage Kha5h]	0
gp_0195	120396	120686	reverse	96	Hypothetical protein	Hypothetical protein KMC25_gp213 [*Escherichia* phage teqhad]	7×10^{-60}
gp_0196	120747	122804	reverse	685	ADP-ribosyltransferase	Alt-like RNA polymerase ADP-ribosyltransferase [*Escherichia* phage slur02]	0
gp_0197	122808	124862	reverse	684	ADP-ribosyltransferase	Alt-like RNA polymerase ADP-ribosyltransferase [*Shigella* phage SH7]	0
gp_0198	124915	125103	reverse	62	Hypothetical protein	Hypothetical protein KMC09_gp210 [*Escherichia* phage vB_EcoM_G50]	2×10^{-36}
gp_0199	125100	126563	reverse	487	DNA ligase	DNA ligase [Enterobacteria phage T6]	0
gp_0200	126563	126829	reverse	88	Hypothetical protein	DUF3045 domain-containing protein [*Escherichia* phage vB_EcoM-G28]	2×10^{-59}
gp_0201	126829	127665	reverse	278	Hypothetical protein	Hypothetical protein AVU02_gp059 [*Escherichia* phage slur07]	0
gp_0202	127662	128042	reverse	126	Hypothetical protein	Hypothetical protein ECML134_199 [*Escherichia* phage ECML-134]	6×10^{-87}
gp_0203	128113	128319	reverse	68	Hypothetical protein	Hypothetical protein bas44_0039 [*Escherichia* phage AdolfPortmann]	7×10^{-42}

gp_0204	128316	128513	reverse	65	Hypothetical protein	Hypothetical protein D862_gp071 [*Escherichia* phage vB_EcoM_ACG-C40]	2×10^{-37}
gp_0205	128513	128800	reverse	95	Hypothetical protein	Hypothetical protein BI016_gp261 [*Escherichia* phage HY03]	4×10^{-64}
gp_0206	128842	129207	reverse	121	Hypothetical protein	Head vertex protein [Enterobacteria phage T6]	9×10^{-85}
gp_0207	129275	129607	reverse	110	Hypothetical protein	Hypothetical protein AVU04_gp041 [*Escherichia* phage slur02]	5×10^{-74}
gp_0208	129718	129936	reverse	72	Hypothetical protein	Hypothetical protein ACQ28_gp199 [*Yersinia* phage PST]	3×10^{-42}
gp_0209	130140	130388	reverse	82	Lysis inhibition accessory protein	MAG: hypothetical protein [Bacteriophage sp.]	4×10^{-51}
gp_0210	130536	130871	reverse	111	Co-chaperonin	Head morphogenesis [Serratia phage PhiZZ30]	$\times10^{-72}$
gp_0211	130928	131236	reverse	102	Hypothetical protein	SH3 beta-barrel fold-containing protein [*Shigella* phage Shfl2]	2×10^{-66}
gp_0212	131237	131473	reverse	78	Hypothetical protein	Hypothetical protein e112_225 [*Escherichia* phage vB_EcoM_112]	6×10^{-49}
gp_0213	131473	132054	reverse	193	Deoxycytidylate deaminase	Deoxycytidylate deaminase [*Escherichia* phage MLP2]	2×10^{-140}
gp_0214	132051	132389	reverse	112	Hypothetical protein	Cd.1 hypothetical protein [*Escherichia* phage T4]	5×10^{-75}
gp_0215	132386	132622	reverse	78	Hypothetical protein	Hypothetical protein ACQ28_gp192 [*Yersinia* phage PST]	2×10^{-47}
gp_0216	132616	133143	reverse	175	Hypothetical protein	Hypothetical protein KMC14_gp197 [*Escherichia* phage vB_EcoM_KAW1E185]	8×10^{-123}
gp_0217	133206	133481	reverse	91	Hypothetical protein	Cd.3 conserved hypothetical protein [*Escherichia* phage T4]	2×10^{-58}
gp_0218	133484	133684	reverse	66	Hypothetical protein	Hypothetical protein KMC11_gp193 [*Escherichia* phage vB_EcoM_G4507]	1×10^{-38}
gp_0219	133677	133874	reverse	65	Hypothetical protein	Hypothetical protein KMC13_gp059 [*Escherichia* phage vB_EcoM_IME537]	3×10^{-40}
gp_0220	133874	134782	reverse	302	Polynucleotide kinase	Polynucleotide kinase [*Shigella* phage A2]	0
gp_0221	134779	135099	reverse	106	Hypothetical protein	Hypothetical protein D862_gp055 [*Escherichia* phage vB_EcoM_ACG-C40]	2×10^{-71}
gp_0222	135103	135327	reverse	74	Hypothetical protein	Hypothetical protein Shfl2p224 [*Shigella* phage Shfl2]	3×10^{-47}
gp_0223	135324	135623	reverse	99	Outer lipoprotein subunit (Spanin)	Rz-like spanin [*Escherichia* phage vB_EcoM_KAW1E185]	2×10^{-66}
gp_0224	135620	135973	reverse	117	Inner membrane subunit (Spanin)	Rz-like spanin [*Shigella* phage Shfl2]	7×10^{-76}
gp_0225	135964	136467	reverse	167	ALP protein	Inhibitor of host transcription [*Shigella* phage pss-1]	1×10^{-120}
gp_0226	136532	137656	reverse	374	RNA ligase	RNA ligase and tail fiber protein attachment catalyst [*Escherichia* phage teqhad]	0

gp_0227	137709	138119	reverse	136	Endonuclease	Endonuclease [Enterobacteria phage Aplg8]	1×10^{-94}
gp_0228	138147	139325	reverse	392	Ribonucleotide reductase (beta subunit)	Ribonucleotide reductase class Ia beta subunit [*Shigella* phage CM8]	0
gp_0229	139377	141641	reverse	754	Ribonucleotide reductase (alpha subunit)	NrdA-like aerobic NDP reductase large subunit [*Escherichia* phage slur07]	0
gp_0230	141912	142175	reverse	87	Hypothetical protein	Putative prohead core scaffold protein [*Escherichia* phage JLBYU24]	1×10^{-56}
gp_0231	142172	143032	reverse	286	Thymidylate synthase	Thymidylate synthase [*Escherichia* phage slur02]	0
gp_0232	143078	143425	reverse	115	Hypothetical protein	Putative dihydrofolate reductase [*Escherichia* phage JLBYU22]	2×10^{-79}
gp_0233	143446	144027	reverse	193	Dihydrofolate reductase	Dihydrofolate reductase [*Escherichia* phage vB_EcoM_NBG2]	9×10^{-139}
gp_0234	144027	144272	reverse	81	Hypothetical protein	Hypothetical protein KMC36_gp074 [*Yersinia* phage PYPS2T]	5×10^{-51}
gp_0235	144281	144616	reverse	111	Hypothetical protein	DNA adenine methyltransferase [*Escherichia* phage Killian]	2×10^{-74}
gp_0236	144627	144869	reverse	80	Hypothetical protein	Hypothetical protein EO1_66 [*Escherichia* phage EO1]	4×10^{-51}
gp_0237	144924	145289	reverse	121	Hypothetical protein	Hypothetical protein KMC27_gp099 [*Escherichia* phage YUEEL01]	5×10^{-81}
gp_0238	145334	145564	reverse	76	Hypothetical protein	Putative ssDNA binding protein [*Shigella* phage vB_SboM_Phaginator]	8×10^{-48}
gp_0239	145709	146617	reverse	302	SsDNA binding protein	Single strand DNA binding protein [*Escherichia* phage vB_EcoM_KAW1E185]	0
gp_0240	146717	147370	reverse	217	Helicase loading protein	DNA helicase loader [*Escherichia* phage vB_EcoM_OE5505]	8×10^{-154}
gp_0241	147367	147705	reverse	112	RNA polymerase	Chain K, RNA polymerase-associated protein Gp33 [*Tequatrovirus* T4]	4×10^{-74}
gp_0242	147683	147952	reverse	89	Double-stranded DNA binding protein	Putative double-stranded DNA-binding protein [*Escherichia* phage U115]	4×10^{-55}
gp_0243	147961	148878	reverse	305	Ribonuclease	Rnase H [*Escherichia* phage vB_EcoM_IME537]	0
gp_0244	148983	152855	forward	1290	Tail fiber (proximal subunit)	Tail fiber protein proximal subunit [*Escherichia* phage T2]	0
gp_0245	152864	153979	forward	371	Tail fiber protein	Hypothetical protein [*Escherichia* phage N2]	0
gp_0246	154044	154709	forward	221	Tail fiber protein	Hinge connector of long tail fiber protein distal connector [*Escherichia* phage T4]	1×10^{-151}
gp_0247	154718	157792	forward	1024	Tail fiber protein	Tail fiber [*Salmonella* phage GRNsp7]	0
gp_0248	157820	158371	forward	183	Tail fiber assembly protein	Tail fiber assembly [*Escherichia* phage T4]	5×10^{-126}
gp_0249	158393	159049	forward	218	Holin	Recname: Full=Holin; altname: Full=Lysis protein [Enterobacteria phage K3]	2×10^{-160}

gp_0250	159050	159322	reverse	90	Transcriptional inhibitor (Anti-Sigma)	Chain I, kda anti-sigma factor [*Tequatrovirus* T4]	3×10^{-56}
gp_0251	159335	159487	reverse	50	Hypothetical protein	Asia.1 hypothetical protein [Enterobacteria phage RB51]	5×10^{-26}
gp_0252	159484	159762	reverse	92	Anti-restriction endonuclease	Inhibitor of mrcbc restriction [Enterobacteria phage Kha5h]	9×10^{-58}
gp_0253	159752	159871	reverse	39	Hypothetical protein	RB32ORF249c hypothetical protein [Enterobacteria phage RB51]	3×10^{-17}
gp_0254	159846	159977	reverse	43	Hypothetical protein	Hypothetical protein [*Escherichia* phage UTI-E4]	3×10^{-23}
gp_0255	160048	160344	reverse	98	Hypothetical protein	Anti-restriction nuclease [*Escherichia* phage vB_Eco_F26]	2×10^{-65}
gp_0256	160344	160805	reverse	153	Hypothetical protein	Hypothetical protein JS09_0192 [*Escherichia* phage vB_EcoM_JS09]	2×10^{-109}
gp_0257	160802	161131	reverse	109	Hypothetical protein	Hypothetical protein ISAACDANIEL_262 [Hafnia phage vB_HpaM_IsaacDaniel]	2×10^{-75}
gp_0258	161142	161777	reverse	211	Transcription factor mota	Activator middle promoter [*Escherichia* phage vB_EcoM_JB75]	1×10^{-149}
gp_0259	161904	162053	reverse	49	Hypothetical protein	Hypothetical protein G2540_00269 [*Escherichia* phage vB_EcoM_G2540]	6×10^{-22}
gp_0260	162050	163378	reverse	442	DNA topoisomerase	DNA topoisomerase II [*Shigella* phage Sf21]	0
gp_0261	163516	163674	reverse	52	Hypothetical protein	Acridine resistance protein [*Escherichia* phage vB_EcoM_SA21RB]	5×10^{-27}
gp_0262	163762	164220	reverse	152	Nuclear disruption protein	Ndd-like nucleoid disruption protein [*Escherichia* phage RB32]	7×10^{-110}
gp_0263	164281	164496	reverse	71	Hypothetical protein	Hypothetical protein AVU04_gp099 [*Escherichia* phage slur02]	7×10^{-44}
gp_0264	164505	164630	reverse	41	Hypothetical protein	Putative outer membrane protein [Enterobacteria phage Aplg8]	3×10^{-20}
gp_0265	164612	164809	reverse	65	Hypothetical protein	Periplasmic protein [*Escherichia* phage vB_EcoM_IME537]	5×10^{-37}
gp_0266	164914	165027	reverse	37	Hypothetical protein	Hypothetical protein Shfl2p273 [*Shigella* phage Shfl2]	5×10^{-17}
gp_0267	164817	164930	reverse	37	Hypothetical protein	Hypothetical protein e112_284 [*Escherichia* phage vB_EcoM_112]	2×10^{-14}
gp_0268	165174	165437	reverse	87	Hypothetical protein	Hypothetical protein AS348_gp143 [*Escherichia* phage slur14]	1×10^{-57}
gp_0269	165511	166068	reverse	185	Endonuclease	DenB-like DNA endonuclease IV [*Citrobacter* phage vB_CroM_crrp10]	3×10^{-136}
gp_0270	166001	166330	reverse	109	Hypothetical protein	Hypothetical protein D862_gp003 [*Escherichia* phage vB_EcoM_ACG-C40]	6×10^{-73}
gp_0271	166369	166563	reverse	64	Hypothetical protein	Hypothetical protein AVU04_gp106 [*Escherichia* phage slur02]	2×10^{-36}
gp_0272	166592	167530	reverse	312	RIIB lysis inhibitor	RIIB lysis inhibitor [*Escherichia* phage slur02]	0

Annex 3

Table F – List of 28 ETEC strains utilized in prophage screening. 27 of the strains' genomic information were retrieved from GenBank, under the respective accession numbers, and EC43 DNA was extracted, isolated and sequenced during this project. All strains' DNA ran though Phaster or Phastest softwares to obtain the number of prophages incorporated in the bacterial DNA.

Strain name	Number of prophages			Accession numbers
	Intact	Incomplete	Questionable	
EC43	3	4	2	-
Escherichia coli ETEC H10407	10	4	0	FN649414.1
Escherichia coli strain FMU073332	4	8	5	CP017844.1
Escherichia coli UMNK88	10	4	2	NC_017641.1
Escherichia coli O139:H28 str. E24377A	6	1	2	NC_009801.1
Escherichia coli str. K-12 substr. MG1655	2	1	6	NC_000913.3
Escherichia coli str. K-12 substr. W3110	2	1	2	AP009048.1
Escherichia coli O169:H41 strain 2014EL-1345-2	3	6	2	CP024223.1
Escherichia coli strain ATCC 43886	7	6	2	CP024256.1
Escherichia coli O182:H21 strain D181	5	4	1	CP024252.1
Escherichia coli O128:H27 strain 90-9281	5	2	1	CP024243.1
Escherichia coli strain 90-9276	5	6	3	CP024299.1
Escherichia coli O114:H49 strain 90-9280	4	6	2	CP024240.1
Escherichia coli strain 90-9269	4	1	2	CP024661.1
Escherichia coli O15:H11 strain 90-9272	4	8	1	CP024239.1
Escherichia coli strain ATCC 43896	10	7	1	CP024278.1
Escherichia coli O6:H16 strain M9682-C1	4	4	4	CP024275.1
Escherichia coli O27:H7 strain B4103-1	1	7	1	CP024245.1
Escherichia coli O169:H41 strain F6326-C1	2	4	2	CP024263.1
Escherichia coli strain F5176C6	4	6	2	CP024667.1
Escherichia coli O25:NM strain 2014EL-1343-2	5	2	1	CP024228.1
Escherichia coli strain F5656C1	4	6	6	CP024260.1
Escherichia coli O6:H16 strain 2014EL-1346-6	4	5	6	CP024232.1
Escherichia coli strain F9792	2	5	1	CP024273.1
Escherichia coli O25:H16 strain F5505-C1	5	6	2	CP024257.1
Escherichia coli O6:H16 strain F6699	3	7	2	CP024266.1
Escherichia coli strain F8111-1SC3	1	3	2	CP024269.1
Escherichia coli strain 00-3279	8	9	3	CP024293.1

Table G – Functional annotation of ETEC strain H10407 (accession numbers on GenBank: FN649414.1). The CDSs sequences were ran through BlastP and HHpred softwares. The functional attribution of CDSs encoded proteins were based on a comparison with a close homolog, considering the E-value, where blank spaces represent irrelevant homolog (E-value ≥ 1×10^{-5}). In the table is also represented the start and stop site (bp) and the size (aa) of the sequences.

Prophage	CDS	Start (bp)	End (bp)	Size (aa)	Function	Closest homolog	E-value
1 (intact)	**001**	297402	298565	387	Integrase	Integrase [*Shigella* phage SfIV]	0
	002	298792	299137	114	Hypothetical protein	Hypothetical protein SfVp28 [Enterobacteria phage SfV]	4×10^{-43}
	003	299133	300012	292	Phosphoadenosine phosphosulfate reductase	Phosphoadenosine phosphosulfate reductase [Enterobacteria phage VT2phi_272]	0
	004	300002	300539	178	5'-deoxynucleotidase	Putative hydrolase of HD superfamily [*Escherichia* phage vB_EcoM-720R4]	2×10^{-128}
	005	300666	301491	274	Hypothetical protein	DUF2303 family protein [*Shigella* phage SfIV]	0
	006	301551	301914	120	Hypothetical protein	Protein YfdP [*Escherichia* phage phiSTEC1575-Stx2k]	3×10^{-84}
	007	302351	302597	81	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM-720R4]	2×10^{-53}
	008	303088	303736	215	Repressor protein	CI-like repressor [*Shigella* phage SfII]	3×10^{-158}
	009	303878	304139	86	Transcriptional regulator	Helix-turn-helix transcriptional regulator [*Shigella* phage SfII]	2×10^{-57}
	010	304131	304683	183	Regulatory protein	Putative protein YmfL [*Escherichia* phage vB_EcoM-689R6]	6×10^{-134}
	011	305027	305969	313	Replication protein	DNA replication protein O [*Escherichia* phage vB_EcoM-720R4]	0
	012	305965	306460	164	Transcriptional regulator	Putative protein YfdN [*Escherichia* phage vB_EcoM-689R6]	1×10^{-118}
	013	306459	307113	217	N-6-adenine-methyltransferase	N-6-adenine-methyltransferase [*Escherichia* phage Lys8385Vzw]	1×10^{-162}
	014	307109	307436	108	Transcriptional regulator	LexA-like protein [*Escherichia* phage vB_EcoS-569R4]	9×10^{-76}
	015	307432	307822	129	Crossover junction endodeoxyribonuclease	Holliday junction resolvase [*Escherichia* phage 1720a-02]	7×10^{-92}
	016	307841	308639	265	DNA-binding protein	Transcriptional regulator [*Shigella* phage SfII]	0
	017	308646	309636	329	Nuclease	HNH endonuclease [Enterobacteria phage mEp460]	0
	018	310047	311379	443	NTPase	KAP P-loop domain-containing protein [*Vibrio* phage VD1]	3×10^{-24}

	019	311675	312002	108	Holin	Holin/anti-holin [*Shigella* phage SfIv]	9×10^{-72}
	020	312005	312482	158	Lysozyme	Endolysin [*Escherichia* phage vB_EcoM-720r4]	2×10^{-113}
	021	312465	312858	130	Hypothetical protein	Putative Rz lytic protein [*Escherichia* phage vB_EcoM-720r4]	3×10^{-72}
	022	313098	313653	184	Hypothetical protein	Hypothetical protein BOW93_gp133 [*Salmonella* phage 118970_sal3]	1×10^{-133}
	023	313900	314179	92	Hypothetical protein		
	024	314179	314674	164	Terminase small subunit	Terminase small subunit [*Escherichia* phage vB_EcoM-689r6]	2×10^{-116}
	025	314670	316404	577	Terminase large subunit	Terminase large subunit [Enterobacteria phage SfV]	0
	026	316400	316562	53	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoS-733r5]	8×10^{-9}
	027	316551	317778	408	Portal protein	Portal protein [Enterobacteria phage SfV]	0
	028	317770	318373	200	Prohead protease	Head maturation protease [Enterobacteria phage SfV]	8×10^{-147}
	029	318383	319613	409	Capsid protein	Major head protein [Enterobacteria phage SfV]	0
1 (intact)	**030**	319691	320015	107	Head-tail connector protein	Head-tail adaptor Ad1 [Enterobacteria phage SfV]	1×10^{-71}
	031	320011	320422	136	Head closure protein	Head-tail adaptor [*Escherichia* phage vB_EcoM-720r4]	2×10^{-93}
	032	320396	320903	168	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM-720r4]	9×10^{-118}
	033	320899	321460	186	Hypothetical protein	Hypothetical protein acq56_gp11 [Enterobacteria phage SfI]	9×10^{-136}
	034	321468	321639	56	Hypothetical protein	DUF2635 domain-containing protein [Enterobacteria phage SfV]	7×10^{-34}
	035	321622	323119	498	Tail protein	Tail sheath [*Shigella* phage SfII]	0
	036	323118	323475	118	Tail protein	Tail tube protein [Enterobacteria phage SfV]	2×10^{-85}
	037	323474	323744	89	Tail assembly protein	Tail protein [*Shigella* phage SfII]	6×10^{-61}
	038	323885	325718	610	Tail tape measure protein	Tail length tape measure protein [*Shigella* phage SfII]	0
	039	325809	326340	176	Hypothetical protein	Hypothetical protein [Caudoviricetes sp.]	1×10^{-124}
	040	326401	327730	442	DNA circularization protein	DNA circularization [Enterobacteria phage SfV]	0
	041	327726	328806	359	Baseplate protein	Baseplate hub [*Shigella* phage SfIv]	0

	042	328805	329354	182	Baseplate assembly protein	Baseplate assembly protein [*Escherichia* phage vB_EcoM-689r6]	4×10^{-132}
	043	329353	329779	141	Baseplate protein	Baseplate wedge subunit [Enterobacteria phage SfV]	1×10^{-96}
	044	329765	330824	352	Baseplate protein	Baseplate protein [*Shigella* phage SfIv]	0
	045	330814	331399	194	Hypothetical protein	Baseplate wedge subunit [Enterobacteria phage SfV]	2×10^{-142}
	046	331402	332230	275	Tail protein	Side tail fiber protein [*Escherichia* phage vB_EcoM-689r6]	2×10^{-114}
	047	332226	332832	201	Tail fiber assembly protein	Tail fiber assembly [*Escherichia* phage p2]	8×10^{-130}
	048	332803	333244	146	Tail fiber assembly protein	Tail fiber assembly protein [*Escherichia* phage vB_EcoM-720r4]	3×10^{-101}
	049	333685	334240	184	DNA recombinase	DNA invertase [*Escherichia* phage vB_EcoM-689R6]	6×10^{-131}
	050	334297	335071	257	Hypothetical protein	Endonuclease [serratia phage vB_SmaM-Kodama]	1×10^{-63}
	051	335896	336640	247	Transcriptional regulator	AraC family transcriptional regulator [*Escherichia* phage ev099]	1×10^{-175}
1 (intact)	**052**	336681	336843	53	Transposase	Transposase [*Escherichia* phage ev099]	4×10^{-32}
	053	337340	338555	404	Integrase	Tyrosine-type recombinase/integrase [*Erwinia* phage phiET88]	5×10^{-138}
	054	339024	339138	37	Hypothetical protein	MAG TPA: hypothetical protein [Caudoviricetes sp.]	3×10^{-20}
	055	339409	339841	143	Hypothetical protein	MAG TPA: hypothetical protein [Caudoviricetes sp.]	1×10^{-104}
	056	339853	340687	277	Antirepressor protein	Anti-repressor [*Escherichia* phage TL-2011c]	4×10^{-28}
	057	340855	341923	355	Ash protein	Transcriptional regulator [Enterobacteria phage SfV]	7×10^{-20}
	058	341915	342110	64	Hypothetical protein	MAG TPA: hypothetical protein [Caudoviricetes sp.]	4×10^{-38}
	059	342106	342370	87	Hypothetical protein	MAG TPA: hypothetical protein [Caudoviricetes sp.]	8×10^{-59}
	060	342366	342588	73	Hypothetical protein	MAG TPA: hypothetical protein [Caudoviricetes sp.]	4×10^{-19}
	061	342580	343183	200	Hypothetical protein	Hypothetical protein [*Aeromonas* phage asxd-1]	2×10^{-34}
	062	343195	345544	782	DNA primase-helicase	DNA primase [*Alteromonas* phage ZP6]	9×10^{-47}
2 (intact)	**001**	862243	863314	356	Integrase	Integrase [*Escherichia* phage HK106]	0

	002	863561	863732	56	Hypothetical protein		
	003	864411	864561	49	Hypothetical protein	Putative membrane protein [*Escherichia* phage 434]	1×10^{-29}
	004	865091	865307	71	Hypothetical protein	Hypothetical protein FDH52_gp17 [*Escherichia* phage PA28]	4×10^{-47}
	005	865383	865497	37	Hypothetical protein	DUF1382 family protein [*Escherichia* phage HK630]	3×10^{-18}
	006	865726	866407	226	Recombinase	Putative exonuclease [*Escherichia* phage phi458]	5×10^{-170}
	007	866403	867189	261	Recombination protein	RecT-like ssDNA annealing protein [*Escherichia* phage sh2026stx1]	0
	008	867194	867491	98	Host-nuclease inhibitor	Gam-like host nuclease inhibitor [*Escherichia* phage 933W]	6×10^{-67}
	009	867566	867710	47	Host cell division inhibitory peptide	Kil protein for bacterial septation inhibition [Enterobacteria phage VT2-Sakai]	6×10^{-28}
	010	867915	868284	122	Hypothetical protein	Hypothetical protein ST933Wp18 [*Escherichia* phage 933W]	1×10^{-85}
	011	868479	868929	149	Hypothetical protein	Hypothetical protein GALLYPH_30 [*Escherichia* phage gally]	6×10^{-105}
	012	868988	869111	40	Antirestriction protein	Restriction alleviation ral [*Escherichia* phage Lambda]	4×10^{-23}
2 (intact)	**013**	869429	869984	184	Super-infection exclusion protein (SieB)	Super-infection exclusion protein B [Caudoviricetes sp.]	4×10^{-136}
	014	870000	870273	90	Hypothetical protein	Early gene regulation protein [*Escherichia* phage JEP4]	7×10^{-58}
	015	870284	870416	43	Hypothetical protein		
	016	870584	870944	119	Transcriptional regulator	LexA-like repressor [*Escherichia* phage Lambda h434 imm21]	9×10^{-83}
	017	871315	871501	61	Transcriptional regulator	Repressor [Enterobacteria phage CUS-3]	9×10^{-38}
	018	871922	872387	154	Replication protein	Replication initiation protein [*Escherichia* phage 434]	1×10^{-84}
	019	872380	872755	124	Replication protein	Replication initiation O-like [*Escherichia* phage ev243]	1×10^{-88}
	020	872751	873453	233	Replication protein	Replication initiation protein [*Escherichia* phage 434]	5×10^{-173}
	021	873449	873740	96	Ren protein	Ren-like exclusion protein [*Escherichia* phage 933W]	1×10^{-66}
	022	873813	874254	146	Recombination protein	Ninb/Orf homologous recombination mediator [*Escherichia* phage HK544]	2×10^{-104}
	023	874250	874778	175	N-6-adenine-methyltransferase	DNA methyltransferase [Enterobacteria phage Sf101]	2×10^{-129}
	024	874953	875295	113	Hypothetical protein	DUF2591 family protein [*Escherichia* phage ev207]	7×10^{-82}

2 (intact)	**025**	875501	875864	120	Crossover junction endodeoxyribonuclease	Holliday junction resolvase / crossover junction endodeoxyribonuclease RusA (ec [*Escherichia* phage 1H12]	2×10^{-86}
	026	875860	876001	46	Hypothetical protein	Ylcg family protein [*Escherichia* phage ev099]	2×10^{-26}
	027	876086	876470	127	Antitermination protein	Anti-termination protein q-like [*Escherichia* phage ev099]	6×10^{-91}
	028	876659	877817	385	Porin	Porin ompc [*Escherichia* phage ev207]	0
	029	878329	878545	71	Lysis protein	Holin [Enterobacteria phage mEp460]	1×10^{-44}
	030	878544	879042	165	Lysozyme	Endolysin r21 like protein [Caudoviricetes sp.]	4×10^{-117}
	031	879038	879503	154	Lysis protein	Endopeptidase Rz, partial [*Escherichia* phage vB_EcoS-689r7]	2×10^{-108}
	032	879640	881098	485	Potassium transporter	Trk potassium uptake system protein [*Escherichia* phage vB_EcoS-683r6]	0
	033	881235	882027	263	Chromosome partitioning system protein	Chromosome (plasmid) partitioning protein [*Escherichia* phage vB_EcoS-473r12]	0
	034	882019	882952	310	Hypothetical protein	Hypothetical protein [*Salmonella* phage SPF_0923]	0
	035	882929	883139	69	Hypothetical protein	Hypothetical protein [*Salmonella* phage SPF_0923]	7×10^{-42}
	036	883142	884237	364	Terminase small subunit	Terminase small subunit [*Escherichia* phage vB_EcoS-473r12]	0
	037	884217	885519	433	Terminase	Putative phage terminase [*Salmonella* phage SPF_0923]	0
	038	885521	886928	468	Hypothetical protein	DUF4055 domain-containing protein [*Escherichia* phage vB_EcoS-473r12]	0
	039	886911	888024	370	Head morphogenesis protein	NAD+-asparagine adp-ribosyltransferase [*Escherichia* phage vB_EcoS-606r9]	0
	040	888128	888893	254	Hypothetical protein	Hypothetical protein [*Salmonella* phage SPF_0923]	0
	041	888991	890131	379	Capsid protein	Coat protein [*Escherichia* phage vB_EcoS-606r9]	0
	042	890173	890350	58	Hypothetical protein	Hypothetical protein [*Salmonella* phage SPF_0923]	3×10^{-37}
	043	890353	890749	131	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoS-606r9]	5×10^{-92}
	044	890748	891132	127	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoS-473r12]	2×10^{-89}
	045	891132	891513	126	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoS-473r12]	1×10^{-88}
	046	891509	891902	130	Hypothetical protein	Hypothetical protein [*Salmonella* phage SPF_0923]	2×10^{-90}

2 (intact)	**047**	891928	892858	309	Tail tube protein	Hypothetical protein [*Salmonella* phage SPF_0923]	0
	048	892950	893400	149	Hypothetical protein	Hypothetical protein [*Salmonella* phage SPF_0923]	3×10^{-106}
	049	893576	893708	43	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoS-666r9]	7×10^{-25}
	050	893873	897107	1077	Tail tape measure protein	Tail length tape-measure protein 1 [*Salmonella* phage SPF_0923]	0
	051	897141	897438	98	Tail protein	Tail tip assembly protein M [*Escherichia* phage vB_EcoS-606r9]	7×10^{-70}
	052	897437	898136	232	Tail protein	Tail tip assembly protein I [*Escherichia* phage vB_EcoS-473r12]	2×10^{-174}
	053	898285	898885	199	Tail tip assembly protein	Tail assembly protein [*Salmonella* phage SPF_0923]	1×10^{-146}
	054	898881	899454	190	Tail assembly protein	Tail assembly protein [*Escherichia* phage 4A7]	2×10^{-129}
	055	899514	903012	1165	Host specificity protein	Tail fiber protein [*Escherichia* phage ev243]	0
	056	903082	903682	199	Outer membrane protein	Lom outer membrane protein [*Escherichia* phage ev243]	3×10^{-145}
	057	903682	903856	57	Hypothetical protein	Hypothetical protein PBV4795_ORF74 [Enterobacteria phage BP-4795]	1×10^{-33}
	058	904243	904981	245	Hypothetical protein	Hypothetical protein [*Escherichia* phage phi458]	2×10^{-29}
	059	905016	905151	44	Hypothetical protein		
	060	905111	907145	677	Tail fiber protein	Tail fiber protein [*Escherichia* phage 2H10]	0
	061	907144	907729	194	Tail fiber assembly protein	Tail fiber assembly [*Escherichia* phage ev099]	1×10^{-135}
	062	907802	909134	443	Diguanylate cyclase	Putative signal transduction protein [*Acinetobacter* phage MD-2021a]	2×10^{-26}
3 (intact)	**001**	1289847	1290966	372	Integrase	Integrase [*Escherichia* phage vB_EcoS-640r1]	0
	002	1290934	1291204	89	Excisionase	Excisionase [*Escherichia* phage vB_EcoS-640r1]	8×10^{-61}
	003	1291265	1293737	823	Exonuclease	Exonuclease viii, ds DNA exonuclease [*Escherichia* phage Tritos]	0
	004	1293830	1294022	63	Hypothetical protein	Uncharacterized protein ydfD [*Escherichia* phage Tritos]	9×10^{-41}
	005	1294018	1294207	62	Cell division inhibition protein	Division inhibition protein dicB [*Escherichia* phage mEp460_ev081]	6×10^{-39}
	006	1294235	1294406	56	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoS-689r5]	5×10^{-20}

3 (intact)	**007**	1294607	1295045	145	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM-606r7]	4×10^{-27}
	008	1295013	1295343	109	Hypothetical protein	Hypothetical protein [Caudoviricetes sp.]	3×10^{-71}
	009	1295354	1295588	77	Hypothetical protein	Head-DNA stabilization protein d [*Escherichia* phage vB_EcoS-689r5]	7×10^{-47}
	010	1295799	1296399	199	Repressor protein	Repressor protein CI [*Escherichia* phage vB_EcoS-640R1]	4×10^{-142}
	011	1296529	1296772	80	DNA transcriptional regulator	Protein DicC [*Escherichia* phage vB_EcoS-640R1]	5×10^{-53}
	012	1296755	1297181	141	YdaT toxin domain-containing protein	Activator protein CII [*Escherichia* phage vB_EcoS-640R1]	6×10^{-101}
	013	1297252	1298323	356	Primosomal protein	DNA replication protein o [*Escherichia* phage vB_EcoS-640r1]	0
	014	1298363	1298786	140	Hypothetical protein	LygF [*Escherichia* phage vB_EcoS-640r1]	4×10^{-99}
	015	1298843	1299200	118	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoS-640r1]	1×10^{-81}
	016	1299359	1299476	38	Hypothetical protein	Hypothetical protein H3H23_gp15 [*Escherichia* phage argo145]	$\times10^{-18}$
	017	1299468	1299645	58	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoS-640r1]	3×10^{-36}
	018	1300128	1300308	59	Hypothetical protein	MAG TPA: hypothetical protein [Caudoviricetes sp.]	1×10^{-35}
	019	1300763	1300943	59	Hypothetical protein	Mobile element protein [*Escherichia* phage vB_EcoM-606r7]	9×10^{-22}
	020	1301163	1302222	352	Hypothetical protein	Putative YdfU family protein [*Escherichia* phage vB_EcoS-640r1]	0
	021	1302222	1302588	121	Endodeoxyribonuclease	Holliday junction resolvase/crossover junction endodeoxyribonuclease RusA [*Escherichia* phage Evi]	8×10^{-76}
	022	1302596	1303139	180	Hypothetical protein	Antitermination protein q [*Escherichia* phage vB_EcoS-666r9]	1×10^{-112}
	023	1303370	1303568	65	Hypothetical protein	TrmB family transcriptional regulator [*Escherichia* phage phiSTEC1575-Stx2k]	2×10^{-42}
	024	1303718	1304768	349	DNA methyltransferase	DNA methyltransferase [Enterobacteria phage phiP27]	0
	025	1305089	1305314	74	Tellurite resistance protein	Hypothetical protein [*Klebsiella* phage 5 lv-2017]	8×10^{-34}
	026	1305310	1305463	50	Hypothetical protein	P1 TciB-like protein [*Escherichia* phage Cyrano]	1×10^{-8}
	027	1305551	1305944	130	Holin	Hypothetical protein [Enterobacteria phage phiP27]	3×10^{-73}
	028	1305933	1306209	91	Holin	Holin [Enterobacteria phage phiP27]	9×10^{-58}

3 (intact)	**029**	1306211	1306589	125	Peptidase	Endolysin [Enterobacteria phage phiP27]	2×10^{-83}
	030	1306863	1307019	51	Hypothetical protein	Hypothetical protein phiSTEC1575Stx2k_54 [*Escherichia* phage phiSTEC1575-Stx2k]	1×10^{-22}
	031	1307193	1307586	130	Hypothetical protein	Hypothetical protein p27p33 [Enterobacteria phage phiP27]	5×10^{-66}
	032	1307969	1308089	39	Hypothetical protein	Hypothetical protein [*Escherichia* phage 1W]	4×10^{-9}
	033	1308170	1308716	181	Terminase small subunit	Terminase small subunit [*Escherichia* phage ev099]	6×10^{-110}
	034	1308690	1310616	641	Terminase large subunit	Terminase large subunit [*Escherichia* phage ev207]	0
	035	1310612	1310819	68	Head completion protein	Head-tail adaptor Ad1 [*Escherichia* phage Lambda]	2×10^{-43}
	036	1310815	1312417	533	Portal protein	Portal protein [*Escherichia* phage 2H10]	0
	037	1312397	1313717	439	Serine protease	S49 family peptidase [*Escherichia* phage ev099]	0
	038	1313726	1314059	110	Head decoration protein	Head decoration [*Escherichia* phage ev099]	3×10^{-75}
	039	1314228	1314906	225	Transposase	Transposase [Stx2-converting phage Stx2a_WGPS2]	4×10^{-28}
	040	1314986	1315166	59	Hypothetical protein		
	041	1315272	1316844	523	Transposase	Transposase [*Escherichia* phage phiV205-1]	0
	042	1316968	1317847	292	Capsid protein	Phage major capsid protein [*Escherichia* phage 2H10]	0
	043	1317888	1318284	131	DNA packaging protein	DNA-packaging protein FI [*Escherichia* phage phiSTEC1575-Stx2k]	2×10^{-87}
	044	1318295	1318649	117	Head to tail attachment protein	Head closure Hc1 [*Escherichia* phage ev207]	3×10^{-80}
	045	1318660	1319239	192	Tail protein	Tail protein [*Escherichia* phage Lambda]	8×10^{-133}
	046	1319235	1319631	131	Tail protein	Tail terminator [*Escherichia* phage 1H12]	1×10^{-91}
	047	1319659	1320379	239	Tail tube protein	Major tail protein [*Escherichia* phage ev207]	4×10^{-168}
	048	1320394	1320817	140	Tail assembly chaperone	Minor tail protein g [*Escherichia* phage ev099]	1×10^{-99}
	049	1320798	1321233	144	Tail protein	Minor tail protein [Enterobacteria phage 0276]	6×10^{-103}
	050	1321225	1323787	853	Tail tape measure protein	Tail length tape measure protein TMP [*Escherichia* phage Perceval]	0

	051	1323783	1324113	109	Tail protein	Minor tail protein [*Escherichia* phage Lambda]	1×10^{-76}
	052	1324112	1324811	232	Tail protein	Tail tip complex protein [*Escherichia* phage 434]	5×10^{-173}
	053	1324960	1325560	199	Tail assembly protein	Phage tail assembly protein [*Escherichia* phage 2H10]	1×10^{-145}
	054	1325556	1326129	190	Tail assembly protein	Tail assembly protein [*Escherichia* phage ev017]	1×10^{-133}
	055	1326189	1329669	1159	Tail protein	Tail tip host specificity protein j [*Escherichia* phage vB_EcoS-640r1]	0
	056	1329736	1330336	199	Outer membrane protein	Host-cell envelope protein [*Escherichia* phage vB_EcoS-640r1]	7×10^{-141}
	057	1330336	1330510	57	Hypothetical protein	Hypothetical protein PBV4795_ORF74 [Enterobacteria phage BP-4795]	1×10^{-33}
	058	1330897	1331635	245	Hypothetical protein	Hypothetical protein [*Escherichia* phage phi458]	4×10^{-30}
3 (intact)	**059**	1331670	1331805	44	Hypothetical protein		
	060	1331765	1333502	578	Tail fiber repeat	Tail fiber protein [Enterobacteria phage P7]	0
	061	1333504	1334038	177	Tail fiber assembly protein	Tail fiber chaperone [*Escherichia* phage 503458]	1×10^{-128}
	062	1334066	1334594	175	Tail fiber assembly protein	Tail fiber assembly protein [*Escherichia* phage 503458]	1×10^{-124}
	063	1334597	1335434	278	Tail fiber protein	Tail fiber protein [*Yersinia* phage HQ103]	0
	064	1335745	1335865	39	Tail fiber assembly protein	Tail fiber assembly protein [*Salmonella* phage SPF_0923]	2×10^{-14}
	065	1335919	1336093	57	Methyltransferase	Class I SAM-dependent methyltransferase [*Escherichia* phage 2G7b]	9×10^{-34}
	066	1337142	1337976	277	Spermidine	Spermidine/putrescine ABC transporter membrane protein [Caudoviricetes sp.]	4×10^{-120}
	067	1337989	1339126	378	ABC transporter protein	Putrescine/spermidine ABC transporter ATPase protein [Caudoviricetes sp.]	0
	001	2119086	2120217	376	Integrase	Integrase [*Synechococcus* phage Yong-I1-251]	1×10^{-16}
	002	2120367	2120568	66	Hypothetical protein		
4 (intact)	**003**	2120557	2121196	212	Hypothetical protein		
	004	2121246	2123847	866	DNA polymerase	MAG TPA: hypothetical protein [Caudoviricetes sp.]	0
	005	2123895	2124018	40	Hypothetical protein		

	006	2124028	2124154	41	Hypothetical protein		
	007	2124600	2125503	300	Transcriptional regulator	Hypothetical protein [*Moraxella* phage Mcat5]	5×10^{-44}
	008	2125564	2126383	272	Hypothetical protein		
	009	2127059	2127428	122	Hypothetical protein		
	010	2127500	2128034	177	Hypothetical protein		
	011	2128059	2128233	57	Hypothetical protein		
	012	2128257	2128374	38	Hypothetical protein		
	013	2128799	2129999	399	Integrase	XerC integrase [uncultured Caudovirales phage]	3×10^{-29}
	014	2130079	2130820	246	Hypothetical protein		
	015	2131222	2132224	333	Integrase	Integrase [*Yersinia* phage vB_YenM_31.17]	3×10^{-144}
	016	2132229	2132577	115	Hypothetical protein	MAG TPA: protein of unknown function (DUF2511) [Caudoviricetes sp.]	2×10^{-78}
4 (intact)	**017**	2132606	2133257	216	Flagella biosynthesis regulator	MAG TPA: flagella biosynthesis regulator [Caudoviricetes sp.]	1×10^{-153}
	018	2133272	2133677	134	Transcriptional regulator	Transcriptional regulator [*Escherichia* phage 520873]	1×10^{-92}
	019	2133975	2134179	67	Excisionase	Cox-like excisionase and repressor [*Escherichia* phage 520873]	9×10^{-42}
	020	2134200	2134551	116	Hypothetical protein	DUF4761 family protein [*Escherichia* phage P88]	4×10^{-70}
	021	2134561	2134849	95	Hypothetical protein	Hypothetical protein HYP18_gp45 [*Xuanwuvirus* P884B11]	2×10^{-65}
	022	2134860	2135103	80	Hypothetical protein	DUF4754 family protein [*Escherichia* phage P88]	2×10^{-52}
	023	2135099	2135213	37	Hypothetical protein	Hypothetical protein HYP18_gp47 [*Xuanwuvirus* P884B11]	4×10^{-18}
	024	2135299	2135503	67	Hypothetical protein	Hypothetical protein HYP18_gp48 [*Xuanwuvirus* P884B11]	7×10^{-44}
	025	2135499	2135745	81	Transcription regulator	Hypothetical protein HYP18_gp49 [*Xuanwuvirus* P884B11]	6×10^{-55}
	026	2135741	2136041	99	Hypothetical protein	Hypothetical protein HYP18_gp50 [*Xuanwuvirus* P884B11]	2×10^{-65}
	027	2136363	2136594	76	Derepression protein	Derepression protein [*Escherichia* phage P88]	2×10^{-41}
	028	2136666	2137032	121	Hypothetical protein	Hypothetical protein SF19_gp15 [*Escherichia* phage P88]	3×10^{-82}

	029	2137038	2139846	935	Replication protein	Replication endonuclease [*Escherichia* phage P88]	0
	030	2139922	2140882	319	Plasmid segregation protein	Plasmid segregation protein ParM [*Escherichia* phage P88]	0
	031	2140886	2141198	103	Plasmid stability protein	Partition protein [*Escherichia* phage 520873]	2×10^{-68}
	032	2141749	2141872	40	Hypothetical protein		
	033	2142120	2142249	42	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM-705r2]	1×10^{-18}
	034	2142392	2142785	130	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM-569r5]	8×10^{-91}
	035	2142868	2142994	41	Hypothetical protein		
	036	2143029	2143533	167	Transposase	Transposase [*Escherichia* phage P1]	1×10^{-123}
	037	2143571	2143694	40	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM-569r5]	3×10^{-22}
	038	2143708	2144755	348	Portal vertex protein	Portal vertex protein [*Escherichia* phage vB_EcoM-569r5]	0
	039	2144754	2146506	583	Terminase	Terminase large subunit [*Escherichia* phage 520873]	0
4 (intact)	**040**	2146660	2147497	278	Capsid scaffolding protein	Capsid scaffolding protein [*Escherichia* phage vB_EcoM-569r5]	0
	041	2147520	2148573	350	Capsid protein	Major capsid protein, P2 family [*Escherichia* phage 503458]	0
	042	2148618	2149419	266	Terminase	Terminase small subunit [*Escherichia* phage P88]	0
	043	2149522	2150017	164	Head completion protein	Head-tail adaptor Ad1 [*Escherichia* phage 503458]	1×10^{-113}
	044	2150016	2150217	66	Tail Protein	Baseplate hub [*Escherichia* phage 500465-2]	3×10^{-43}
	045	2150219	2150543	107	Holin	Holin/anti-holin [*Xuanwuvirus* P884B11]	9×10^{-73}
	046	2150539	2150932	130	Peptidase endolysin	Endolysin [*Escherichia* phage P88]	1×10^{-92}
	047	2150928	2151336	135	Hypothetical protein	DUF2570 domain-containing protein [*Escherichia* phage 503458]	6×10^{-95}
	048	2151473	2151941	155	Tail completion protein	Tail protein [*Escherichia* phage P88]	3×10^{-113}
	049	2151924	2152569	214	Morphogenesis protein	Virion morphogenesis protein [*Escherichia* phage P88]	9×10^{-152}
	050	2152565	2153147	193	Baseplate assembly protein	Baseplate assembly protein [*Escherichia* phage 503458]	4×10^{-137}
	051	2153143	2153494	116	Baseplate wedge subunit	Baseplate wedge subunit [*Escherichia* phage P88]	2×10^{-80}

	052	2153497	2154394	298	Baseplate assembly protein	Baseplate protein [*Escherichia* phage 500465-2]	0
	053	2154386	2154917	176	Tail formation protein	Tail protein [*Xuanwuvirus* P884B11]	5×10^{-126}
	054	2154919	2157277	785	Tail fiber protein	Tail fiber protein [*Escherichia* phage P88]	0
	055	2157279	2157813	177	Tail fiber assembly protein	Tail fiber chaperone [*Escherichia* phage 503458]	1×10^{-128}
	056	2157841	2158369	175	Tail fiber assembly protein	Tail fiber assembly protein [*Escherichia* phage 503458]	4×10^{-124}
	057	2158372	2159209	278	Tail fiber protein	Tail fiber protein [*Xuanwuvirus* P884B11]	0
	058	2159313	2159901	195	Serine recombinase	DNA invertase [*Escherichia* phage 500465-2]	3×10^{-141}
	059	2159936	2160425	162	Hypothetical protein	Tail protein [*Escherichia* phage P88]	9×10^{-119}
	060	2160437	2163245	935	Tail tape measure protein	Tail tape measure protein [*Escherichia* phage 503458]	0
	061	2163231	2163360	42	Tail protein	Tail protein [*Escherichia* phage P88]	1×10^{-23}
	062	2163395	2163761	121	Tail assembly chaperone	Tail protein [*Escherichia* phage 500465-2]	1×10^{-82}
4 (intact)	**063**	2163815	2164328	170	Tail tube protein	Head closure [*Escherichia* phage P88]	1×10^{-121}
	064	2164327	2165512	394	Tail sheath protein	Tail sheath family protein [*Escherichia* phage 500465-2]	0
	065	2165491	2165623	43	Hypothetical protein	Hypothetical protein SF19_gp50 [*Escherichia* phage P88]	4×10^{-25}
	066	2165669	2165999	109	Tail protein	Late control d family protein [*Escherichia* phage 520873]	6×10^{-54}
	067	2165949	2167101	383	Transposase	Hypothetical protein [*Lacticaseibacillus* phage P2.4]	5×10^{-62}
	068	2167279	2168008	242	Tail protein	Tail protein [*Xuanwuvirus* P884B11]	6×10^{-179}
	069	2168050	2168311	86	Transcriptional regulator	Transcriptional regulator [*Escherichia* phage 500465-2]	1×10^{-60}
	070	2168500	2168641	46	Toxin-antitoxin system Hok/gef	Mokw-like host killing [*Escherichia* phage P88]	1×10^{-26}
	071	2168862	2168979	38	Hypothetical protein		
	072	2169455	2169632	58	Hypothetical protein		
	073	2170106	2170592	161	Phosphatidylglycerophosphate synthase	MAG TPA: phosphatidylglycerophosphate synthetase [Caudoviricetes sp.]	1×10^{-100}
	074	2170711	2172544	610	Excinuclease	Excinuclease ABC subunit c [*Acinetobacter* phage MD-2021a]	0

4 (intact)	**075**	2172540	2173197	218	Response regulator	Response regulator UvrY [*Acinetobacter* phage MD-2021a]	2×10^{-66}
	076	2173435	2173564	42	Hypothetical protein		
	077	2173655	2173880	74	Hypothetical protein		
	078	2173946	2174669	240	Transcriptional regulator	DNA-binding transcriptional activator SdiA [*Acinetobacter* phage MD-2021a]	3×10^{-28}
	079	2174898	2175651	250	Amino acid transport system	ABC transporter ATP-binding protein [*Acinetobacter* phage MD-2021a]	4×10^{-75}
5 (incomplete)	**001**	2271592	2272459	288	Transposase	Transposase [*Acinetobacter* phage ab105-1phi]	7×10^{-65}
	002	2272455	2272755	99	Transposase	MAG TPA: transposase [Caudoviricetes sp.]	2×10^{-65}
	003	2272875	2273007	43	Hypothetical protein		
	004	2273018	2273735	238	Transposase	MAG TPA: transposase [Caudoviricetes sp.]	3×10^{-10}
	005	2274540	2274951	136	Hypothetical protein	MAG TPA_asm: putative zinc-ribbon [Caudoviricetes sp.]	5×10^{-10}
	006	2275136	2275616	159	Transposase	MAG TPA: transposase [*Siphoviridae* sp. CtBLh2]	3×10^{-10}
	007	2275682	2276105	140	Transposase	MAG TPA: transposase [Caudoviricetes sp.]	3×10^{-24}
	008	2276672	2277242	189	Replication protein	MAG TPA: inovirus Gp2, partial [Caudoviricetes sp.]	2×10^{-49}
	009	2277460	2277724	87	Hypothetical protein		
	010	2277957	2278098	46	Hemolysin expression modulator	MAG TPA: gene expression modulator [Caudoviricetes sp.]	5×10^{-12}
	011	2278841	2279048	68	Transcriptional regulator	MAG TPA_asm: putative transcriptional regulator [Caudoviricetes sp.]	4×10^{-33}
6 (intact)	**001**	2287578	2288052	157	Antirestriction protein	MAG TPA: antirestriction protein [Caudoviricetes sp.]	4×10^{-74}
	002	2288067	2288544	158	DNA repair protein	MAG TPA: DNA repair protein [Caudoviricetes sp.]	2×10^{-93}
	003	2288606	2288828	73	Hypothetical protein	Hypothetical protein A1q_00002 [*Klebsiella* phage vlcpia1q]	8×10^{-16}
	004	2288827	2288941	37	Hypothetical protein	MAG TPA: hypothetical protein [Caudoviricetes sp.]	7×10^{-6}
	005	2289065	2289359	97	Antitoxin (CbeA)	MAG TPA: antitoxin [Caudoviricetes sp.]	2×10^{-43}
	006	2289448	2289823	124	Toxin (CbeA)	MAG TPA: toxin [Caudoviricetes sp.]	5×10^{-44}
	007	2289822	2290026	67	Hypothetical protein		

6 (intact)	**008**	2290105	2290246	46	Hypothetical protein		
	009	2290272	2291547	424	Transposase	Transposase [*Escherichia* phage 4A7]	0
	010	2291737	2292013	91	Transposase	Transposase [Enterobacteria phage BP-4795]	2×10^{-61}
	011	2292012	2292900	295	Transposase	Is3 family transposase [Enterobacteria phage BP-4795]	0
	012	2292902	2293199	98	Transposase	Mobile element protein [*Escherichia* phage D2]	3×10^{-47}
	013	2293229	2293580	116	Transposase	Transposase [*Escherichia* phage 4A7]	2×10^{-81}
	014	2293576	2293939	120	Transposase	Transposase [*Escherichia* phage 4A7]	4×10^{-83}
	015	2294239	2295391	383	Hypothetical protein	Hypothetical protein [*Lacticaseibacillus* phage P2.4]	5×10^{-62}
	016	2295530	2296349	272	Transposase	Transposase [*Acinetobacter* phage ab105-1phi]	3×10^{-83}
	017	2296494	2296815	106	Tyrosine recombinase	Integrasse [*Escherichia* phage ger2]	3×10^{-47}
7 (intact)	**001**	2300741	2302823	693	Outer membrane protein	MAG TPA: outer membrane protein, partial [*Siphoviridae* sp. Ct5lu19]	1×10^{-45}
	002	2304018	2305557	512	Transposase	Transposase [*Escherichia* phage phiV205-1]	0
	003	2305606	2305954	115	Transposase	Transposase [Stx2-converting phage 1717]	3×10^{-79}
	004	2306454	2308026	523	Transposase	Transposase [*Escherichia* phage phiV205-1]	0
	005	2308045	2308393	115	Transposase	Transposase [Stx2-converting phage Stx2a_WGPS2]	1×10^{-58}
	006	2308392	2309040	215	Hypothetical protein	Transposase [Stx2-converting phage Stx2a_WGPS2]	3×10^{-28}
	007	2309110	2310115	334	Transposase	IS21-like element IS100 family transposase [*Escherichia* phage 503458]	0
	008	2310114	2310891	258	Transposase	IS21-like element IS100kyp family helper ATPase IstB [*Escherichia* phage 520873]	0
	009	2311227	2311347	39	Hypothetical protein		
	010	2311447	2311777	109	Hypothetical protein		
	011	2311948	2312695	248	Membrane protein	MAG TPA: putative membrane protein [Caudoviricetes sp.]	9×10^{-87}
	012	2312779	2313079	99	Transposase	MAG TPA: transposase [Caudoviricetes sp.]	1×10^{-64}
	013	2313075	2313942	288	Transposase	MAG TPA: Mos transposase [Caudoviricetes sp.]	0

	001	2839109	2839262	50	Hypothetical protein	Hypothetical protein F365_gp57 [*Escherichia* phage TL-2011b]	3×10^{-27}
	002	2839279	2839471	63	Hypothetical protein	DUF2633 family protein [*Escherichia* phage TL-2011b]	7×10^{-39}
	003	2839541	2839670	42	Hypothetical protein		
	004	2839786	2840302	171	Outer membrane lipoprotein	DNA methyltransferase [*Escherichia* phage TL-2011b]	1×10^{-121}
	005	2840317	2840857	179	Hypothetical protein	DUF5384 family protein [*Escherichia* phage TL-2011b]	3×10^{-125}
	006	2841074	2841557	160	Hypothetical protein	Rz-like spanin [*Salmonella* phage epsilon15]	5×10^{-102}
	007	2841553	2842180	208	Chitinase	Endolysin [Enterobacter phage Tyrion]	6×10^{-141}
	008	2842464	2842782	105	Holin	Membrane protein [*Salmonella* phage epsilon15]	9×10^{-71}
	009	2843003	2844458	484	Tail protein	Tail protein [*Escherichia* phage vB_EcoM_ECOO78]	0
	010	2844527	2846099	523	Transposase	Transposase [*Escherichia* phage phiV205-1]	0
	011	2846118	2846466	115	Transposase	Transposase [Stx2-converting phage Stx2a_WGPS2]	1×10^{-58}
8 (intact)	**012**	2846465	2847113	215	Transposase	Transposase [Stx2-converting phage Stx2a_WGPS2]	3×10^{-28}
	013	2847192	2848008	271	Tail fiber protein	Tail fiber protein [*Escherichia* phage phiV10]	3×10^{-136}
	014	2848296	2848461	54	Hypothetical protein	Hypothetical protein F365_gp47 [*Escherichia* phage TL-2011b]	3×10^{-33}
	015	2849606	2850299	230	Antirepressor protein	MAG TPA: repressor domain protein [Caudoviricetes sp.]	2×10^{-171}
	016	2850496	2850715	72	Hypothetical protein	MAG TPA: hypothetical protein [Caudoviricetes sp.]	5×10^{-31}
	017	2851176	2851338	53	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoP_zx5]	3×10^{-31}
	018	2851369	2851666	98	Hypothetical protein	Hypothetical protein [*Escherichia* phage phiV142-3]	2×10^{-60}
	019	2851861	2854336	824	Structural protein	Hypothetical protein [*Escherichia* phage vB_EcoP_zx5]	0
	020	2854340	2856143	600	Hypothetical protein	Hypothetical protein SPN1S_0018 [*Salmonella* phage SPN1S]	0
	021	2856139	2858650	836	Structural protein	Head protein [Enterobacter phage Tyrion]	0
	022	2858662	2859205	180	Hypothetical protein	Hypothetical protein [*Salmonella* virus pat1]	1×10^{-90}
	023	2859204	2859669	154	Hypothetical protein	Hypothetical protein [Enterobacter phage buct554]	4×10^{-105}

	024	2859668	2862146	825	Hypothetical protein	Hypothetical protein [Enterobacter phage buct554]	0
	025	2862145	2862751	201	Hypothetical protein	Tail protein [*Escherichia* phage phiV10]	4×10^{-150}
	026	2862750	2863074	107	Hypothetical protein	Hypothetical protein [*Escherichia* phage phiV142-3]	2×10^{-73}
	027	2863124	2863460	111	Hypothetical protein	Hypothetical protein f365_gp31 [*Escherichia* phage TL-2011b]	1×10^{-76}
	028	2863470	2863908	145	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoP_zx5]	8×10^{-101}
	029	2863959	2864946	328	Capsid protein	Major head protein [*Escherichia* phage TL-2011b]	0
	030	2864960	2865656	231	Protease	Protease [*Escherichia* phage TL-2011b]	6×10^{-166}
	031	2865658	2865955	98	Hypothetical protein	Hypothetical protein f365_gp27 [*Escherichia* phage TL-2011b]	7×10^{-68}
	032	2865951	2867631	559	Head to tail connecting protein	Head to tail connecting protein [*Salmonella* virus pat1]	0
	033	2867645	2867852	68	Hypothetical protein	Hypothetical protein epsilon15p03 [*Salmonella* phage epsilon15]	5×10^{-42}
	034	2868506	2869112	201	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoP_pas61]	7×10^{-74}
8 (intact)	**035**	2869155	2870637	493	Terminase large subunit	Terminase large subunit [*Salmonella* phage SPN1S]	0
	036	2870633	2871305	223	Terminase small subunit	Terminase small subunit [*Salmonella* virus pat1]	3×10^{-162}
	037	2871332	2871671	112	Hypothetical protein	Hypothetical protein [*Salmonella* virus pat1]	5×10^{-69}
	038	2871860	2872715	284	Hypothetical protein	MAG TPA: protein of unknown function (DUF551), partial [Caudoviricetes sp.]	4×10^{-96}
	039	2872711	2873077	121	Endonuclease	HNH endonuclease [*Escherichia* phage argo145]	2×10^{-86}
	040	2873078	2873465	128	Hypothetical protein	Ead/ea22-like family protein [*Escherichia* phage rcs47]	1×10^{-35}
	041	2873772	2873970	65	Hypothetical protein	MAG TPA: hypothetical protein [Caudoviricetes sp.]	5×10^{-30}
	042	2873966	2874440	157	Hypothetical protein	Hypothetical protein [*Salmonella* virus pat1]	2×10^{-54}
	043	2874611	2874956	114	Endonuclease	DUF1064 domain-containing protein [*Escherichia* phage phiV10]	4×10^{-80}
	044	2875073	2875859	261	Replication protein	Putative replication protein p [*Escherichia* phage phiV10]	0
	045	2875855	2876671	271	Replication protein	Replication initiation protein [*Escherichia* phage phiV10]	0
	046	2876686	2876887	66	Hypothetical protein	Hypothetical protein phiV10p42 [*Escherichia* phage phiV10]	4×10^{-44}

8 (intact)	**047**	2877037	2877268	76	Hypothetical protein	Hypothetical protein f365_gp11 [*Escherichia* phage TL-2011b]	9×10^{-53}
	048	2877422	2878007	194	Repressor protein	Helix-turn-helix domain-containing protein [*Escherichia* phage phiV10]	3×10^{-143}
	049	2878160	2878313	50	Hypothetical protein	Hypothetical protein f365_gp09 [*Escherichia* phage TL-2011b]	7×10^{-31}
	050	2878315	2878615	99	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoP_pas61]	5×10^{-65}
	051	2878814	2879636	273	Exodeoxyribonuclease	Hypothetical protein ydc107_5455 [*Escherichia* phage ydc107_2]	0
	052	2879632	2880574	313	DNA annealing protein	RecT family protein [*Escherichia* phage YDC107_2]	0
	053	2880623	2880872	82	Transcriptional regulator	Alpa family regulatory protein [*Escherichia* phage phiV10]	2×10^{-55}
	054	2881029	2881281	83	Transcriptional activator	Perc transcriptional activator family protein [*Escherichia* phage ydc107_2]	5×10^{-55}
	055	2881570	2881924	117	DNA methyltransferase	Adenine DNA methyltransferase [*Escherichia* phage vB_EcoP_ZX5]	3×10^{-82}
	056	2881920	2882580	219	Metallophosphatase	Serine/threonine protein phosphatase [*Escherichia* phage vB_EcoP_pas61]	8×10^{-164}
	057	2882582	2883839	418	Integrase	Integrase [*Escherichia* phage vB_EcoP_pas61]	0
	058	2884031	2885609	525	GMP synthase	GMP synthase domain protein [*Klebsiella* phage ST11-OXA245phi3.2]	0
	059	2885677	2887213	511	IMP dehydrogenase	Inosine-5'-monophosphate dehydrogenase [*Klebsiella* phage ST512-KPC3phi13.3]	0
9 (intact)	**001**	2953837	2954053	71	Recombinase	MAG TPA: gamma delta resolvase, site specific recombination [Caudoviricetes sp.]	3×10^{-41}
	002	2954037	2954391	117	DNA invertase	DNA-invertase [*Escherichia* phage Cartapus]	2×10^{-68}
	003	2954700	2954820	39	Tail protein	MAG TPA: putative tail fiber protein [*Myoviridae* sp. Ct8ar17]	9×10^{-21}
	004	2954823	2955243	139	Tail fiber assembly protein	MAG TPA: tail fiber assembly protein [*Myoviridae* sp. Ct8ar17]	6×10^{-97}
	005	2955214	2955808	197	Tail fiber assembly protein	Tail fiber assembly [*Escherichia* phage 500465-2]	2×10^{-139}
	006	2955807	2956677	289	Tail fiber protein	Side tail fiber protein [*Escherichia* phage vB_EcoM-606r2]	1×10^{-136}
	007	2956676	2957357	226	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM-606r2]	4×10^{-164}
	008	2957353	2958553	399	Baseplate protein	Baseplate J domain-containing protein [*Escherichia* phage vB_EcoM-606r2]	0
	009	2958552	2958906	117	Tail sheath initiator protein	Hypothetical protein [*Escherichia* phage vB_EcoM-606r2]	4×10^{-80}
	010	2958905	2959658	250	Baseplate component	Hypothetical protein [*Escherichia* phage vB_EcoM-606r2]	2×10^{-180}

	011	2959720	2959891	56	Hypothetical protein		
	012	2959894	2960116	73	Hypothetical protein		
	013	2960456	2960804	115	Hypothetical protein	MAG TPA: hypothetical protein [*Myoviridae* sp. Ctnhb8]	2×10^{-67}
	014	2960806	2961871	354	Tail protein	Hypothetical protein [*Escherichia* phage vB_EcoM-689r2]	0
	015	2961873	2962176	100	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM-606r2]	1×10^{-68}
	016	2962175	2962763	195	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM-606r2]	2×10^{-137}
	017	2962762	2964751	662	Structural protein	MAG TPA: virion protein [Caudoviricetes sp.]	0
	018	2964928	2965381	150	Tail assembly chaperone protein	Hypothetical protein [*Escherichia* phage vB_EcoM-606r2]	1×10^{-106}
	019	2965384	2965825	146	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM-606r2]	4×10^{-105}
	020	2965835	2966981	381	Tail sheath protein	Hypothetical protein [*Escherichia* phage vB_EcoM-606r2]	0
	021	2966984	2967533	182	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM-683r1]	7×10^{-133}
9 (intact)	**022**	2967522	2967912	129	Head closure protein	Hypothetical protein [*Escherichia* phage vB_EcoM-606r2]	5×10^{-90}
	023	2967898	2968453	184	Morphogenesis protein	Hypothetical protein [*Escherichia* phage vB_EcoM-606r2]	1×10^{-127}
	024	2968449	2968857	135	Head to tail adaptor	Hypothetical protein [*Escherichia* phage vB_EcoM-606r2]	6×10^{-94}
	025	2968822	2969191	122	Hypothetical protein	MAG TPA: SlyX-like protein [Caudoviricetes sp.]	3×10^{-83}
	026	2969231	2970173	313	Capsid protein	Hypothetical protein [*Escherichia* phage vB_EcoM-606r2]	0
	027	2970184	2970691	168	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM-606r2]	4×10^{-118}
	028	2970694	2971915	406	Prohead serine protease	Hypothetical protein [*Escherichia* phage vB_EcoM-606r2]	0
	029	2971929	2972379	149	Hypothetical protein	Head morphogenesis protein [*Escherichia* phage vB_EcoM-606r2]	3×10^{-104}
	030	2972554	2974021	488	Portal protein	Hypothetical protein [*Escherichia* phage vB_EcoM-606r2]	0
	031	2974020	2975643	540	Terminase large subunit	Terminase large subunit [*Escherichia* phage vB_EcoM-606r2]	0
	032	2975645	2976218	190	Terminase small subunit	Terminase small subunit [*Escherichia* phage vB_EcoM-606r2]	1×10^{-139}
	033	2976279	2976804	174	Hypothetical protein	Endopeptidase Rz [*Escherichia* phage vB_EcoM-606r2]	1×10^{-124}

	034	2976787	2977264	158	Lysozyme	Endolysin [*Escherichia* phage vB_EcoM-606r2]	4×10^{-116}
	035	2977267	2977609	113	Holin	Holin/antiholin component S [*Escherichia* phage vB_EcoM-606r2]	2×10^{-76}
	036	2977879	2977996	38	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM-611r5]	6×10^{-21}
	037	2978054	2978396	113	Hypothetical protein	TPA: hypothetical protein [Caudoviricetes sp.]	3×10^{-81}
	038	2978427	2978850	140	Antitermination protein	Antitermination protein Q [*Escherichia* phage vB_EcoM-606r2]	2×10^{-91}
	039	2979134	2981057	640	Virulence associated protein	TPA: virulence associated protein e [Caudoviricetes sp.]	0
	040	2981042	2981480	145	Hypothetical protein		
	041	2981891	2982341	149	Repressor protein	Repressor protein [*Escherichia* phage vB_EcoM-606r2]	8×10^{-109}
	042	2982398	2982542	47	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM-606r2]	5×10^{-25}
	043	2982759	2983221	153	Hypothetical protein	Hypothetical protein [*Escherichia* phage vb_ecop-720r6]	1×10^{-37}
	044	2983508	2983658	49	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM-606r2]	4×10^{-26}
9 (intact)	**045**	2983654	2984557	300	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM-606r2]	0
	046	2984559	2985861	433	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM-606r2]	0
	047	2985876	2986425	182	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM-606r2]	1×10^{-131}
	048	2986477	2987107	209	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM-606r2]	8×10^{-145}
	049	2987153	2987516	120	DNA polymerase	DNA polymerase [*Escherichia* phage vB_EcoM-606R2]	1×10^{-73}
	050	2987677	2989228	516	DNA polymerase	DNA polymerase [*Escherichia* phage vB_EcoM-606R2]	0
	051	2989233	2989449	71	Transcriptional regulator	MAG TPA: perc transcriptional activator [Caudoviricetes sp.]	1×10^{-45}
	052	2989445	2989745	99	Hypothetical protein	TPA: nucleoside 2-deoxyribosyltransferase like protein [Caudoviricetes sp.]	2×10^{-68}
	053	2989734	2990028	97	Nuclease	TPA: nuclease [Caudoviricetes sp.]	2×10^{-61}
	054	2990000	2991014	337	Cytosine methyltransferase	TPA: cytosine specific methyltransferase [Caudoviricetes sp.]	0
	055	2991016	2991946	309	Hypothetical protein	TPA: transcriptional regulator [Caudoviricetes sp.]	2×10^{-147}
	056	2992056	2992980	307	DNA helicase	DNA helicase [*Escherichia* phage vB_EcoM-606R2]	0

9 (intact)	**057**	2993046	2993451	134	ATPase	DNA helicase [*Escherichia* phage vB_EcoM-606R2]	1×10^{-93}
10 (incomplete)	**001**	3041923	3043285	453	Signal recognition protein	Signal recognition particle protein [*Escherichia* phage vB_EcoM-613r3]	0
	002	3043376	3044243	288	Inner membrane protein	Inner membrane protein [*Escherichia* phage vB_EcoM-613r3]	0
	003	3044287	3045550	420	Transporter protein	Putative membrane protein [*Escherichia* phage vB_EcoM-613r3]	0
	004	3045604	3046198	197	Heat shock protein	Heat shock protein [*Escherichia* phage vB_EcoM-613r3]	2×10^{-140}
	005	3046194	3046332	45	Hypothetical protein	Putative cytoplasmic protein [*Escherichia* phage vB_EcoM-613r3]	2×10^{-24}
	006	3046392	3047199	268	Inorganic polyphosphate	NAD kinase [*Escherichia* phage vB_EcoM-613r3]	0
	007	3047284	3048946	553	DNA repair protein	DNA repair protein recn [*Escherichia* phage 500465-1]	0
	008	3049094	3049436	113	Outer membrane protein	Outer membrane protein assembly factor bame [*Escherichia* phage 500465-1]	2×10^{-80}
	009	3049497	3049788	96	Antitoxin component (RatB)	Hypothetical protein EC_CP1639_68 [Enterobacteria phage CP-1639]	6×10^{-64}
	010	3049777	3050167	129	Ribosome association toxin (RatA)	Ribosome association toxin [*Escherichia* phage vB_EcoM-613R3]	9×10^{-94}
	011	3050385	3050868	160	tmRNA-binding protein	Ssra-binding protein smpb [*Escherichia* phage 500465-1]	2×10^{-117}
	012	3051568	3052840	423	Integrase	Prophage CP4-57 integrase [*Escherichia* phage phi467]	0
11 (incomplete)	**001**	4101828	4102488	219	Outer membrane protein	Putative outer membrane protein [*Acinetobacter* phage MD-2021a]	2×10^{-47}
	002	4102591	4103566	324	Glyoxylate/hydroxypyruvate reductase	2-oxo-carboxylic acid reductase [*Acinetobacter* phage MD-2021a]	3×10^{-89}
	003	4103615	4104326	236	Hypothetical protein		
	004	4104759	4105050	96	Hypothetical protein		
	005	4105330	4105543	70	Cold shock protein	Cold shock-like protein cspb [*Escherichia* phage Tritos]	3×10^{-35}
	006	4105730	4105883	50	Hypothetical protein		
	007	4105962	4106484	173	Transposase	TPA: Mos transposase [Caudoviricetes sp.]	2×10^{-17}
	008	4106480	4107332	283	Transposase	TPA: Mos transposase [Caudoviricetes sp.]	4×10^{-116}
12 (intact)	**001**	4684594	4685104	169	Hypothetical protein	Hypothetical protein Lys8385Vzw_28 [*Escherichia* phage Lys8385Vzw]	3×10^{-123}
	002	4685358	4685520	53	Hypothetical protein	Hypothetical protein Lys8385Vzw_29 [*Escherichia* phage Lys8385Vzw]	5×10^{-32}

	003	4685559	4685733	57	Hypothetical protein	Hypothetical protein F853_gp24 [Enterobacteria phage mEp460]	3×10^{-32}
	004	4685871	4686906	344	Hypothetical protein		
	005	4687083	4687227	47	Hypothetical protein		
	006	4687416	4687989	190	Exoribonuclease	Exonuclease [Enterobacteria phage mEp460]	3×10^{-138}
	007	4687988	4688798	269	Hypothetical protein	Hypothetical protein sb31 [*Salmonella* phage st64b]	6×10^{-93}
	008	4688797	4689139	113	Hypothetical protein	TPA: protein of unknown function (DUF551) [Caudoviricetes sp.]	5×10^{-79}
	009	4689150	4689687	178	Hydrolase	Hypothetical protein [*Escherichia* phage 1720a-02]	9×10^{-133}
	010	4689814	4690639	274	Hypothetical protein	Fig00642676: hypothetical protein [*Escherichia* phage 2H10]	0
	011	4690704	4691067	120	Hydrolase	Protein YfdP [*Escherichia* phage phiSTEC1575-Stx2k]	9×10^{-83}
	012	4691524	4691887	120	Repressor protein	Repressor protein CI [*Escherichia* phage vB_EcoM-720r4]	2×10^{-84}
	013	4692273	4692471	65	Transcriptional regulator	Hypothetical protein Lys8385Vzw_42 [*Escherichia* phage Lys8385Vzw]	6×10^{-43}
12 (intact)	**014**	4692498	4693083	194	Regulatory protein	Fig00639586: hypothetical protein [*Escherichia* virus mEp460_4f5]	5×10^{-143}
	015	4693427	4694369	313	Replication protein	DNA replication protein O [*Escherichia* phage vB_EcoM-720r4]	0
	016	4694365	4694860	164	Transcriptional activator	Putative protein YfdN [*Escherichia* phage vB_EcoM-689r6]	1×10^{-117}
	017	4694859	4695513	217	N-6-adenine-methyltransferase	N-6-adenine-methyltransferase [*Escherichia* phage Lys8385Vzw]	1×10^{-162}
	018	4695509	4695836	108	Transcriptional repressor	Transcriptional repressor [Stx2-converting phage Stx2a_WGPS2]	1×10^{-74}
	019	4695832	4696222	129	Endodeoxyribonuclease	RusA family crossover junction endodeoxyribonuclease [*Shigella* phage SfII]	3×10^{-92}
	020	4696241	4697051	269	Hypothetical protein	Hypothetical protein SJJBTUD_0047 [*Escherichia* phage Ayreon]	0
	021	4697058	4698048	329	Hypothetical protein	HNH endonuclease [Enterobacteria phage mEp460]	0
	022	4698061	4698814	250	Antitermination protein	Anti-termination protein q-like [Enterobacteria phage mEp460]	0
	023	4698964	4699222	85	Hypothetical protein	Hypothetical protein BOW93_gp067 [*Salmonella* phage 118970_sal3]	1×10^{-52}
	024	4699301	4699688	128	Holin	Holin family protein [*Salmonella* phage 118970_sal3]	2×10^{-82}
	025	4699955	4700570	204	Chitinase	Endolysin [*Salmonella* phage 118970_sal3]	7×10^{-150}

	026	4700572	4700959	128	Hypothetical protein	Rz-like spanin [*Salmonella* phage 118970_sal3]	6×10^{-82}
	027	4701199	4701367	55	Hypothetical protein		
	028	4701618	4701969	116	Endonuclease	HNH endonuclease [Enterobacteria phage SfI]	7×10^{-80}
	029	4702094	4702589	164	Terminase small subunit	Terminase small subunit [Enterobacteria phage SfV]	5×10^{-117}
	030	4702585	4704319	577	Terminase large subunit	Terminase large subunit [Enterobacteria phage SfV]	0
	031	4704315	4704477	53	Hypothetical protein	TPA: protein of unknown function (DUF1056) [Caudoviricetes sp.]	2×10^{-28}
	032	4704466	4705693	408	Portal protein	Portal protein [Enterobacteria phage SfV]	0
	033	4705685	4706288	200	Protease	Head maturation protease [Enterobacteria phage SfV]	8×10^{-147}
	034	4706298	4707528	409	Capsid protein	Major head protein [Enterobacteria phage SfV]	0
	035	4707606	4707930	107	Head-tail connector	Head-tail adaptor Ad1 [Enterobacteria phage SfV]	5×10^{-71}
	036	4707926	4708337	136	Head-tail adaptor	Head closure Hc1 [*Shigella* phage Sflv]	5×10^{-95}
12 (intact)	**037**	4708308	4708833	174	Structural protein	Tail completion or neck1 protein [*Shigella* phage Sflv]	1×10^{-122}
	038	4708829	4709390	186	Hypothetical protein	Hypothetical protein SfVp09 [Enterobacteria phage SfV]	9×10^{-137}
	039	4709398	4709569	56	Hypothetical protein	DUF2635 domain-containing protein [Enterobacteria phage SfV]	2×10^{-33}
	040	4709552	4711049	498	Tail sheath protein	Tail sheath protein [*Escherichia* phage vB_EcoM-720r4]	0
	041	4711048	4711405	118	Tail tube protein	Tail tube protein [Enterobacteria phage SfV]	2×10^{-85}
	042	4711401	4711725	107	Tail assembly chaperone protein	Small tail protein e [*Escherichia* phage vB_EcoM-720r4]	4×10^{-73}
	043	4711809	4713711	633	Tail tape measure protein	Tape measure protein [*Escherichia* phage vB_EcoM-720r4]	0
	044	4713742	4714264	173	Hypothetical protein	Hypothetical protein [*Escherichia* phage vB_EcoM-720r4]	3×10^{-115}
	045	4714407	4715694	428	DNA circularization protein	Tail/DNA circulation protein [*Escherichia* phage vB_EcoM-720R4]	0
	046	4715690	4716770	359	Tail protein	Baseplate hub [*Shigella* phage Sflv]	0
	047	4716769	4717318	182	Baseplate assembly protein	Baseplate assembly protein [*Escherichia* phage vB_EcoM-689r6]	4×10^{-113}
	048	4717317	4717743	141	Hypothetical protein	Baseplate wedge subunit [Enterobacteria phage SfV]	2×10^{-98}

12 (intact)	**049**	4717729	4718788	352	Baseplate protein	Baseplate protein [Enterobacteria phage SfI]	0
	050	4718778	4719363	194	Hypothetical protein	Baseplate wedge subunit [Enterobacteria phage SfV]	1×10^{-143}
	051	4719366	4720188	273	Tail fiber protein	TPA: tail fiber protein [Caudoviricetes sp.]	1×10^{-97}
	052	4720187	4720781	197	Tail fiber assembly protein	Tail fiber assembly [*Escherichia* phage 500465-2]	2×10^{-139}
	053	4720752	4721172	139	Tail fiber assembly protein	TPA: tail fiber assembly protein [*Myoviridae* sp. Ct8ar17]	6×10^{-97}
	054	4721150	4721282	43	Hypothetical protein		
	055	4721608	4722172	187	Serine recombinase	TPA: gamma delta resolvase, site specific recombination [*Myoviridae* sp. Ct8ar17]	2×10^{-137}
	056	4722315	4722564	82	DNA damage-inducible protein	RecA filament binding protein DinI [Enterobacteria phage mEp460]	3×10^{-54}
	057	4722625	4723723	365	Integrase	Integrase family protein [*Escherichia* phage Lys8385Vzw]	0
	058	4723811	4724849	345	TRNA-dihydrouridine synthase	TPA: tRNA-dihydrouridine synthase A [*Myoviridae* sp. Ct8ar17]	0
	059	4724982	4725225	80	Shock protein	TPA: shock protein G [*Myoviridae* sp. Ct8ar17]	6×10^{-47}
	060	4725390	4726374	327	Quinone oxidoreductase	Zinc-containing alcohol dehydrogenase superfamily protein [stenotrophomonas phage ts-12]	1×10^{-53}
	061	4726456	4727872	471	DNA helicase	Hypothetical protein JK004_33 [Cronobacter phage JK004]	0
13 (incomplete)	**001**	4829797	4830643	281	Hypothetical protein	TPA: protein of unknown function (DUF4942) [Caudoviricetes sp.]	3×10^{-118}
	002	4830727	4831150	140	Hypothetical protein		
	003	4831146	4831524	125	Toxin (cbta)	TPA: toxin [Caudoviricetes sp.]	7×10^{-44}
	004	4831613	4831862	82	Antitoxin (cbea)	TPA: antitoxin [Caudoviricetes sp.]	7×10^{-32}
	005	4831997	4832642	214	Hypothetical protein	TPA: hypothetical protein [Caudoviricetes sp.]	2×10^{-91}
	006	4832660	4832882	73	Hypothetical protein	TPA: protein of unknown function (DUF987) [Caudoviricetes sp.]	3×10^{-30}
	007	4832944	4833421	158	DNA repair protein	TPA: DNA repair protein [Caudoviricetes sp.]	2×10^{-93}
	008	4833435	4833909	157	Antirestriction protein	TPA: antirestriction protein [Caudoviricetes sp.]	7×10^{-74}
	009	4834210	4835029	272	Hypothetical protein	TPA: protein of unknown function (DUF932) [Caudoviricetes sp.]	4×10^{-152}

13 (incomplete)	**010**	4835131	4835320	62	Hypothetical protein		
	011	4835791	4836097	101	Transposase	Transposase [Stx2-converting phage Stx2a_WGPS2]	2×10^{-20}
	012	4836177	4836357	59	Hypothetical protein		
	013	4836463	4838035	523	Transposase	Transposase [Stx2-converting phage Stx2a_WGPS2]	0
14 (intact)	**001**	5072386	5073610	407	Integrase	Site-specific integrase [Enterobacteria phage cdti]	0
	002	5073792	5077620	1275	Anti-phage defense-associated sirtuin	Anti-phage defense-associated sirtuin dsr1 [*Escherichia* coli]	0
	003	5078016	5078637	206	Hypothetical protein	Fig00639676: hypothetical protein [*Escherichia* phage 2H10]	2×10^{-142}
	004	5078636	5078999	120	Hypothetical protein	TPA: protein of unknown function (DUF551) [Caudoviricetes sp.]	5×10^{-84}
	005	5078989	5079553	187	Hypothetical protein	Hypothetical protein [*Escherichia* phage 1720a-02]	2×10^{-132}
	006	5079653	5080478	274	Hypothetical protein	DUF2303 family protein [*Shigella* phage Sflv]	0
	007	5080543	5080924	126	Hypothetical protein	Protein YfdP [*Escherichia* phage phiSTEC1575-Stx2k]	4×10^{-88}
	008	5081574	5082249	224	Repressor protein	Umud-like protein [*Escherichia* phage 2H10]	3×10^{-169}
	009	5082339	5082540	66	Hypothetical protein	Transcriptional repressor [Enterobacteria phage SfV]	1×10^{-42}
	010	5082583	5083135	183	Regulatory protein	Hypothetical protein SJJBTUD_0040 [*Escherichia* phage Ayreon]	1×10^{-133}
	011	5083131	5083968	278	Transcriptional regulator	Transcriptional regulator [Enterobacteria phage SfV]	0
	012	5083960	5084197	78	Hypothetical protein	Hypothetical protein SJJBTUD_0042 [*Escherichia* phage Ayreon]	1×10^{-50}
	013	5084193	5085012	272	Replication protein	Replication protein [Enterobacteria phage SfV]	0
	014	5085014	5085503	162	Transcriptional activator	Fig00641663: hypothetical protein [*Escherichia* virus mEp460_4f5]	2×10^{-118}
	015	5085502	5086156	217	N-6-adenine-methyltransferase	N-6-adenine-methyltransferase [*Escherichia* phage Lys8385Vzw]	7×10^{-162}
	016	5086152	5086479	108	Transcriptional repressor	Transcriptional repressor [Enterobacteria phage mEp460]	8×10^{-76}
	017	5086475	5086865	129	Endodeoxyribonuclease	RusA family crossover junction endodeoxyribonuclease [Enterobacteria phage cdti]	3×10^{-90}
	018	5086884	5087007	40	Hypothetical protein	Phage-related protein [*Escherichia* virus mEp460_4f5]	1×10^{-9}
	019	5087048	5087414	121	Hypothetical protein	Unknown [*Shigella* phage sfx]	1×10^{-83}

	020	5087437	5088277	279	Transposase	TPA: Mos transposase [Caudoviricetes sp.]	0
	021	5088388	5089030	213	Hypothetical protein	Hypothetical protein SJJBTUD_0047 [*Escherichia* phage Ayreon]	4×10^{-160}
	022	5089037	5090027	329	Endonuclease	HNH endonuclease [Enterobacteria phage mEp460]	0
	023	5090040	5090793	250	Antitermination protein	Antitermination protein Q [*Escherichia* phage Lys8385Vzw]	0
	024	5091180	5091318	45	Hypothetical protein		
	025	5091400	5092453	350	DNA methyltransferase	DNA methyltransferase [Enterobacteria phage cdti]	0
	026	5092529	5092736	68	Holin	Holin [Enterobacteria phage mEp460]	3×10^{-43}
	027	5092735	5093233	165	Endolysin	Lysin [*Escherichia* phage Ayreon]	2×10^{-119}
	028	5093229	5093673	147	Lysis protein	Rz-like spanin [*Escherichia* phage 4A7]	6×10^{-92}
	029	5093791	5093905	37	Hypothetical protein		
	030	5094711	5095257	181	Terminase small subunit	Terminase small subunit [*Escherichia* phage ev207]	1×10^{-130}
14 (intact)	**031**	5095231	5097157	641	Terminase large subunit	Terminase large subunit [*Escherichia* phage ev207]	0
	032	5097153	5097360	68	Head-tail adaptor	Head-tail adaptor Ad1 [*Escherichia* phage Lambda]	1×10^{-42}
	033	5097356	5098958	533	Portal protein	Portal protein [*Escherichia* phage phiSTEC1575-Stx2k]	0
	034	5098938	5100258	439	Peptidase	Phage capsid and scaffold [*Escherichia* phage 2H10]	0
	035	5100267	5100600	110	Head decoration protein	Head decoration [*Escherichia* phage 2G7b]	4×10^{-75}
	036	5100628	5101681	350	Capsid protein	Major capsid protein [*Escherichia* phage Lambda]	0
	037	5101722	5102118	131	DNA packaging protein	DNA packaging chaperone FI [*Escherichia* phage 1H12]	1×10^{-86}
	038	5102129	5102483	117	Head-Tail attachment protein	Head closure Hc1 [*Escherichia* phage 2B8]	5×10^{-80}
	039	5102494	5103073	192	Tail protein	Tail protein [*Escherichia* phage Lambda]	5×10^{-133}
	040	5103376	5104948	523	Transposase	Transposase [Stx2-converting phage Stx2a_WGPS2]	0
	041	5104967	5105315	115	Transposase	Transposase [Stx2-converting phage Stx2a_WGPS2]	1×10^{-58}
	042	5105314	5105992	225	Transposase	Transposase [Stx2-converting phage Stx2a_WGPS2]	4×10^{-28}

	043	5105968	5106172	67	Tail protein	Tail terminator [*Escherichia* phage ev017]	2×10^{-21}
	044	5106179	5106920	246	Tail protein	Major tail protein [*Escherichia* phage 2G7b]	6×10^{-177}
	045	5106935	5107358	140	Tail assembly chaperone	Minor tail protein g [*Escherichia* phage phiSTEC1575-Stx2k]	1×10^{-100}
	046	5107339	5107774	144	Tail assembly protein	Minor tail protein [Enterobacteria phage O276]	2×10^{-102}
	047	5107766	5110346	859	Tail tape measure protein	Phage tail tape measure protein, Lambda family [*Escherichia* phage ydc107_1]	0
	048	5110342	5110672	109	Tail protein	Phage minor tail family protein [*Escherichia* phage ydc107_1]	1×10^{-76}
	049	5110671	5111370	232	Tail protein	Minor tail protein [*Escherichia* phage 2G7b]	1×10^{-173}
	050	5111375	5112119	247	Tail tip assembly protein	C40 family peptidase [*Escherichia* phage ev017]	5×10^{-176}
	051	5112151	5112658	168	Tail tip assembly protein	Tail tip assembly protein i [*Escherichia* phage vB_EcoM-606r7]	5×10^{-120}
	052	5112718	5116198	1159	Tail protein	Tail tip host specificity protein j [*Escherichia* phage vB_EcoS-640r1]	0
	053	5116265	5116865	199	Outer membrane protein	Host-cell envelope protein [*Escherichia* phage vB_EcoS-640r1]	7×10^{-141}
14 (intact)	**054**	5116865	5117039	57	Hypothetical protein	Hypothetical protein PBV4795_ORF74 [Enterobacteria phage BP-4795]	1×10^{-33}
	055	5117426	5118176	249	Hypothetical protein	Hypothetical protein [*Escherichia* phage phi458]	3×10^{-30}
	056	5118199	5118334	44	Hypothetical protein		
	057	5118294	5119206	303	Tail fiber protein	Tail fiber protein [*Escherichia* phage 2H10]	0
	058	5119283	5120288	334	Tail fiber protein	Tail fiber protein [*Escherichia* phage 2H10]	0
	059	5120287	5120860	190	Tail fiber assembly protein	TPA: tail fiber assembly protein [*Podoviridae* sp. Ct3k57]	6×10^{-135}
	060	5120870	5121416	181	Serine acetyltransferase	TPA: hypothetical protein [*Podoviridae* sp. Ct3k57]	6×10^{-127}
	061	5121732	5121981	82	DNA damage-inducible protein	RecA filament binding protein DinI [Enterobacteria phage cdti]	1×10^{-54}
	062	5122095	5122260	54	Hypothetical protein	TPA: hypothetical protein [Caudoviricetes sp.]	2×10^{-33}
	063	5122212	5123787	524	Peptide chain release factor	Peptide chain release factor 3 [*Acinetobacter* phage MD-2021a]	0
	064	5124179	5124785	201	Osmotically-inducible protein	TPA: periplasmic protein [Caudoviricetes sp.]	3×10^{-91}
	065	5124911	5125073	53	Hypothetical protein	Hypothetical protein phynn_138 [pantoea phage phynn]	7×10^{-11}

Table H – Functional annotation of ETEC strain EC 43. The CDSs sequences were ran through BlastP and HHpred softwares. The functional attribution of CDSs encoded proteins were based on a comparison with a close homolog, considering the E-value, where blank spaces represent irrelevant homolog (E-value ≥ 1×10^{-5}). In the table is also represented the start and stop site (bp) and the size (aa) of the sequences. Functional annotation was not performed on the two questionable prophages (5 and 8).

Prophage	CDS	Start (bp)	End (bp)	Size (aa)	Function	Closest homolog	E-value
1 (intact)	**001**	210385	211293	302	Transcriptional regulator	Transcriptional regulator [Acinetobacter phage MD-2021a]	4×10^{-72}
	002	211393	211983	196	NADPH-quinone reductase	TPA: hypothetical protein [Caudoviricetes sp.]	3×10^{-31}
	003	212065	212862	265	Pyruvate formate lyase	Pyruvate formate-lyase activating enzyme [*Escherichia* phage vB_EcoM-569R5]	0
	004	212894	213889	331	Integrase	Integrase [*Escherichia* phage pro483]	0
	005	214391	214747	118	Transcriptional regulator	Transcriptional regulator [*Escherichia* phage pro483]	2×10^{-84}
	006	214758	214928	56	Hypothetical protein	TPA: hypothetical protein [Caudoviricetes sp.]	4×10^{-35}
	007	214925	215425	166	Replication protein	Putative replication gene B protein [*Yersinia* phage P37]	1×10^{-121}
	008	215489	215713	74	Hypothetical protein	DUF2732 family protein [*Salmonella* phage FSLSP004]	5×10^{-46}
	009	215713	216015	100	Hypothetical protein	Minor tail protein [*Escherichia* phage P2_AC1]	5×10^{-66}
	010	216015	216239	74	Hypothetical protein	Dksa-like zinc-finger protein [Bacteriophage P2]	4×10^{-49}
	011	216236	216511	91	Hypothetical protein	DUF5405 family protein [*Yersinia* phage L-413C]	5×10^{-63}
	012	216501	218777	758	Replication protein	Nicking at origin of replication [*Escherichia* phage vB_EcoM-12474III]	0
	013	219080	219193	37	Hypothetical protein	Hypothetical protein KANBJGNJ_00004 [*Salmonella* phage MET_P1_103_31]	2×10^{-16}
	014	219942	221375	477	ATPase	AAA family ATPase [Vibrio phage 1.202.O._10N.222.45.E8]	2×10^{-41}
	015	221999	223027	342	Portal protein	Portal protein [*Escherichia* phage pro483]	0
	016	223027	224799	590	Terminase large subunit	Terminase large subunit [Bacteriophage P2]	0
	017	224973	225827	284	Capsid scaffolding protein	Head scaffolding protein [Bacteriophage P2]	0
	018	225886	226959	357	Capsid protein	Major head protein [Peduovirus P2]	0

	019	226963	227706	247	Terminase small subunit	Small terminase subunit [Enterobacteria phage 299]	1×10^{-171}
	020	227842	228315	157	Head completion protein	Head-tail adaptor Ad1 [Enterobacteria phage fiAA91-ss]	3×10^{-112}
	021	228315	228518	67	Tail protein	Phage Tail Protein X [*Salmonella* phage STYP1]	5×10^{-20}
	022	228522	228803	93	Holin	Holin [Bacteriophage P2]	2×10^{-58}
	023	228803	229300	165	Endolysin	Endolysin [Bacteriophage P2]	8×10^{-122}
	024	229315	229740	141	Antiholin	Endolysin [*Yersinia* phage P37]	1×10^{-92}
	025	229728	230171	147	Lysis regulatory protein	Rz-like spanin [*Yersinia* phage vB_YpM_46]	6×10^{-101}
	026	230140	230298	52	Lysis system protein	Rz-like spanin [*Escherichia* phage P2]	3×10^{-31}
	027	230261	230728	155	Tail completion protein	Tail terminator [*Escherichia* phage vB_EcoM-12474III]	7×10^{-110}
	028	230721	231173	150	Tail completion protein	Virion morphogenesis protein [*Yersinia* phage L-413C]	2×10^{-106}
	029	231240	231875	211	Baseplate assembly protein	Baseplate assembly protein [Peduovirus P22H1]	1×10^{-150}
1 (intact)	**030**	231872	232219	115	Baseplate wedge subunit	Baseplate wedge subunit [Bacteriophage P2]	2×10^{-80}
	031	232224	233132	302	Baseplate assembly protein	Baseplate protein [*Escherichia* phage Wphi]	0
	032	233125	233736	203	Tail protein	Tail protein [*Escherichia* phage pro483]	5×10^{-146}
	033	233733	235544	603	Tail fibre protein	TPA: Baseplate wedge protein [Caudoviricetes sp.]	0
	034	235544	235954	136	Tail fibre assembly protein	TPA: tail fiber assembly protein [Caudoviricetes sp.]	9×10^{-90}
	035	236055	236228	57	Hypothetical protein		
	036	236285	237475	396	Tail sheath protein	Tail sheath protein [*Escherichia* phage pro483]	0
	037	237488	238006	172	Tail tube protein	Head closure [Bacteriophage P2]	2×10^{-125}
	038	238063	238338	91	Tail protein	Tail protein [*Yersinia* phage L-413C]	9×10^{-59}
	039	238483	240930	815	Tail tape measure protein	Tail length tape measure protein [Peduovirus P24C9]	0
	040	240945	241424	159	Tail protein	Tail protein [Enterobacteria phage fiAA91-ss]	8×10^{-112}
	041	241424	242587	387	Tail protein	Tail protein [Peduovirus P22H1]	0

1 (intact)	**042**	242633	242887	84	Transcriptional regulator	Transcriptional regulator [Peduovirus P22H4]	2×10^{-59}
	043	243206	245488	760	Pyruvate formate lyase	Pyruvate formate-lyase [*Escherichia* phage vB_EcoM-569R5]	0
2 (incomplete)	**001**	2658993	2661422	809	Reductase	TPA: trimethylamine N-oxide reductase I catalytic subunit [Caudoviricetes sp.]	0
	002	2661587	2662558	323	Methyltransferase	TPA: tRNA mo(5)U34 methyltransferase [Caudoviricetes sp.]	2×10^{-155}
	003	2662555	2663298	247	Methyltransferase	TPA: tRNA (cmo5u34)-methyltransferase [Caudoviricetes sp.]	7×10^{-47}
	004	2663339	2663734	131	Hypothetical protein		
	005	2663787	2664566	259	Hypothetical protein	TPA: Protein of unknown function DUF72 [Caudoviricetes sp.]	0
	006	2664563	2665822	419	Integrase	Integrase [*Klebsiella* phage Mulock]	0
	007	2665865	2666056	63	Excisionase	Excisionase [*Klebsiella* phage Mulock]	1E-$\times10^{-41}$
	008	2666132	2666266	44	Hypothetical protein	TPA: hypothetical protein [Caudoviricetes sp.]	5×10^{-26}
	009	2666270	2666605	111	Hypothetical protein	Hypothetical protein [*Escherichia* phage N7]	2×10^{-62}
	010	2666607	2667014	135	Hypothetical protein	Hypothetical protein [*Escherichia* phage phion-2011]	3×10^{-94}
	011	2667312	2667479	55	Hypothetical protein	DUF2737 family protein [*Shigella* phage Sf6]	7×10^{-34}
	012	2667591	2667743	50	Hypothetical protein		
	013	2667822	2668304	160	Host-nuclease inhibitor	Hypothetical protein Kapi1_011 [*Escherichia* phage vB_EcoP_Kapi1]	1×10^{-109}
	014	2668288	2669199	303	DNA recombinase	Recombination protein [*Escherichia* phage vB_EcoP-683R2]	0
	015	2669196	2669504	102	Hypothetical protein	Hypothetical protein GALLYPH_29 [*Escherichia* phage Gally]	9×10^{-71}
	016	2669485	2669607	40	Hypothetical protein	Hypothetical protein Kapi1_014 [*Escherichia* phage vB_EcoP_Kapi1]	8×10^{-18}
	017	2669589	2669741	50	Protein kil	Kil protein for bacterial septation inhibition [*Escherichia* phage HK97]	1×10^{-29}
	018	2669726	2669860	44	Regulatory protein	CIII anti-termination [*Escherichia* phage HK97]	2×10^{-25}
	019	2669944	2670192	82	Hypothetical protein	Anti-restriction protein [*Escherichia* phage HK446]	1×10^{-52}
	020	2670367	2670990	207	DNA modification	TPA: Nucleotide modification associated domain 5 [Caudoviricetes sp.]	4×10^{-150}
	021	2671002	2671328	108	Antitermination protein	Antitermination protein N [*Escherichia* phage HK633]	1×10^{-73}

2 (incomplete)	**022**	2671742	2672779	345	Hypothetical protein	Hypothetical protein [Enterobacter phage ENC13]	3×10^{-102}
	023	2672802	2673425	207	Transcriptional repressor	TPA: SOS-response transcriptional repressors (reca-mediated autopeptidases) [Caudoviricetes sp.]	1×10^{-154}
	024	2674064	2674201	45	Regulatory protein	CII protein [*Escherichia* phage phi458]	1×10^{-25}
	025	2674559	2675446	295	DNA Replication protein	Phage replication protein [Enterobacteria phage CUS-3]	0
	026	2675443	2676819	458	Replicative DNA helicase	Replicative DNA helicase [Enterobacteria phage YYZ-2008]	0
	027	2676934	2677098	54	Hypothetical protein	Hypothetical protein [Enterobacteria phage Lahn2]	8×10^{-27}
	028	2677116	2677436	106	Hypothetical protein	TPA: hypothetical protein [Caudoviricetes sp.]	3×10^{-73}
	029	2677439	2677603	54	Hypothetical protein	TPA: hypothetical protein [Caudoviricetes sp.]	7×10^{-33}
	030	2677581	2678054	157	DNA Recombination protein	Hypothetical protein HK022_61 [*Escherichia* phage HK022]	1×10^{-101}
	031	2678243	2679079	278	Hypothetical protein		
	032	2679345	2680100	251	Peptidase	TPA: lipoprotein [Caudoviricetes sp.]	4×10^{-46}
3 (intact)	**001**	3174598	3175344	248	Dehydrogenase	Bifunctional NAD-dependent-3-hydroxypropionate dehydrogenase [*Klebsiella* phage ST13-OXA48phi12.4]	7×10^{-166}
	002	3175433	3176119	228	Transcriptional regulator	GntR family transcriptional regulator [*Klebsiella* phage ST13-OXA48phi12.4]	3×10^{-137}
	003	3176296	3176499	67	Hypothetical protein	Putative selenium delivery protein YdfZ [*Salmonella* phage SSU5]	5×10^{-18}
	004	3176534	3177994	486	Oxidoreductase	Fructuronate reductase [*Klebsiella* phage ST13-OXA48phi12.4]	0
	005	3178083	3179366	427	Transport protein	TPA: shia-like protein [Caudoviricetes sp.]	0
	006	3180278	3180430	50	Hypothetical protein		
	007	3180703	3181293	196	Serine recombinase	DNA invertase [*Escherichia* phage 2H10]	4×10^{-141}
	008	3181391	3181966	191	Tail fibre assembly protein	Phage tail fiber assembly protein [*Escherichia* phage mEp460_ev081]	1×10^{-139}
	009	3181966	3183675	569	Tail fibre protein	Tail fiber protein [*Salmonella* phage SPF_0923]	0
	010	3183636	3183770	44	Hypothetical protein		
	011	3183806	3184543	245	Hypothetical protein	Tail fiber protein [*Escherichia* phage vB_EcoS-569R4]	2×10^{-88}

	012	3184931	3185104	57	Hypothetical protein	Hypothetical protein PBV4795_ORF74 [Enterobacteria phage BP-4795]	4×10^{-32}
	013	3185105	3185704	199	Outer membrane protein	Lom membrane protein [*Escherichia* phage Perceval]	2×10^{-146}
	014	3185774	3189271	1165	Tail fiber protein	Host specificity protein J [*Escherichia* phage phiSTEC1575-Stx2k]	0
	015	3189332	3189904	190	Tail assembly protein	Tail protein [*Escherichia* phage phiSTEC1575-Stx2k]	2×10^{-133}
	016	3189901	3190536	211	Tail tip assembly protein	Tail protein [*Escherichia* phage phiSTEC1575-Stx2k]	2×10^{-158}
	017	3190650	3191348	232	Tail protein	Minor tail protein [*Escherichia* phage Lambda]	8×10^{-174}
	018	3191348	3191677	109	Tail protein	Minor tail protein [*Escherichia* phage Lambda]	1×10^{-76}
	019	3191674	3194235	853	Tail tape measure protein	Tail length tape measure protein [*Escherichia* phage ev207]	0
	020	3194228	3194662	144	Tail protein	Minor tail protein [Enterobacteria phage 0276]	3×10^{-102}
	021	3194644	3195066	140	Tail protein	Minor tail protein G [*Escherichia* phage phiSTEC1575-Stx2k]	5×10^{-102}
	022	3195082	3195822	246	Tail tube protein	Major tail protein [*Escherichia* phage HK629]	2×10^{-177}
3 (intact)	**023**	3195830	3196225	131	Tail protein	Tail terminator [*Escherichia* phage Lambda]	1×10^{-91}
	024	3196222	3196800	192	Tail protein	Tail protein [*Escherichia* phage Lambda]	1×10^{-133}
	025	3196812	3197165	117	Head-Tail attachment protein	Head closure Hc1 [*Escherichia* phage Lambda]	5×10^{-80}
	026	3197177	3197503	108	DNA packaging protein	DNA packaging protein [*Escherichia* phage 434]	5×10^{-70}
	027	3197545	3198570	341	Capsid protein	Major capsid protein [*Escherichia* phage Lambda]	0
	028	3198627	3198959	110	Head decoration protein	Head decoration protein [*Escherichia* phage ydc107_1]	9×10^{-75}
	029	3198969	3200288	439	Capsid assembly protease	Capsid protease/scaffolding protein [*Escherichia* phage 434]	0
	030	3200269	3201870	533	Portal protein	Portal protein [*Escherichia* phage phiSTEC1575-Stx2k]	0
	031	3201867	3202073	68	Head-to-tail joining protein	Head-tail adaptor ad1 [*Escherichia* phage Lambda]	2×10^{-43}
	032	3202070	3203995	641	Terminase (large subunit)	Terminase large subunit [*Escherichia* phage ev207]	0
	033	3203970	3204515	181	Terminase (small subunit)	Terminase small subunit [*Escherichia* phage ev207]	1×10^{-130}
	034	3204655	3204798	47	Hypothetical protein	TPA: hypothetical protein [Caudoviricetes sp.]	8×10^{-27}

3 (intact)	**035**	3205757	3205930	57	Hypothetical protein	Protein GnsB [*Escherichia* phage Tritos]	8×10^{-33}
	036	3206604	3206816	70	Cold shock protein	Cold shock protein [*Escherichia* phage 2H10]	1×10^{-45}
	037	3207179	3207661	160	Hypothetical protein	DUF2514 protein, putative peptidoglycan endopeptidase [*Escherichia* phage Tritos]	4×10^{-113}
	038	3207673	3208206	177	Lysozyme	Endolysin [*Escherichia* phage Tritos]	5×10^{-131}
	039	3208203	3208514	103	Hypothetical protein	TPA: Protein of unknown function (DUF1327) [Caudoviricetes sp.]	3×10^{-71}
	040	3208519	3208725	68	Holin	Lysis protein S [*Escherichia* phage Tritos]	2×10^{-43}
	041	3209488	3209703	71	Cold shock protein	Cold shock-like protein CspB [*Escherichia* phage Tritos]	5×10^{-45}
	042	3210004	3210216	70	Cold shock protein	Cold shock protein [*Escherichia* phage 2H10]	3×10^{-45}
	043	3210451	3210606	51	Hypothetical protein	TPA: hypothetical protein [Caudoviricetes sp.]	2×10^{-28}
	044	3210638	3211390	250	Antitermination protein	Antitermination protein Q [*Escherichia* phage Tritos]	0
	045	3211404	3212453	349	Hypothetical protein	Phage antitermination protein Q [*Escherichia* phage mEp460_ev081]	0
4 (incomplete)	**001**	3267042	3267701	219	Outer membrane protein	Putative outer membrane protein [Acinetobacter phage MD-2021a]	3×10^{-47}
	002	3267805	3268779	324	Reductase	2-oxo-carboxylic acid reductase [Acinetobacter phage MD-2021a]	1×10^{-88}
	003	3268829	3269539	236	Hypothetical protein		
	004	3269656	3269808	50	Hypothetical protein		
	005	3269973	3270263	96	Hypothetical protein		
	006	3270544	3270756	70	Cold shock protein	Putative portal protein [*Klebsiella* phage pJN2-26]	1×10^{-35}
	007	3270943	3271095	50	Hok/gef family	MokW-like host killing [*Escherichia* phage TL-2011c]	2×10^{-12}
	008	3271175	3271696	173	Transposase	TPA: Mos transposase [Caudoviricetes sp.]	2×10^{-17}
	009	3271693	3272544	283	Transposase	TPA: Mos transposase [Caudoviricetes sp.]	4×10^{-116}
6 (incomplete)	**001**	3788765	3789925	386	Integrase	Site-specific integrase [*Escherichia* phage HK446]	0
	002	3790742	3791020	92	Hypothetical protein	DUF4752 family protein [*Escherichia* phage HK106]	6×10^{-64}
	003	3791020	3792273	417	Hypothetical protein	DUF551 domain-containing protein [Enterobacteria phage mEp460]	7×10^{-103}

6 (incomplete)	**004**	3792403	3792714	103	Hypothetical protein	FIG00639155: hypothetical protein [*Escherichia* virus mEp460_4F5]	3×10^{-65}
	005	3792711	3792935	74	Hypothetical protein	Hypothetical protein PhiV10p47 [*Escherichia* phage phiv10]	1×10^{-48}
	006	3792932	3793099	55	Hypothetical protein	DUF2737 family protein [*Shigella* phage Sf6]	1×10^{-32}
	007	3793096	3793386	96	Hypothetical protein	TPA: protein of unknown function (DUF5405) [Caudoviricetes sp.]	4×10^{-65}
	008	3793397	3793690	97	Hypothetical protein	Abc2 anti-RecBCD [*Escherichia* phage HK633]	3×10^{-68}
	009	3793714	3794097	127	Hypothetical protein	Hypothetical protein F850_gp39 [Enterobacteria phage mEp043 c-1]	1×10^{-90}
	010	3794097	3794702	201	DNA recombination protein	Erf-like ssDNA annealing protein [Enterobacteria phage Sf101]	5×10^{-147}
	011	3794804	3794962	52	Recombinase	Recombinase [Enterobacteria phage Sf101]	4×10^{-31}
	012	3794959	3795111	50	Protein kil	TPA: Kil protein [Caudoviricetes sp.]	6×10^{-30}
	013	3795252	3796271	339	Hypothetical protein	Hypothetical protein F865_gp37 [*Escherichia* phage HK544]	0
	014	3796270	3796404	44	Rz lysis protein	TPA: Rz lysis protein [Caudoviricetes sp.]	7×10^{-24}
	015	3796556	3797152	198	Hypothetical protein	TPA: hypothetical protein [Caudoviricetes sp.]	4×10^{-100}
	016	3797213	3797953	246	Hypothetical protein	TPA: hypothetical protein [Caudoviricetes sp.]	5×10^{-178}
	017	3797957	3799288	443	Terminase large subunit	Putative terminase [*Klebsiella* phage vb_Kpn_Chronis]	0
	018	3799300	3800715	471	Hypothetical protein	Putative DUF1073 domain containing protein [*Klebsiella* phage vb_Kpn_Chronis]	0
	019	3800712	3801533	273	Hypothetical protein	Hypothetical protein CHRON_3 [*Klebsiella* phage vb_Kpn_Chronis]	0
	020	3801546	3803159	537	Hydrolase (nudix)	Putative NUDIX hydrolase [*Klebsiella* phage vb_Kpn_Chronis]	0
7 (intact)	**001**	3817561	3819822	753	Tail fiber protein	Tail fiber protein [*Klebsiella* phage PKP126]	0
	002	3819837	3820595	252	Hypothetical protein	Hypothetical protein BI015_gp60 [*Klebsiella* phage PKP126]	1×10^{-94}
	003	3820681	3821673	330	Tail fiber protein	TPA: tail fiber protein [Caudoviricetes sp.]	0
	004	3822163	3823326	387	Integrase	Integrase [*Shigella* phage SfiV]	0
	005	3823553	3823858	101	Hypothetical protein	Hypothetical protein SfVp28 [Enterobacteria phage SfV]	1×10^{-67}
	006	3823858	3824199	113	Hypothetical protein	Protein [Enterobacteria phage SfV]	4×10^{-80}

7 (intact)	**007**	3824211	3824747	178	Hydrolase	Deoxyribonucleoside 5' monophosphate phosphatase [*Shigella* phage SfiV]	6×10^{-133}
	008	3824875	3825699	274	Hypothetical protein	DUF2303 family protein [*Shigella* phage SfiV]	0
	009	3825765	3826127	120	Hypothetical protein	Hypothetical protein SJJBTUD_0035 [*Escherichia* phage Ayreon]	9×10^{-84}
	010	3826850	3827497	215	Transcriptional regulator	XRE family transcriptional regulator [Enterobacteria phage mEp460]	2×10^{-157}
	011	3827640	3827900	86	Transcriptional regulator	Helix-turn-helix transcriptional regulator [Enterobacteria phage mEp460]	8×10^{-58}
	012	3827893	3828444	183	Hypothetical protein	Hypothetical protein PcdtI_gp43 [Enterobacteria phage cdtI]	6×10^{-133}
	013	3828441	3829592	383	Regulatory protein	Transcriptional regulator [Stx2-converting phage Stx2a_WGPS2]	0
	014	3829589	3829813	74	Regulatory protein	Hypothetical protein PcdtI_gp45 [Enterobacteria phage cdtI]	2×10^{-49}
	015	3829816	3830628	270	Replication protein	Replication protein [Enterobacteria phage SfV]	0
	016	3830868	3831119	83	Transcriptional regulator	Perc family transcriptional regulator [*Shigella* phage SfiV]	2×10^{-52}
	017	3831119	3831772	217	Methyltransferase	N-6-adenine-methyltransferase [*Escherichia* phage Lys8385Vzw]	5×10^{-163}
	018	3831769	3832095	108	Transcriptional repressor	Transcriptional repressor [Enterobacteria phage cdtI]	1×10^{-75}
	019	3832092	3832301	69	Holliday junction resolvase	RusA family crossover junction endodeoxyribonuclease [*Shigella* phage SfII]	1×10^{-42}
	020	3832363	3832482	39	Holliday junction resolvase	Holliday junction resolvase / Crossover junction endodeoxyribonuclease rusA (EC [*Escherichia* phage 2H10]	7×10^{-21}
	021	3832502	3833311	269	Hypothetical protein	Hypothetical protein SJJBTUD_0047 [*Escherichia* phage Ayreon]	0
	022	3833391	3834308	305	Hypothetical protein	Protein YdfU family [*Escherichia* phage vB_EcoM-689R6]	0
	023	3834689	3835150	153	Hypothetical protein	Hypothetical protein H3V23_gp26 [Stx2-converting phage Stx2a_WGPS2]	1×10^{-105}
	024	3835221	3836252	343	Hypothetical protein	Hypothetical protein H3V23_gp27 [Stx2-converting phage Stx2a_WGPS2]	0
	025	3836533	3836859	108	Holin	Holin/anti-holin [*Shigella* phage SfiV]	9×10^{-73}
	026	3836863	3837339	158	Lysozyme	TPA: lysozyme [Caudoviricetes sp.]	1×10^{-113}
	027	3837336	3837797	153	Rz lysis protein	I-spanin [*Escherichia* phage 434]	5×10^{-101}
	028	3837829	3838122	97	Outer membrane protein	Outer membrane lipoprotein complement inhibitor [*Escherichia* phage 21]	1×10^{-65}

7 (intact)	**029**	3838805	3839299	164	Terminase small subunit	Terminase small subunit [*Escherichia* phage ayreon]	3×10^{-116}
	030	3839299	3841401	700	Terminase large subunit	Terminase large subunit [*Escherichia* phage vB_EcoS-813R6]	0
	031	3841398	3841610	70	Head-to-tail joining protein	Putative head-to-tail joining protein W [*Escherichia* phage vB_EcoS-813R6]	6×10^{-46}
	032	3841610	3843085	491	Portal protein	Portal protein B [*Escherichia* phage vB_EcoS-813R6]	0
	033	3843132	3845090	652	Protease	Putative protease [*Escherichia* phage phi458]	0
	034	3845177	3845500	107	Hypothetical protein	Hypothetical protein SJJBTUD_0006 [*Escherichia* phage Ayreon]	7×10^{-71}
	035	3845493	3845768	91	Hypothetical protein	Minor tail protein [Enterobacteria phage mEp460]	1×10^{-61}
	036	3845780	3846358	192	Tail protein	Tail protein [Enterobacteria phage cdtI]	2×10^{-136}
	037	3846355	3846756	133	Tail protein	Tail component U [*Escherichia* phage vB_EcoS-569R4]	7×10^{-96}
	038	3846768	3847511	247	Tail protein	Major tail protein [*Escherichia* phage Ayreon]	2×10^{-177}
	039	3847572	3847958	128	Tail assembly chaperone	Putative tail component [*Escherichia* phage phi458]	5×10^{-90}
	040	3848021	3848296	91	Tail assembly protein	Putative minor tail protein [*Escherichia* phage phi458]	1×10^{-61}
	041	3848268	3851333	1021	Tail length tape measure protein	Phage tail length tape-measure protein 1 [*Escherichia* phage mEp460_ev081]	0
	042	3851333	3851662	109	Tail protein	Minor tail protein [Enterobacteria phage cdtI]	6×10^{-78}
	043	3851672	3852370	232	Tail protein	Minor tail protein [Enterobacteria phage cdtI]	2×10^{-175}
	044	3852484	3853119	211	Peptidase	C40 family peptidase [Enterobacteria phage mEp460]	6×10^{-156}
	045	3853116	3853664	182	Tail assembly protein	Tail assembly protein [Enterobacteria phage cdtI]	6×10^{-126}
	046	3853725	3857204	1159	Tail tip assembly protein	Tail tip host specificity protein J [*Escherichia* phage vB_EcoS-640R1]	0
	047	3857272	3857871	199	Outer membrane protein	Host-cell envelope protein [*Escherichia* phage vB_EcoS-640R1]	3×10^{-143}
	048	3857936	3861694	1252	Tail fiber protein	Side tail fiber protein [*Escherichia* phage vB_EcoS-640R1]	0
9 (incomplete)	**001**	4555570	4557549	659	Restriction-modification system, DNA methylase	TPA: type I restriction-modification system methyltransferase [Caudoviricetes sp.]	6×10^{-152}
	002	4557616	4558113	165	Hypothetical protein		

9 (incomplete)	**003**	4558614	4559273	219	Chloramphenicol acetyltransferase	Type A-1 chloramphenicol O-acetyltransferase [Cloning vector plyss] [*Escherichia* phage vB_EcoP_24B]	4×10^{-168}
	004	4559474	4559851	125	Acetyltransferase	TPA: acetyltransferase domain containing protein [Caudoviricetes sp.]	3×10^{-56}
	005	4559918	4562887	989	transposase	Mobile element protein [*Escherichia* phage P1]	0
	006	4562890	4563447	185	Serine Recombinase	Resolvase [*Escherichia* phage P1]	4×10^{-109}
	007	4563753	4564766	337	Integrase	Integrase [*Escherichia* phage P1]	0
	008	4564926	4565399	157	Dihydrofolate reductase	Dihydrofolate reductase [*Salmonella* phage spastu]	8×10^{-26}