Development of a tool that detects anti-restriction mechanisms in phage genomes

Joana Araújo, Ana Oliveira, Hugo Oliveira, and Óscar Dias

Centre of Biological Engineering, University of Minho, Portugal

Abstract. The emergence of antibiotic-resistant bacteria has become a significant public health concern. In light of the recognized antibiotic potential of bacteriophages, this study aims to elucidate the anti-restriction mechanisms employed by phages to counteract bacterial defense systems and ensure their survival within the host. The project involved the implementation of Python scripts to understand the susceptibility of phage genomes to cleavage by *Escherichia coli* restriction enzymes, allowing for the inference of their prevalence within the host. Additionally, an extensive investigation was conducted to identify and analyze the diverse array of anti-restriction enzymes developed by phages. The findings confirmed that phages have evolved mechanisms involving anti-restriction enzymes to effectively combat bacterial defense systems. Notably, the analysis indicated that T7 phages exhibit a lower probability of being recognized and cleaved by the host's restriction enzymes, suggesting their potential as favorable candidates for phage therapy. Consequently, further exploration of T7-like phages is necessary to expand the pool of viable candidates for effective phage therapy.

Keywords: Bacteriophages · Restriction-Modification System · Anti-restriction mechanisms · Recognition sites · Phage Therapy.

1 Introduction

1.1 Motivation and Context

The increasing diversity of bacterial strains and multidrug-resistant phenotypes prompted the community to seek safer global solutions encompassing environmental, animal, and human health ("One Health" strategy)[1]. The antibacterial potential of bacteriophages, viruses that infect and lyse bacteria, is generally acknowledged[2, 3]. When choosing phages for therapy, several factors must be considered: they must be safe, capable of replicating in the host, and ideally combined in cocktails targeting distinct receptors to limit bacterial resistance[4].

Bacteriophages have proven to be promising alternative or complements for treatment of bacterial infections[4, 5]. However, the effectiveness of phage therapy may be limited by bacterial defense mechanisms, such as restriction-modification systems. In response, bacteriophages have developed their own defense mechanisms to overcome bacterial defense strategies[6].

One of these defense mechanisms employed by bacteriophages is the avoidance of recognition sites for bacterial restriction enzymes. By doing so, phages are able to evade bacterial defense and continue to infect and replicate within the host bacterium[6]. Additionally, bacteriophages have also developed anti-restriction enzymes that protect them against bacterial defense mechanisms[7].

These adaptations demonstrate the ongoing evolutionary arms race between bacteriophages and bacteria, as each seeks to gain the upper hand in the battle for survival. Overall, understanding these defense mechanisms and their interplay is critical for the development of effective phage therapies against bacterial infections.

1.2 Objectives

This project aims to address the challenges and limitations of existing methods in bacteriophage therapy and contribute to the development of more effective and efficient solutions. With this goal in mind, the primary objectives of the project are:

1. Identify bacteriophages that are less probable of cleavage by bacterial restriction enzymes, enabling them to avoid the host bacterium's defense mechanisms. The avoidance of recognition sites of restriction-modification systems is a widespread anti-restriction strategy for prokaryotic viruses. This will involve the screening of a wide variety of bacteriophage genomes and the analysis of their proteome, in order to identify potential candidates that are not recognized and, thereby, not affected by restriction enzymes.
2. Identify *Enterobacteriaceae* bacteriophages encoding anti-restriction enzymes which can protect them from bacterial defense systems. The results of this research will be compiled into a document, that can be later used to construct a comprehensive database, in order to inform future studies in the field of bacteriophage therapy.

Ultimately, this may allow a stricter bacteriophage selection envisioning bacteriophage therapy effectiveness.

2 Background and State-of-the-Art

2.1 Antibiotic Resistance

Antimicrobial resistance is on the rise worldwide, pressing the healthcare services. Longer hospital stays, higher medical expenses and increased mortality may result from the increasing occurrence of difficult-to-treat infections. Each year, approximately 700.000 deaths are recorded worldwide as a result of multidrug-resistant infections, and a death toll of up to 10 million people per year is expected by the year of 2050[8].

2.2 Bacteriophages

Bacteria with antibiotic resistance mechanisms have an edge in the selection, particularly in environments that are advantageous for them, namely clinical settings. Given the problem of the increased emergence of infection and drug-resistant pathogens in science and public health[9, 10], bacteriophages have been considered important candidates for alternative or complementary therapy[11]. Bacteriophages, or just phages, are naturally occurring viruses that are isolated from different ecological specimens, namely sewage, water and soil[12–14].

In their mechanism of action, phages recognize specific receptors from the bacterial cell wall (adsorption), inject their DNA, and use the bacteria's biochemical machinery to produce multiple new copies and assemble new phages. After reaching the bacterium's "critical mass", the bacterial cell wall lyses and the releasing of the new phage progeny is triggered, to reinitiate a new lytic cycle (**Fig. 1a.)** [15, 16].

Several advantages arise from the usage of bacteriophages. Besides specificity, targeting exclusively bacteria, avoiding the surrounding microbiota[17], phages can also self-amplify, and be self-limiting in the lack of the targeted bacteria [4, 17].

2.3 Restriction-Modification Systems

The entire history of life is the tale of host-parasite co-evolution, with the continuous "arms race" playing a crucial role [18, 19]. Due to the arms race with parasites, cells developed highly complex and diverse defense systems that operate through distinct strategies[20]. The defense mechanisms of bacteria and archaea can be classified into three general categories, based on their principle of action: 1) resistance based on virus receptor variation; 2) immunity, and 3) dormancy induction and programmed cell death[21, 22]. Particularly for immunity, from a wide range of mechanisms bacteria have, Restriction-Modification (R-M) systems are one of the most important means to acquire resistance to phages.

R-M systems work as prokaryotic immune systems, destroying foreign DNA that enters the cell[23]. These systems typically contain enzymes that perform two opposing functions: a restriction endonuclease (REase) that is responsible for the recognition of a specific DNA sequence for cleavage, and a cognate methyltransferase (MTase) that confers cleavage resistance to bacterial DNA by methylating adenine or cytosine bases within the same recognition sequence. The absence of characteristic modification within specific recognition sites allows REases to identify foreign (or "non-self") DNA (**Fig. 1b.)**, such as that of phage and plasmids[24]. Endonucleotidic cleavage is then used to inactivate foreign DNA. Following replication, host DNA generally becomes methylated by the MTase, whereas invading non-self DNA is not.

The R-M systems can be divided into three main restriction-modification enzyme types: Type I, containing two REases subunits required for DNA cleavage using ATP, one Specifity subunit (S), that specifies the DNA sequence that is

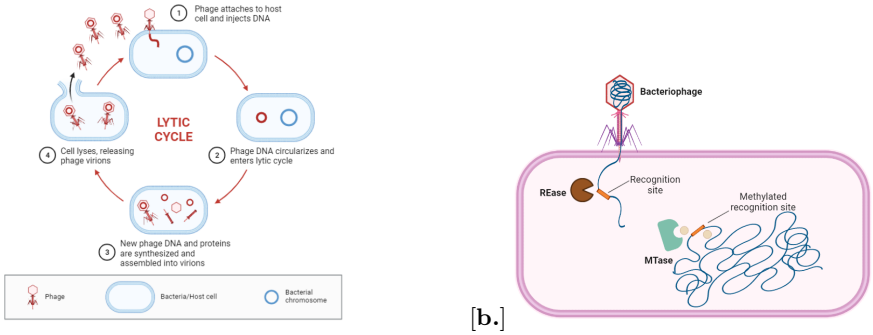

Fig. 1. [a.] Bacteriophage lytic cycle. The bacteriophage attaches itself to the specific bacterial membrane and injects its genetic material into the bacterium, which will replicate via bacteria enzymes and also synthesize proteins and lipids needed to form capsids. After assembling these different components to form virions, the bacteria is lysed, releasing new phages, which can attach themselves to new bacteria and start the cycle all over again. **[b.] Restriction-Modification system.** Recognition sequences are recognized as non-self DNA and cleaved by the restriction endonuclease (REase). The methylated sequences are recognized as self-DNA to protect themselves against cleavage. This methylation status is maintained by the cognate methyltransferate (MTase) on the R-M system.

being recognized, and two MTase subunits (M) which are responsible for the catalyzes of the methylation reaction[19, 25]; Type II enzymes, based on cleavage specificity[26] and have REase and MTase activities, and are usually dependent on Mg2+ and SAM, respectively; Type III enzymes, hetero-oligomers comprising a "mod" subunit that is mandatory for substrate recognition and modification, and also a restriction subunit "res" that is solely active when combined with the "mod" subunit on a res2mod2 complex[27]. These later complexes maintain contact with their specific recognition sequence in order to proceed with the cleavage[25, 28].

2.4 Anti-Restriction System

Counteracting the system described above, the ability of phages to acquire host modification and protect themselves from restriction indicates that R-M systems are not flawless, and emphasizes an evolutionary arms race between bacterial genomes (R-M systems) and parasitic DNA molecules (anti-restriction systems)[25]. Some of the anti-restriction mechanisms of bacteriophages are described below:

- **Changes in DNA sequence:** this process involves some different strategies as:
  - Removal of recognition sites from phage genomes, allowing them to evade restriction. Sometimes, a reduction of recognition sites is enough to enable phage to avoid restriction[29, 30]. For example, the restriction en-

donuclease EcoRII, which requires the coordinated presence of at least two recognition sites for the cleavage to occur, cannot proceed to cleave DNA in T3 and T7 phages because the distance of EcoRII sites is very large[29, 31]. This can show that it's greatly likely that these phages have lost their restriction sites as a result of counter-selection imposed by host R-M systems[32].
  - Strand-biased asymmetrical sequences, where recognition sites are in the same orientation, instead of "head-to-head" formation, which is demanded for the cleavage to occur. For example, EcoP1I sites in T7 phage genomes contain that feature, enabling the phage to avoid restriction[33].
  - Integration of unusual DNA bases in bacteriophages' genome. As an example, many *Bacillus subtilis* bacteriophages replace thymine with 5-hydroxy-
    methyluracil[34].
  - Encoding of MTases to modify and protect phage genome. For example, SPβ phage encodes and MTase that modifies bases in *B. subtilis* BsuRI recognition sites[34].
- **Inhibition of R-M enzymes by the development of anti-restriction enzymes:** The most widely studied anti-restriction protein, the gene 0.3 protein of phage T7 (also known as the overcome classical restriction [Ocr] protein), inhibits Type I R-M systems through direct interaction. Ocr is the first product expressed by T7-phage as it infects a bacterium[7, 35]. It obstructs the DNA binding location of resident Type I R-M systems, allowing the phage to multiply[36]. In (**Table 1**) we can observe some anti-restriction enzymes present in *Enterobacteriaceae* phages.

There are currently no phage therapy products authorized for human use neither in the European Union nor the United States. However, in the food industry, there are several commercial phage preparations used for the biocontrol of bacterial pathogens that have been approved by the Food and Drugs Administration (FDA) and classified as "generally considered as safe"[37]. These preparations have been used to control foodborne pathogens such as *Salmonella spp.*, *Escherichia coli* O157:H7, *Mycobacterium tuberculosis*, *Listeria monocytogenes*, *Pseudomonas syringae*, among other bacteria[38–41].

The published literature on the use of phage and phage-derived proteins for combating bacterial infections, particularly those caused by multidrug-resistant bacteria, shows growing promise for phage therapy as an alternative or complement to antibiotics. Yet, inconsistencies in recent findings on immunomodulatory effects, host range, and the potential gene transfer make it obvious that before implementing phage therapy on a large scale, there is a need for a better understanding of the interaction between phage, microbiome and human host[37].

3 Methodology

In order to accomplish the stipulated objectives, a well-defined methodology was implemented to guide our study, incorporating algorithms and specialized tools to analyze and explore the complexities of our subject matter.

Table 1. Examples of anti-restriction enzymes and respective functions of bacteriophages that infect Enterobacteria. This information was retrieved from the UniProt database.

UniProt Entry	Protein names	Organism	Function
P04392	DNA adenine methylase	Enterobacteria phage T4	Recognizes the double-stranded sequence 5'-GATC-3' and methylates it
P04519	DNA alpha-glucosyltransferase	Enterobacteria phage T4	Protects the phage genome against its own nucleases
P04547	DNA beta-glucosyltransferase	Enterobacteria phage T4	Protects the phage genome against its own nucleases and CRISPR-Cas9 defense system
P08794	Methylcarbamoylase mom	Escherichia phage D108	Iron-binding protein turning viral DNA resistant to several host type I and II RE by modifying some of the DNA adenine residues
P12427	DNA adenine methylase	Enterobacteria phage T2	Recognizes the double-stranded sequence 5'-GATC-3' and methylates A-2 on both strands
Q38156	DNA N-6-adenine-methyltransferase	Escherichia phage T1	Methylates adenine residues in the dsDNA sequence 5'-GATC-3', preventing degradation of viral DNA by the host R-M systems
P03703	Restriction inhibitor protein	Escherichia phage Lambda	Interferes with DNA cleavage by *E.coli* EcoK R-M system.
P03775	Protein Ocr	Escherichia phage T7	Prevents both degradation and modification of T7 DNA by the host restriction-modification complex
P39510	Anti-restriction endonuclease	Enterobacteria phage T4	Inhibition of the host R-M system by cleavage of 5-methyl and 5-hydroxymethylcytosines
P62765	T4 Suppressor of prr	Enterobacteria phage T4	Inactivation of several host DNA R-M systems including EcoprrI or EcoR124I

3.1 Identifying phages that use the avoidance strategy

To identify bacteriophages that can evade bacterial restriction enzymes, we propose to use the REBASE database, which is widely used in molecular biology and genetics. This database contains a vast collection of references, recognition and cleavage sites, sequences, and structures of restriction enzymes. By selecting bacterial restriction enzymes and comparing their cutting sequence with phage genomes, we aim to identify phages that will potentially be affected.

Conversely, phages that exhibit avoidance strategies are expected to lack the recognition sequence of the corresponding bacterial restriction enzyme. This approach will enable us to determine the prevalence of phages employing this type of primary restriction-modification defense system.

3.1.1. Tools and implemented algorithm Initially, the genomes of bacteriophages were acquired from the Nucleotide database of the National Center for Biotechnology Information (NCBI). These phages were isolated from *Escherichia coli*, a bacterial specie widely investigated in the field of microbiology.

The developed Python code, presented on GitHub, implements an algorithm for analyzing a nucleotide sequence using restriction enzymes. The algorithm aims to identify specific sites in the sequence where these enzymes can cut.

At the beginning of the code, essential modules are imported, including 'sys' for command-line arguments and 're' for regular expressions. These modules provide the necessary functionality for handling input/output, pattern matching and data processing.

The code defines several functions that play key roles in the algorithm. The 'format' function takes a nucleotide sequence and organizes it into a specific structure, ensuring a specified number of base pairs per group and a certain number of characters per line. This formatting is important as it enhances the readability and organization of the sequence, making it easier to analyze and interpret.

Next, the function 'lines_to_line' joins multiple lines of a FASTA file containing the nucleotide sequence into a single continuous string. This consolidation of the sequence simplifies subsequent operations and enables a more streamlined analysis.

The main body of the code revolves around reading input files and processing the nucleotide sequence. It reads the sequence and extracts relevant information such as the sequence name and length. The output later includes this information to provide context and reference for the analyzed sequence.

Then it proceeds to read the restriction enzymes and their corresponding cutting patterns from a separate file, by employing the 'get_pattern' function to handle ambiguous nucleotides in the enzyme patterns, allowing for flexibility in recognizing variations in the sequence. The script searches for matches between the enzyme patterns and the nucleotide sequence using regular expressions, identifying potential cutting sites.

If cutting sites are found, information about the number of cutting sites and the resulting sequence fragments are written to an output file. These fragments, representing distinct portions of the sequence resulting from the enzyme cuts, are appropriately formatted using the 'format' function to ensure readability and clarity.

Throughout the analysis, the code keeps track of each enzyme's cutting sites, using a dictionary called 'enzyme_counts'. This tracking allows for subsequent visualization and analysis of the enzyme distribution and cutting frequencies.

Towards the end, the results are summarized and visualized. This includes identifying the enzymes with the most and least cutting sites, grouping enzymes based on the number of cuts they generate (no cuts, 1 to 10, 10 to 100 or more than 100), and generating a result file that provides counts for each group. These visualizations and summary statistics provide insights into the characteristics of

the analyzed nucleotide sequence and the behavior of the different restriction enzymes.

This approach facilitates the identification of phage prevalence within the bacterial population. If the phages are extensively targeted and cleaved by bacterial enzymes, they are unlikely to persist in the host and thus, not suitable candidates for phage therapy. Conversely, phages exhibiting limited or no cleavage events are more likely to prevail and might be considered promising candidates.

This process will help determine the presence or absence of recognition sites for the stored restriction enzymes, and accelerate the identification of bacteriophages that can evade bacterial defense mechanisms. Besides, this will facilitate the creation of a comprehensive database of such phages. Ultimately, this research will contribute to the development of more effective bacteriophage therapies for the treatment of bacterial infections.

3.2 Anti-restriction enzyme search and Database development

In order to identify bacteriophages that employ anti-restriction enzymes as a mean to evade bacterial defense systems, our proposed approach involves conducting a systematic search for related proteins within the UniProt database. UniProt is a widely recognized and extensively utilized resource that serves as a centralized repository for comprehensive protein sequence data, as well as detailed information pertaining to protein function, interaction, and expression.

3.2.1. Tools and Keyword search By employing specific keywords and filters to refine our search criteria, we aim to extract pertinent information from the UniProt database, focusing on the organisms and functions of interest, respectively *Escherichia coli* and 'Restriction-modification system evasion by virus'. This comprehensive analysis will enable us to identify and compile relevant data regarding bacteriophages and their utilization of anti-restriction enzymes, contributing to a deeper understanding of their mechanisms for evading bacterial defenses.

Overall, this approach will enable us to systematically identify and record phage anti-restriction enzymes and provide a valuable resource for the development of more effective phage therapies for the control of bacterial infections.

4 Results

4.1 Screening of phages' genome

Using the developed Python scripts, and following the analysis of 48 *E. coli* phage genomes, it was observed that the enzyme EcoBLMcrX exhibited the highest number of cleavage sites within the phage genomes. Upon analyzing its recognition sequence, it became evident that the sequence contained several ambiguous nucleotides, such as "R", which can encode either an "A" or "G", and "S", which encodes "C" and "G", enabling it to recognize a diverse range

of DNA sequences and expanding the repertoire of recognized sequences within phage genomes.

This finding demonstrates that enzymes with a higher number of ambiguous nucleotides in their recognition sequences possess a greater capacity for recognition and cleavage within phage genomes, thereby extending phages' potential to be recognized in a much broader range of DNA sequences.

The presence and prevalence of the EcoBLMcrX recognition sequence in phage genomes suggest its significant role in shaping the dynamics of these genomes. The cleavage activity facilitated by the recognition sequence of EcoBLMcrX may contribute to the adaptive processes and survival of phages in the host.

This observation highlights the ability of enzymes with a greater number of ambiguous nucleotides in their recognition sequences to exhibit enhanced recognition and cleavage capabilities within phage genomes. Such enzymes broaden the range of sequences that can be recognized, thus influencing the dynamics and diversity of phage genomes. Further research is warranted to elucidate the functional implications of these findings and their implications for phage-host interactions and evolutionary processes in *E. coli*.

4.2 T-like phages

In the subsequent phase of our investigation, we focused on T-like phages and classified the percentage of enzymes capable of cleaving these phages into distinct categories: no cleavage (category 1), 1 to 10 cleavages (category 2), 10 to 100 cleavages (category 3), and over 100 cleavages (category 4). Our primary aim was to determine which phage exhibited the highest proportion of enzymes that performed minimal cleavages, specifically falling into categories 1 or 2 (**Figure 2**).

Upon analyzing the graphical representation, it becomes evident that, among a total of 404 enzymes, approximately 10% of them did not execute any cleavages on the T7-phage. Moreover, around 70% of the enzymes were found to be responsible for only 1 to 10 cleavages, while a mere 20% demonstrated the capability to induce more than 11 cleavages within this particular phage. Consequently, it is reasonable to suggest that phages of this nature may possess a reduced susceptibility to restriction-modification systems, thereby rendering them highly promising candidates for phage therapy (considering this restriction mechanism only). The limited extent of cleavage observed suggests their potential to effectively infect and successfully replicate within the host.

In contrast, T2, T4, T5, and T6-phages exhibited a substantial percentage (approximately 50%) of enzymes that had over 100 cleavages in their genomes. This notable finding implies that the genomes of these phages are likely to be readily recognized and cleaved by the host bacterium's restriction enzymes. Consequently, such phages can be considered unpromising candidates for phage therapy, due to the heightened vulnerability resulting from extensive enzymatic cleavage.

These findings underscore the differential susceptibility of T-like phages to restriction-modification systems.

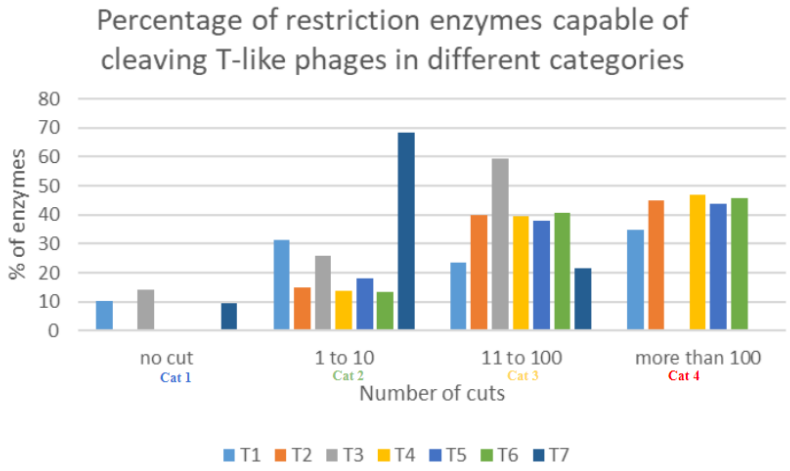

Fig. 2. Graphical representation of the percentage of enzymes performing different numbers of cleavages on T phages, categorized into four groups: no cleavage, 1-10 cleavages, 10-100 cleavages, and over 100 cleavages.

4.3 Presence of anti-restriction enzymes

Through our examination of the UniProt database, we identified a diverse range of anti-restriction enzymes associated with *E. coli* phages. This observation indicates a high degree of genetic diversity within phage populations. The presence of various anti-restriction enzymes suggests that different phages have evolved different strategies to counteract the host bacterium's restriction-modification systems.

One of the main findings was the Ocr protein, an anti-restriction enzyme found in T7-phages, which primary function is to counteract the host bacterium's restriction-modification systems. The Ocr protein accomplishes its role by mimicking the structure of DNA and binding to the Type I restriction enzymes, thereby preventing them from recognizing and cleaving the phage DNA. This interaction effectively neutralizes the host's defense system, allowing the T7-phage to evade restriction and proceed with its replication cycle within the host bacterium.

Another notable anti-restriction enzyme that emerged from our investigation is the ArdA protein. Found in certain strains of *E. coli*, ArdA serves as a potent defense mechanism against the host bacterium's restriction-modification systems. Unlike the Ocr protein in T7-phages, ArdA operates through a different mechanism. It interferes with the Type I restriction enzymes by binding to them and altering their activity, inhibiting their ability to recognize and cleave the phage DNA. By disrupting the host's restriction machinery, ArdA enables the phage to evade the defense mechanisms and establish successful infections within the *E. coli* host.

The identification of these anti-restriction enzymes highlights the ongoing arms race between phages and bacteria. Phages constantly evolve and adapt to evade the host's defense mechanisms, while bacteria develop new strategies to defend against phage attacks. The diversity of anti-restriction enzymes reflects the genetic battles and ongoing evolutionary dynamics between phages and their hosts.

5 Conclusion and Future Work

In conclusion, this study has provided valuable insights into the susceptibility of phages to bacterial enzyme cleavage, allowing for the determination of their prevalence within the host. Furthermore, the identification of phage anti-restriction enzymes has emerged as a critical factor for phage survival and persistence within the bacterial host.

Our findings highlight the heterogeneous nature of phage susceptibility to restriction enzyme cleavage. Notably, T2, T4, T5, and T6-like phages demonstrated a higher propensity for recognition and cleavage by restriction enzymes. Conversely, T7-like phages exhibited a diminished susceptibility to enzyme recognition, thereby enhancing their potential for sustained presence within bacterial populations. Consequently, T7-like phages may hold promise as viable candidates for phage therapy against *E. coli*-related diseases, considering the role of restriction mechanisms only. However, it is imperative to conduct further extensive laboratory investigations to confirm and validate these observations.

This research underscores the significance of understanding the interplay between phages and bacterial restriction-modification systems. The knowledge gained from studying phage susceptibility to restriction enzymes can inform the development of targeted phage-based therapeutic interventions, ultimately contributing to the mitigation of *E. coli* infections. Future investigations should explore additional phage-host interactions and conduct comprehensive laboratory analyses to deepen our understanding of phage biology and optimize their clinical applications.

This study also underlines that anti-restriction enzymes play a critical role in the successful replication and survival of phages within the host organism. By countering the host's defense mechanisms, these enzymes allow phages to overcome barriers and effectively propagate their genetic material. This underscores the importance of anti-restriction enzymes in the lifecycle of phages and their ability to navigate the hostile environment within the host.

The characterization and understanding of anti-restriction enzymes in *E. coli* phages have implications for phage therapy. The identification of these enzymes provides valuable insights into the potential success or limitations of using phages as therapeutic agents against bacterial infections. By deciphering the mechanisms employed by phages to overcome host defense, we can enhance the development of effective phage-based treatments.

The successful completion of the defined objectives in this phase of our work opens up several promising prospects for future exploration, which have the

potential to enhance the scope of our research and elevate the developed tool to greater heights.

One compelling prospect for further investigation involves the establishment of a comprehensive database. Currently, our approach relies on the utilization of *E. coli*-derived enzymes, assuming their universal efficacy in cleaving phages. However, for a more realistic understanding, it is essential to discern the specific repertoire of restriction enzymes present within the target bacterial strain. This would demand conducting a genomic analysis of the bacterial genome, enabling the identification of the most prevalent and biologically relevant restriction enzymes. Such a comprehensive database would provide more precise insights into the understanding of the interplay between enzymes and phages.

Another interesting prospective for exploration lies in the identification of phage-encoded anti-restriction enzymes. This would involve the analysis of the genomes of the phages under scrutiny. Of particular interest is the investigation of these enzymes in T phages, given their high degree of genetic similarity to T-like phages. By extending our findings from T phages to their T-like counterparts, which share a genetic resemblance of over 95%, we can assemble a more expansive and inclusive database without compromising its reliability.

These future endeavors hold significant promise, as they have the potential to refine our approach and deepen our understanding of the dynamics between bacteria and phages. By pursuing these ideas, we can advance our research, leading to more effective strategies for combating bacterial infections and informing the development of targeted therapeutic interventions.

6 Code availability

The code developed for this research work has been entirely implemented in Python and is freely available on GitHub. It can be accessed through the following GitHub link: https://github.com/jaraujo09/Project1stYear.

References

1. H. K. Allen, M. Merabishvili, E. M. Kutter, G. Bao, J. Zhao, Y. Liu, C. Xiao, S. He, and H. Yao, "Efficacy of phage therapy in controlling rabbit colibacillosis and changes in cecal microbiota," 2017.
2. T. Azimi, M. Mosadegh, M. J. Nasiri, S. Sabour, S. Karimaei, and A. Nasser, "Phage therapy as a renewed therapeutic approach to mycobacterial infections: a comprehensive review," 2019.
3. J. D. Nosanchuk, D. V. Duin, R. Aminov, S. T. Abedon, P. García, and P. Mullany, "Editorial: Phage therapy: Past, present and future," *Frontiers in Microbiology — www.frontiersin.org*, vol. 1, p. 981, 2017.
4. C. Loc-Carrillo and S. T. Abedon, "Pros and cons of phage therapy," *Bacteriophage*, vol. 1, pp. 111–114, 3 2011.
5. C. Rohde, J. Wittmann, and E. Kutter, "Invited reviews bacteriophages: A therapy concept against multi-drug-resistant bacteria,"

6. S. J. Labrie, J. E. Samson, and S. Moineau, "Bacteriophage resistance mechanisms," *Nature Reviews Microbiology 2010 8:5*, vol. 8, pp. 317–327, 3 2010.
7. D. H. Kruger, C. Schroeder, M. Santibanez-Koref, and M. Reuter, "Avoidance of dna methylation a virus-encoded methylase inhibitor and evidence for counterselection of methylase recognition sites in viral genomes," *Cell Biophysics*, vol. 87, pp. 9–9.
8. S. Wall, "Prevention of antibiotic resistance – an epidemiological scoping review to identify research categories and knowledge gaps," 2019.
9. J. Y. Sharahi, T. Azimi, A. Shariati, H. Safari, M. K. Tehrani, and A. Hashemi, "Advanced strategies for combating bacterial biofilms," 2019.
10. T. K. Lu and J. J. Collins, "Engineered bacteriophage targeting gene networks as adjuvants for antibiotic therapy,"
11. U. Bajpai and A. K. Mehta, "Isolation and characterization of bacteriophages from india, with lytic activity against mycobacterium tuberculosis. 2 3."
12. S. Guo and Z. Ao, "Phage in the diagnosis and treatment of tuberculosis," 2012.
13. A. Nasser, R. Azizian, M. Tabasi, J. K. Khezerloo, F. S. Heravi, T. Kalani, N. Sadeghifard, R. Amini, I. Pakzad, A. Radmanesh, and F. A. Jalilian, "Article history: Specification of bacteriophage isolated against clinical methicillin-resistant staphylococcus aureus osong public health and research perspectives," *Public Health Res Perspect*, vol. 10, pp. 20–24, 2019.
14. R. Azizian, A. Nasser, H. Askari, M. T. Kalani, N. Sadeghifard, I. Pakzad, R. Amini, A. S. M. Nejad, and F. A. Jalilian, "Sewage as a rich source of phage study against pseudomonas aeruginosa pao," *Biologicals*, vol. 43, pp. 238–241, 7 2015.
15. M. G. Weinbauer, "Ecology of prokaryotic viruses," *FEMS Microbiology Reviews*, vol. 28, pp. 127–181, 2004.
16. R. Young, "Phage lysis: do we have the hole story yet?," *Current Opinion in Microbiology*, vol. 16, pp. 790–797, 12 2013.
17. A. Forde and C. Hill, "Phages of life – the path to pharma," *British Journal of Pharmacology*, vol. 175, pp. 412–418, 2 2018.
18. P. Forterre and D. Prangishvili, "The great billion-year war between ribosome-and capsid-encoding organisms (cells and viruses) as the major source of evolutionary novelties,"
19. E. V. Koonin and Y. I. Wolf, "Evolution of microbes and viruses: a paradigm shift in evolutionary biology?," 2012.
20. E. V. Koonin, K. S. Makarova, and Y. I. Wolf, "Evolutionary genomics of defense systems in archaea and bacteria *," 2017.
21. K. S. Makarova, V. Anantharaman, L. Aravind, and E. V. Koonin, "Live virus-free or die: coupling of antivirus immunity and programmed suicide or dormancy in prokaryotes," *Biology Direct*, vol. 7, p. 1, 2012.
22. K. S. Makarova, Y. I. Wolf, and E. V. Koonin, "Comparative genomics of defense systems in archaea and bacteria," *SURVEY AND SUMMARY*.
23. G. Bertani and J. J. Weigle, "Host controlled variation in bacterial viruses',"
24. A. Pingoud and A. Jeltsch, "Structure and function of type ii restriction endonucleases," *SURVEY AND SUMMARY*, vol. 29, pp. 3705–3727, 2001.
25. D. T. F. Dryden, N. E. Murray, and D. N. Rao, "Nucleoside triphosphate-dependent restriction enzymes," *SURVEY AND SUMMARY*, vol. 29, pp. 3728–3741, 2001.
26. R. J. Roberts, M. Belfort, T. Bestor, A. S. Bhagwat, T. A. Bickle, J. Bitinaite, R. M. Blumenthal, S. K. Degtyarev, D. T. F. Dryden, K. Dybvig, K. Firman, E. S.

Gromova, R. I. Gumport, S. E. Halford, S. Hattman, J. Heitman, D. P. Hornby, A. Janulaitis, A. Jeltsch, J. Josephsen, A. Kiss, T. R. Klaenhammer, I. Kobayashi, H. Kong, D. H. K. ger, S. Lacks, M. G. Marinus, M. Miyahara, R. D. Morgan, N. E. Murray, V. Nagaraja, A. Piekarowicz, A. Pingoud, E. Raleigh, D. N. Rao, N. Reich, V. E. Repin, E. U. Selker, P.-C. Shaw, D. C. Stein, B. L. Stoddard, W. Szybalski, T. A. Trautner, J. L. V. Etten, J. M. B. Vitor, G. G. Wilson, and S. yong Xu, "A nomenclature for restriction enzymes, dna methyltransferases, homing endonucleases and their genes," *SURVEY AND SUMMARY Nucleic Acids Research*, vol. 31, 2003.

27. J. P, S. U, S. MD, and B. TA, "Subunit assembly and mode of dna cleavage of the type iii restriction endonucleases," *J Mol Biol*, vol. 306, pp. 417–431, 2001.
28. R. S, G. I, R. M, R. JP, and K. DH, "Scanning force microscopy of dna translocation by the type iii restriction enzyme ecop15i," *Journal of Molecular Biology*, vol. 341, pp. 337–343, 2004.
29. D. H. Kruger and T. A. Bickle, "Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts," *Microbiological Reviews*, vol. 47, pp. 345–360, 9 1983.
30. T. A. Bickle' and D. H. Kruger2, "Biology of dna restriction," *Microbiological Reviews*, vol. 57, pp. 434–450, 6 1993.
31. D. H. Kruger, G. J. Barcakl, M. Reuter, and H. O. Smithl, "Nucleic acids research ecorii can be activated to cleave refractory dna recognition sites," vol. 16, 1988.
32. M. R. Tock and D. T. Dryden, "The biology of restriction and anti-restriction," 8 2005.
33. A. Meisel, T. A. Bickle, D. H. Kriiger, and C. Schroeder, "Type iii restriction enzymes need two inversely oriented recognition sites for dna cleavage," *Nature 1992 355:6359*, vol. 355, pp. 467–469, 1992.
34. W. RA, "Modified bases in bacteriophage dnas," *Annual Review of Microbiology*, vol. 34, pp. 137–158, 1980.
35. F. W. Studier, "Analysis of bacteriophage t7 early rnas and proteins on slab gels," *Journal of Molecular Biology*, vol. 79, pp. 237–248, 9 1973.
36. P. K. Bandyopadhyay, F. W. Studier, D. L. Hamilton, and R. Yuan, "Inhibition of the type i restriction-modification enzymes ecob and ecok by the gene 0.3 protein of bacteriophage t7," *Journal of Molecular Biology*, vol. 182, pp. 567–578, 4 1985.
37. D. Lin, B. Koskella, and H. Lin, "Phage therapy: An alternative to antibiotics in the age of multi-drug resistance," *World J Gastrointest Pharmacol Ther*, vol. 8, pp. 162–173, 2017.
38. A. B. Monk, C. D. Rees, P. Barrow, S. Hagens, and D. R. Harper, "Bacteriophage applications: where are we now?," *Letters in Applied Microbiology*, vol. 51, pp. 363–369, 2010.
39. R. Nannapaneni and K. Soni, "Use of bacteriophages to remove biofilms of listeria monocytogenes and other foodborne bacterial pathogens in the food environment," 2015.
40. L. Endersen, J. O'mahony, C. Hill, R. P. Ross, O. Mcauliffe, and A. Coffey, "Phage therapy in the food industry," 2014.
41. D. A. Schofield, N. J. Sharp, J. Vandamm, I. J. Molineux, K. A. Spreng, C. Rajanna, C. Westwater, and G. C. Stewart, "Bacillus anthracis diagnostic detection and rapid antibiotic susceptibility determination using 'bioluminescent' reporter phage," *Journal of Microbiological Methods*, vol. 95, pp. 156–161, 11 2013.