Bacteriophages for gut-associated bacteria

Miles Oladapo Oke

1st October 2023

This thesis is submitted for the Master of Science by Research.

University of East Anglia

Quadram Institute Biosciences

Abstract

The human gut microbiome hosts a complex and dynamic community of microorganisms that have been shown to play key roles in many biological processes. Amongst these microorganisms are bacteriophages (hereafter also referred to as phages), bacterial-infecting viruses that drive bacterial composition and diversity. Phages have seen a renewed interest within the last two decades as an alternative treatment for combatting antimicrobial resistance (AMR). Much remains to be discovered about phage dynamics, their interactions within the human microbiome and more. They currently represent promising potential in phage therapy against bacterial infections, AMR, and food biocontrol. Bacteriophages, and the wider gut virome, represent a pivotal, albeit complicated, part of shedding light on the human gut microbiome and delivering new therapies for human health.

The aim of my research described in this thesis was two-fold: (1) to isolate and characterise novel bacteriophages for *Enterobacter cloacae* (*E. cloacae*) NC10005 and (2) to apply the previous methodology to isolate and characterise novel bacteriophages for *Ruminococcus gnavus* (*R. gnavus*), and to use these phages to investigate their use in treating associated bacterial infections and modulating the gut microbiota.

In summary, I have isolated and characterised three phages targeting *E. cloacae* NC10005: MO1, MO2 and MO3. I have also characterised these phages by demonstrating an expanded host-range; assembling and annotating their genomes; proposing taxonomy and undergoing a comparative genomics analysis. I have also developed a protocol for the isolation of phages for *R. gnavus*, of which I used to successfully isolate four phages targeting *R. gnavus* CC55_001C: MOR1, MOR2, MOR3, and MOR4. The work undertaken provides scope for expanding knowledge of bacteriophages of *E. cloacae* and providing a foundation for future isolation of phages targeting *R. gnavus* and anaerobic bacteria.

Contents

List of Figures

1 Introduction

4 Results

List of Tables

4 Results

Acknowledgements

A number of people have helped me during my Master's that I am grateful for and would like to acknowledge and thank.

Firstly, I would like to thank my primary supervisor, Dr. Evelien Adriaenssens, for giving me the opportunity to research within her lab and all the opportunities that came during my time. Each one was a chance for me to develop and grow both academically and personally, to see a side of science that I was not able to otherwise. Her support was invaluable during my time, and I would like to thank her for her guidance, ideas and help within the research project and beyond it. Without her supervision, I would not be able to complete the project or develop into the scientist that I am becoming, and I am forever grateful for her mentorship. I am also deeply grateful to Dr. Nathalie Juge for providing me with the opportunity and space to work with *Ruminococcus gnavus* and for her support during the project, along with Dr. Emmanuelle Crost who introduced me to working with *Ruminococcus gnavus* so I may have a foundation to build my project on.

I want to thank all the members of the Adriaenssens groups, current and former, for their unconditional support, guidance, and laughter. My time during my research project allowed me to form a great relationship with them as colleagues and as friends. Thank you to Dr. Teagan Brown for getting me started to work with bacteriophages and her supervision. James Docherty for the constant bioinformatics support, funny lunchtimes and moments shared together. Claire Elek for her support and guidance during the duration of my project. Dr. Oliver Charity, may he rest in peace, for his encouragement for me to question things and explore all possible options. Dr. Ryan Cook for the unforgettable moments of laughter and support as well as encouraging me to pursue other opportunities. Dr. Hannah Pye for her supervision, funny moments, and assistance in learning how to produce the figures in the latter portion of the thesis. Luke Acton for his support and guidance. Xena Dyball for her continued support and contributions to my project, her input was tremendously helpful, and I would not have been able to complete the work on *Ruminococcus gnavus* without her.

A special thanks to all my family and friends that supported me during both the good and bad. Especially to my parents, who always gave their all to support me as best they could and instilled the value of working hard within me. My sister, Elizabeth, and my brother, Duke, who kept encouraging me during hard times and reminded me to keep on working. Finally, I would like to express my infinite gratitude to my former supervisor Dr. Anastasia Sobolewski, for providing me with the best foundation I could have as a researcher and encouraging me to see what's available for me and try, regardless of the outcome. I would not be pursuing this Master's were it not for her.

1. Introduction

1.1 Gut Microbiome – Overview

The human gut microbiota refers to the population of microorganisms within the human gastrointestinal tract (GIT) comprising bacteria, viruses, archaea and eukarya. Of these microorganisms inhabiting a "healthy" (absent of any overt disease, as defined by Aagaard and Petrosino, 2013) GIT, bacteria make up 98% of them and the majority belong to one of the four phyla Firmicutes, Bacteroidetes, Actinobacteria, and Proteobacteria (Aagard, et. al., 2012; Das & Nair, 2019). The microbiota has evolved to develop a mutually beneficial relationship with their human hosts beginning as far back as we can remember; symbiotic in nature, the relationship between host and gut microbiota can be characterised by cross-talking interactions spanning the metabolic, immune, and neuroendocrine systems (Kho and Lal, 2018). These systems are reliant on microbial metabolites produced that assist with various host functions, including maintaining host homeostasis (Jandhyala, et al., 2015). Dysbiosis of this relationship has been associated with various diseases such as Crohn's disease, type 1 diabetes and asthma in infants and inflammatory bowel diseases (IBDs), type 2 diabetes and colorectal cancer in adults (DeGruttola, et. al., 2016). Though bacteria dominate the cellular fraction of the microbiome, viruses are an essential aspect of the gut and maintaining balance, driving microbial composition and diversity despite remaining one of the least understood aspects of the human gut microbiome. Most of the viral component (virome) of the gut is made up of bacteriophages (hereafter also referred to as phages), bacterial-infecting viruses that have seen a particular renewed interest within the last two decades due to their potential to treat bacterial infections (Lin, et. al., 2017).

Initially, knowledge of the microbiome's composition and diversity was derived from extensive culture-dependent methods (Kellenger, 2001). Over the last two decades this has been superseded by advances in sequencing technology creating low-cost, high-throughput culture-independent methods that have been instrumental in providing greater insight into the gut microbiota. Advances have also been made in culturing the human gut microbiome in tandem with sequencing using "culturomics", that have been instrumental in demonstrating that many of what was previously thought of as being 'unculturable' is, in fact, able to be cultured (Browne, et. al., 2016). The United States focused Human Microbiome Project and the European MetaHIT project are two key examples of dedicated studies involving metagenomics that have shed more insight into the human gut microbiome (Ehrlich, et. al., 2011; Turnbaugh, et. al., 2007). Though these studies revealed greater diversity than previously detected with culturing, the inability to translate these findings to culturing of said bacteria still represents a major barrier to fully understanding the microbiome's composition, diversity and, thus, function. A single healthy individual's gut microbiome can harbour between 150 to 400 species, differing greatly in taxonomy between individuals (including twins) and even across an individual's lifespan due to factors such as age, mode of delivery, diet, exercise, antibiotics use and the cross-interactions of all of these (Lloyd-Price, et. al., 2016; The Human Microbiome Project Consortium, 2012). Bearing this is mind, everyone can be considered to have their own unique microbiome and microbial identity/signature (Gilbert, 2015). Attempts have been made to establish a "core" microbiome, described as the stable components of the microbiome between individuals and over time, but remains a complex challenge.

1.2 Gut Microbiome Development

The origin of microbial colonization is thought to begin at birth as it has been long believed that the infant GIT was relatively sterile (though this has been heavily challenged) or has negligible levels of microbes at birth (Jimenez, et. al., 2008; Malmuthuge & Griebel, 2018). The relatively low biomass of the fetal microbiome has made it particularly challenging to address relevant questions and further

examine the origins of the microbiome (Kennedy, et. al., 2023). The delivery of the neonate exposes it to the mother's microbial population to initiate a vertical transmission of the microbiome from mother to child that is responsible for establishing a portion of the gut microbiome (Dominguez-Bello, et. al., 2010). The type of birth, caesarean or vaginal, can also influence the initial microbial population with infants born vaginally presenting more significant microbial diversity including a greater concentration of Bacteroides, Bifidobacteria and Lactobacillus, the latter contributing to inhibiting the growth of pathogenic bacteria (Coelho, et. al., 2021; Huurre, et. al., 2008). In contrast, caesarean-section born infant microbiomes tend to be dominated by *Staphylococcus*, *Streptococcus* and *Clostridium* and a lack of microbial diversity that has been associated with an increased chance of developing immune-related disorders, though the associations are not always consistent across studies (Bernstein, et. al., 2016; Black, et. al., 2016; Kristensen & Henriksen, 2016). Following birth, colonisation is not fully settled, and the development of the microbiome can take several years until it is adult-like. After birth, being breast-fed or formula-fed is the key determinant in colonization of the infant gut (Jost, et. al., 2013; Solís, et. al., 2010). Formula-fed individuals tend to present more diversity that is like that of an adult microbiome with breast-fed individuals. Stewart, et. al., reported observation of three distinct phases of microbiome progression from birth to early childhood: i) a developmental phase (months 3–14), ii) a transitional phase (months 15–30), and iii) a stable phase (months 31–46) (Stewart, et. al., 2018). The advancement of phases was marked by changes to the top 5 genera (Bifidobacteria, Enterococcus, Vellonella, Streptococcus and Enterobacter) and Shannon diversity index over time, namely a large significant change to all 5 phyla and the index, a change to two phyla (Bacteroidetes and Pseudomonadota, renamed from Proteobacteria in 2021; Oren and Garrity, 2021), and no change at all, respectively (Oren and Garrity, 2021). The decrease in gut bacteriome diversity, and subsequent stabilisation, can be attributed to certain bacteria only transiently colonising the infant GIT but being easily lost as "true" colonizers stabilise diversity and abundance (Ferretti, et. al., 2018). While the infant bacteriome is developing, it is generally believed that the infant gut contains no viruses at birth and is instead colonized by induced prophages of gut bacteria vertically transmitted from the mother during birth and from breastfeeding (Lim, et. al., 2015). Lim, et. al., observed the first month of life being dominated by class *Caudoviricetes* (previously named of families *Myoviridae, Podoviridae*, and *Siphoviridaeclass* abolished as of 2023 by the ICTV and replaced with class *Caudoviricetes* to represent all tailed bacterial and archaeal viruses with icosahedral capsids and a double-stranded DNA genome; Turner, et. al, 2023) and being the period of greatest bacteriophage richness (Lim, et. al., 2015) (Turner, et. al., 2023). In the following 24 months of life observed, bacteriophage richness and diversity decrease and a marked shift towards an abundance of *Microviridae* can be observed along with an inverse correlation to *Caudoviricetes* abundance. Liang, et. al., saw a similar pattern with *Caudoviricetes* being most abundant in the first month of life and observing the appearance of families such as *Microviridae* and *Inoviridae* during later months (Liang et al., 2020). Given the transient bacterial colonization and high bacteriophage diversity at birth, it can be believed that the phage population is starts out as unsustainable, and the resulting effect is the observed decrease in phage diversity and richness. Subsequently, the "true" bacterial colonizers are able establish themselves and then drive bacteriophage composition to form the gut virome.

1.3 Diet and Microbiome

Beyond the initial establishment of the gut microbiome, more prominent factors affect the gut microbiota – both extrinsic (e.g., diet, lifestyle, medication) and intrinsic (e.g., genetics, immune function, metabolism) shape the gut microbiota. Of these, diet has been studied the most and can be considered as being the most significant (Rothschild, et. al., 2018). Perhaps the most apparent example is the difference in microbial composition in individuals that tend to consume a modern

Western diet, and those that consume others, e.g., a rural diet, and the effects this can have on disease and disease protection. Filippo, et. al., compared the diets of children in Europe (EU) to those of children in a rural village in Burkina Faso (BF) (Filippo, et. al., 2010). BF children were found to show a unique abundance of bacteria from the genera Prevotella and Xylanibacter, both of which are known to contain genes for hydrolysis of xylan and cellulose but were completely absent in EU children. Conversely, EU children had higher levels of Shigella and Escherichia that were vastly underrepresented in BF children and are genera that are closely genetically related with several disease-causing species between them. The difference in diets goes beyond microbial composition however, as these bacteria also produce (or don't) specific compounds, chemicals and nutrients that contribute to bodily functions. BF children's diets that were rich in starch, fiber, plant polysaccharides, and predominantly vegetarian that allowed them to cultivate anaerobic bacteria that, in the process of fermenting these foods, produce short-chain fatty acids (SCFAs). SCFAs such as butyrate and lactate have been shown to mediate inflammation within the gut (Venegas, et. al., 2019; Vinolo, et. al., 2011). Despite the small sample size, this coincides with the reduced levels of anaerobic bacteria present in many individuals with inflammatory diseases, particularly the ratio between Bacteroides and Firmicutes (the former was significantly more enriched than the latter in BF children) as the two most important bacterial phyla in the GIT (Stojanov, et. al., 2020). Despite controlling many diseases and illnesses that still plague rural areas across the world, the developing world suffers from an increasing prevalence in inflammation and inflammatory bowel disease (IBD), and a high intake of fats, sugar and meat has been positively associated with their development while intakes of dietary fibre and fruits have shown a negative association (Chapman-Kiddell, et. al., 2010; Hou, et. al., 2011). The combination of diet-specific factors leading to a shift in microbial composition, and the resulting compounds available (or not) for use within the body represent a marked influence that the microbiome has in the development of disease, in this case IBD. Though evidence begins to elucidate the role diet plays in the development of IBD, the role of specific bacteria remains unclear.

1.4 *Ruminococcus gnavus* – role as commensal and in IBD

Ruminococcus gnavus (*R. gnavus*) is a Gram-positive, strictly anaerobic, mucin-degrading, gut bacteria belonging to the phyla Firmicutes and is prevalent in the human GIT with a relatively high abundance (Ludwig, et. al., 2009). *R. gnavus* is part of the family *Lachnospiraceae* and though it retains the genus *Ruminococcus* for research purposes (and will be referred to as such throughout this thesis), it has recently been reclassified as belonging to the genus *Meditarraneibacter* (Togo, et. al., 2018). *R. gnavus* was (and still is, predominantly) considered to be a commensal bacterium, part of the "normal" healthy gut flora of humans; studies have shown evidence that it is prevalent from early life, both from vertical transmission from mother and possibly from diet (Sagheddu, et. al., 2016; Tannock, et. al., 2013). It's been well-documented to produce beneficial metabolites, such as butyrate and lactate, and has been reported to have a positive function within food digestion and exhibiting anti-tumour effects against colorectal cancer (Crost, et. al., 2018; Zhang, et. al., 2023). While playing a vital role in many positive functions, recent years have begun to showcase the negative associations of *R. gnavus* on human health. Namely, it has been reported to cause bacteraemia and bloodstream infections and, most notably, has been associated with the development of IBD and Crohn's disease (Hall, et. al., 2017; Henke, et. al., 2017; Kim, et. al., 2017). It should be noted that some of these reports have been ambiguous and involved a previously underlying condition that the patient had, in the case of the infections, which could have contributed to the development of infection.

1.5 Enterobacter cloacae – resistance and prevalence as an ESKAPE pathogen

Of the phylum *Pseudomonadota*, *Enterobacteriaceae* are a large family of over 50 genera and 210 species of Gram-negative, non-spore forming bacteria present across multiple diverse ecological

niches (Rock, et. al., 2014). Members of the *Enterobacteriaceae* family are facultative anaerobes that are mostly commensal within the human gut; as one of the first colonisers of the gut they work to deplete the environment of oxygen and create the necessary conditions for stricter anaerobes to colonise the GIT (Yang, et. al., 2021). Alongside this, they are also frequently associated with intestinal and extraintestinal diseases; within the *Enterobacteriaceae* family and the context of public health, *Escherichia*, *Enterobacter*, *Salmonella* spp, *Klebsiella* and *Shigella* spp are especially relevant for being opportunistic pathogens, rarely causing disease in healthy individuals but capable of causing disease should the optimal conditions arise, e.g., a weakened immune system. As a genus, *Enterobacter* species are of note due to their extensive resistance to antibiotics, earning them their role as part of the ESKAPE group of pathogens - *Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa*, and Enterobacter species - extensively multidrug resistant isolates responsible for most nosocomial infections (Santajit and Indrawattana, 2016). The *Enterobacter cloacae* (E. cloacae) complex is especially relevant, as a group composed of 6 closely related species with similar resistance profiles to antibiotics; *E. cloacae, E. asburiae, E. hormaechei, E. kobei,* and *E. ludwigii*. Distinguishing between the complex at the species level can be difficult, hence their clustering together. As a group, they are responsible for 65-75% of *Enterobacter* infections and can cause pneumonia, sepsis, urinary tract infections, respiratory infections, meningitis, and more (Annavajhala, et. al., 2019). *E. cloacae* is the species of focus (from the *E. cloacae complex*) within this thesis. Alongside causing many infections within hospitals, it is also extensively resistant to antibiotics and drugs. Specifically, it is resistant to penicillins and first- and second-generation cephalosporin antibiotics due to the low expression of ampC genes; these encode an inducible AmpC-type cephalosporinase that inactivates beta-lactam within antibiotics. Amongst third-generation cephalosporins and aztreonam, resistance can still arise due to mutations, usually in ampD, leading to hyperproduction of AmpC. As part of the Enterobacter species it is resistant to carbapenems, fourth-generation cephalosporins, fluoroquinolones, β-lactams, polymyxins, and is both multi- and pan-drug resistant (De Oliveira, et. al., 2020). *E. cloacae* and *Enterobacter* species have become key targets for the use of phages as an alternative to antibiotics.

1.6 Antibiotics, rise of resistance

Alexander Fleming's discovery of antibiotics in 1928 ushered in a new era of combatting bacterial infections, though he warned overuse could lead to "self-medication and all its abuses"; a warning that came to fruition in the coming decades (Fleming, 1929; 1945). As the original antibiotic penicillin was used as the core structure from which further antibiotic analogues were developed from, however this would perpetuate the rise of penicillin-resistant bacteria, as well as generalised antibiotic resistance (ABR), and by the late 1960s it was estimated that over 80 percent of community and nosocomial *Staphylococcus aureus* (*S. aureus*) strains were penicillin-resistant (Lowy, 2003). Further penicillin analogues aimed to counter penicillin resistance by expanding antimicrobial activity beyond Gram-positive bacteria but were met with evolving resistance. Methicillin (Celbenin), first introduced to the UK in 1959 to counter penicillin-resistant *S. aureus*, encountered resistance only two years later as the first methicillin-resistant *Staphylococcus aureus* (MRSA) strain was reported following screens of over 5000 *S. aureus* strains in the UK (Knox, 1960; Jevons, 1961). 20 years later methicillin resistance would become endemic in the US, making up 29 percent *of S. aureus*-infected hospitalizations, and has spread globally since (Panlilio, et. al., 1992; Kaur and Chate, 2015). Ampicillin, a broad-spectrum penicillin-class antibiotic, was developed to be effective against both Gram-positive and Gram-negative bacteria, treating infections caused by *S. aureus, Escherichia coli* (*E. coli*), *Haemophilus influenzae* (*H. influenzae*) and *Streptococcus pneumoniae* (*S. pnuemoniae*) (Kaushik, et. al., 2014). It was also used as a combination therapy with co-trimoxazole to treat enteric fever by *Salmonella enterica* serovar Typhi that was already resistant to the initially used chloramphenicol (Herzog, 1976).

Widespread use of ampicillin not only conferred resistance in the previously mentioned pathogens but also resulted in developing multi-drug resistant (MDR) *S. typhi* strains, causing several strains to exhibit resistance to ampicillin, chloramphenicol, co-trimoxazole and more drugs (Saga and Yamaguchi, 2009).

1.7 Bacteriophages: Biology, Diversity, Genomics

The initial discovery of bacteriophages can be attributed to Frederik Twort in 1915 and Félix d'Hérelle in 1917; Twort was the first to observe a "transparent material" that inhibited the growth of bacteria while d'Hérelle later isolated an "anti-Shigella microbe" and coined the term bacteriophage or "bacteria-eater", realising their therapeutic potential (Twort, 1915; d'Hérelle, 1917). This prompted further studies into early attempts of phage therapy, as d'Hérelle would use phages to successfully treat Salmonella gallinarum-infected chickens and Shigella dysenteriae-infected patients from 1919 to 1921 (d'Hérelle, 1926; Ho, 2001). Subsequent studies continued investigating the effectiveness of phage therapy and, though encouraging, also exposed many challenges to overcome. Namely; (1) most bacteriophages infect a narrow host-range due to binding to host-specific cell receptors, making them highly selective for certain genera, species or strains; (2) purification limitations at the time resulted in bulk titres of phages being contaminated with bacteria and creating further infections; (3) pharmacokinetics reported inefficient efficacy due to rapid removal of phages from the body; and (4) the constant evolution between bacteria and phage for mechanisms of defence and infection respectively created bacteria that could readily develop resistance to phages (Smith, 1924; Krestownikowa and Gubin, 1925; Hadley, 1928; Cowie and Hicks, 1932; Luria and Delbrück, 1943). These barriers to overcome coupled with the discovery and widespread use of antibiotics in 1928 onwards shifted focus away from bacteriophages and phage therapy within the Western world until the 1980s. At the same time the USSR, Poland, and, mostly notably, Georgia continued phage research with the latter's legacy present today as the George Eliava Institute of Bacteriophages, Microbiology and Virology (Chanishvili and Sharp, 2009). Following World War II, the institute prepared various phage treatments in many forms against bacterial pathogens across dermatology, gynaecology, paediatrics, surgery and more. Despite extensive preclinical and clinical trials, the language barriers, non-compliant standards and political tensions caused distrust when considering international approval of bacteriophages for treatment (Kutateladze and Adamia, 2010).

Bacteriophages are the most abundant biological entity on Earth, numbering upwards of 10^{31} particles globally and outnumber bacteria approximately 10 to 1 (Comeau, et. al., 2008). This estimation is frequently reported across phage literature, often without sources, but Mushegian, 2020 has reaffirmed that the predicted number of phages remains close to 10^{31} by comparing recent and varied assessments of the estimate (Mushegian, 2020). Phages are ubiquitous within nature, occupying virtually every ecological niche and shaping microbial communities ranging from animal microenvironments to soil and oceans across the world (Clokie, et. al., 2011). Phages, their genomes, and structures are varied and diverse (Figure 1) – their genomic content is composed of either double-stranded (ds) or single-stranded (ss) DNA or RNA which is packaged into a capsid (Ackermann, 2009; Dion, et. al., 2020). The capsid shape and structure ranges from polyhedral, filamentous, pleomorphic, or attached to a tail: at present, most isolated and characterised phages contain dsDNA and are tailed phages (Ackermann, 2007). Phages can vary greatly in genome size alone. The smallest recorded phage to date is the *Leuconostoc* phage L5 with 2435 bp. In contrast, there has been an increasing number of phages with >200 kb being reported and referred to as "jumbo phages", displaying unique genomic features such as atypical genome organization and genes involved in replication and nucleotide metabolism not present in smaller phage genomes (Yuan and Gao, 2017). These are second in size to a group of phages referred to as "megaphages", with genomes >540 kb and predicted to

infect *Prevotella* species (Devoto, et. al., 2019). Though found in the gut microbiomes of humans, baboons and pigs, these phages were found to have fragmented genomes and use a repurposed stop codon as an alternative genetic code so require further study. With such diversity, phage classification has been a point of contention over the last few decades and their taxonomy is carried out by the International Committee on Taxonomy of Viruses (ICTV) (Adams, et. al., 2017). Historically characteristics such as morphology and host range originally served as the defining point of reference for phage classification, however recent advances in large-scale viral metagenomics, genome sequencing, and improved culturing efforts have drastically overhauled our knowledge and resulted in a shift in focus to now prioritise factors such as genomics and proteomics instead (Adriaenssens and Brister, 2017; Turner, et.al., 2021).

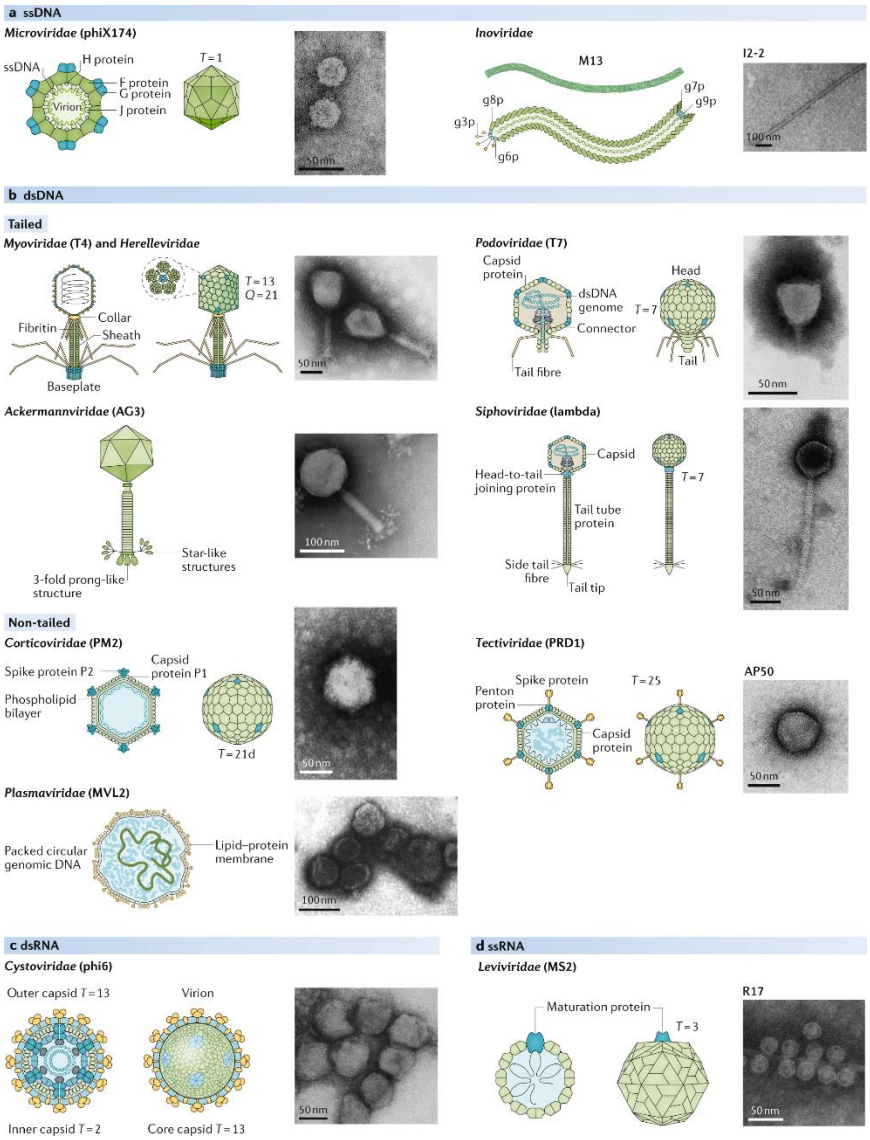


Figure 1 – Classification of phages based on genome type and morphology, schematic representations and transmission electron microscopy are shown for each phage type. dsDNA phages currently represent the most abundant characterized phages. Figure taken from: Dion, M.B., et. al. (2020). Phage diversity, genomics and phylogeny, *Nature Reviews Microbiology*, 18, pp,125-138.

Prior to the abolishment of morphology-based taxonomy in 2022 many tailed phages were reported to belong to the now defunct *Caudovirales* order, previously made up of five families: *Myoviridae, Podoviridae, Ackermannviridae, Herelleviridae* and *Siphoviridae*. Despite being classified based on

morphology, phages in these families tended to differ greatly in capsid size, tail type, and tail length. It was instead decided that these families and the *Caudovirales* order would be abolished in favour genome-based classification, and the groups were instead replaced by the class *Caudoviricetes* to group all tailed bacterial and archaeal viruses with icosahedral capsids and dsDNA genomes. This family, and others forthcoming, should better reflect shared evolutionary histories and lineage of the phages that could better inform future classification. Bacteria and archaea all contain the 16S rRNA gene with a subset of hypervariable regions; the gene allows for phylogenetic analysis, taxonomic classification, and being used an alternative sequencing method (Johnson, et. al., 2019). The presence of such a gene with a varying degree of conservation allows for easier identification and characterization of diverse bacterial communities. In comparison, to date there has been no single gene shared amongst most isolated bacteriophages. The lack of a consistent gene creates additional hurdles in further understanding bacteriophage genomics alongside developing and utilising tools for resolving bacteriophage genomes. Turner, et. al., outlined a guide for assembly and high-quality annotation of bacteriophage genomes to lay a foundation for phage genomics for researchers to draw on, defining questions ranging from assembly conditions to genome annotation methods and taxonomy (Figure 2; Turner, et. al., 2021). In doing so, many challenges associated with phage genomics have come to light and demonstrate the gaps in knowledge/technology that remain to be resolved for more efficient phage genome resolution. For instance, the type of genome assembly undertaken is a key factor to be considered in phage genome assembly and annotation. Illumina or Oxford Nanopore Technology are the two main types of sequencing currently employed, for short and long read data respectively, and Illumina is more commonly used of the two. In comparison, relatively few phages have been reported to be sequenced with Oxford Nanopore Technology alone and are typically used in hybrid assemblies instead – those that incorporate both short-read and long-read data. Shovill (Seeman, 2017) is typically used for assembling phage genomes from short-read data, however all assemblies using short-read data will inevitably be prone to low coverage and generating homopolymeric contigs that will require either filtering or further inspection. Elek, et. al., reported a hybrid, polypolish workflow for an as complete and accurate assembly of bacteriophages as possible (Elek, et. al., 2023). This method may prove to be the 'gold standard' for assembling bacteriophage genomes; using both short and long read data can cover discrepancies that relying on one type alone generate. Alongside this, multiple polishing steps were used to ensure the genomes were as refined as possible. This method generated ten high-quality and accurate phage genomes that could then be used for a better understanding of phage biology, taxonomic classification, and comparative genomic analyses.

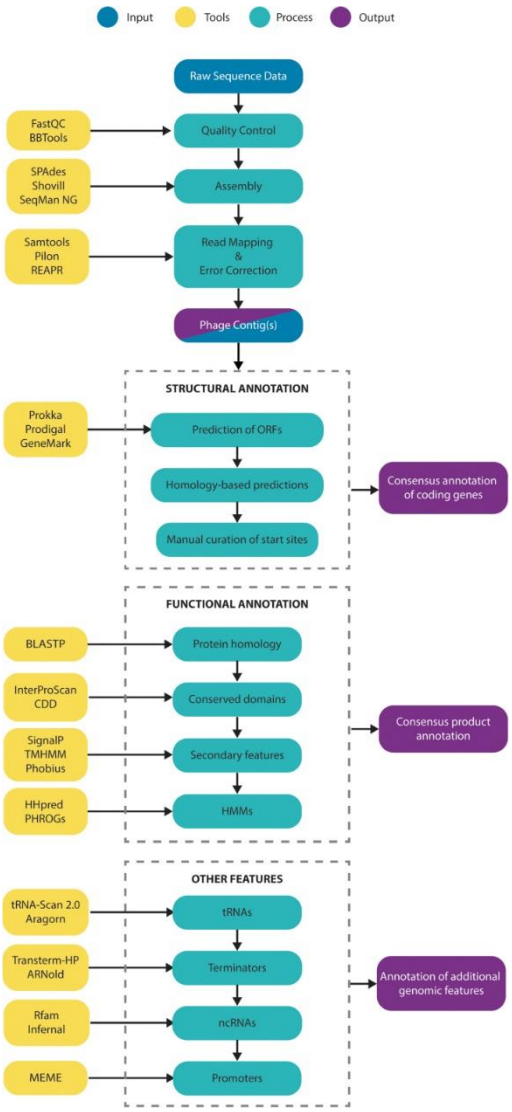


Figure 2 - A proposed and recommended workflow for the annotation of phages, including structure, functions, and other features when completing assembly of bacteriophage genomes. Yellow - examples of tools recommended tools for different processes. CDD, Conserved Domains Database; HMMs, Hidden Markov Models; ncRNAs, noncoding RNAs; ORFs, open reading frames; PHROGs, Prokaryotic virus Remote HOmologous Groups; tRNAs, transfer RNAs. Source: Turner, D., Adriaenssens, E.M., Tolstoy, I., and Kropinski, A.M. (2021). Phage Annotation Guide: Guidelines for Assembly and High-Quality Annotation, *Phage (New Rochelle),* 2(4), pp.170-182.

Following successful genome assembly, annotation of the phage genome is the next hurdle to overcome in generating a high-quality phage genome. Annotation can be divided into two categories: (1) structural annotation that identifies coding sequences (CDS) through gene calling, and (2) functional annotation that is the identification of gene products and predicted CDS. Most programs that perform phage genome annotation can do both types of annotation, though the exact methodology may differ. Most of these programs, however, were initially designed to be used to annotate bacterial genomes and are not optimized for phages. One caveat to bear in mind when using these tools is that bacterial genomes tend to have larger open reading frames (ORFs) compared to viruses, so annotation with these tools can lead to many CDS being missed, particularly smaller ones, or are just unable to be identified. For example, the prokaryotic genome annotation tool Prokka (Seeman, 2014) is itself made up of other dependencies and tools that all come together. Structural annotation is performed using Prodigal (Hyatt, et. al., 2010) to identify protein-coding sequences, itself relying on factors such as start codon usage that vary greatly between bacteria and phages.

Following structural annotation, functional annotation is performed by searching databases to assign functions to the predicted CDS. These databases are as follows: (1) ISFinder, composed of insertion sequences isolated from bacteria and archaea (Siguier, et. al., 2006); (2) the NCBI Bacterial Antimicrobial Resistance Reference Gene Database (NCBI Resource Coordinators, 2016), composed of antimicrobial resistance genes; and (3) UniProtKB (SwissProt), with >95% of its protein sequences provided by the International Nucleotide Sequence Database Collaboration (INSDC) (The UniProt Consortium, 2021). Most of the data contained within these databases pertain to bacterial proteins and genes, typically resulting in majority of phage genomes being annotated as "hypothetical proteins" due to the inability of these databases to resolve them. More recently, the development of bacteriophage-focused databases and tools has begun to emerge and greatly assist with phage genome annotation. Terzian, et. al., developed the Prokaryotic Virus Remote Homologous Groups (PHROG) database, that is a library of >930,000 proteins built on >17,000 reference viruses of prokaryotes (Terzian, et. al., 2021). This database is currently the largest and only one tailored to prokaryotic viruses and serves to better annotate prokaryotic viral sequences. It has been implemented in the fast phage annotation tool Pharokka (Bouras, et. al., 2023), a successor of Prokka that is specialised for the annotation of bacteriophage genomes and metagenomes. For structural annotation, Pharokka uses PHANOTATE (McNair, et. al., 2018) which is currently the only gene prediction tool tailored to bacteriophages. Functional annotations are then assigned by matching predicted CDS to the PHROG database alongside the Comprehensive Antibiotic Resistance Database (CARD; Alcock, et. al., 2020) and the Virulence Factor Database (VFDB; Chen, et. al., 2005). Pharokka represents a bacteriophage-specific tool that is optimal for curating annotated bacteriophage genomes and should be favoured over bacteria-based tools that may lead to inaccurate or missing CDS and annotations. The development of tools in the same vein are not just ideal but essential for elucidating bacteriophage genomics and better curating annotated bacteriophage genomes for others to draw on and utilise.

Bacteriophages broadly operate in one of two life cycles, depending on whether they exclusively use the lytic cycle (virulent phages) or use the lysogenic cycle and then the lytic cycle (temperate phages) (Figure 3). Lytic lifecycles begin with phages irreversibly binding to a surface receptor, utilizing specific receptor binding proteins to attach to the cell surface before penetrating the cell wall and injecting their genetic material into the bacteria. Inside, the viral genome begins synthesising endonucleases and exonucleases for host genome degradation to then utilize host cell machinery to begin genome replication, particle synthesis and assembly. Newly replicated phage genomes, via polymerases (DNA or RNA), synthesize phage particles such as the capsid, its subunits, and tail fibres, which are then assembled into mature phage. The head and tail structures begin independent assembly, with the capsid subunits and proteins forming empty heads to be packed with condensed phage genetic material, fusing with the tails to form matured phage progeny. Phage-produced enzymes weaken the host cell wall as the phage progeny accumulate within the cell, eventually lysing the bacteria and releasing progeny to seek out other susceptible bacteria in the surrounding environment. In contrast, lysogenic phages do not necessarily immediately kill their hosts, though it is sometimes possible - following injection of their genetic material, phage nucleic acid integrates itself into the host genome and is now dubbed a 'prophage' (Saussereau and Debarbieux, 2012). The prophage is replicated in tandem with the bacterial genome, lying dormant within the daughter cell and future generations. Exposure to stressors (UV light, low nutrient conditions, chemicals) and subsequent damage to DNA can cause the prophage to make a 'switch', entering 'phage induction' in which prophages excise from the host genome, then entering the lytic cycle (Figure 3).

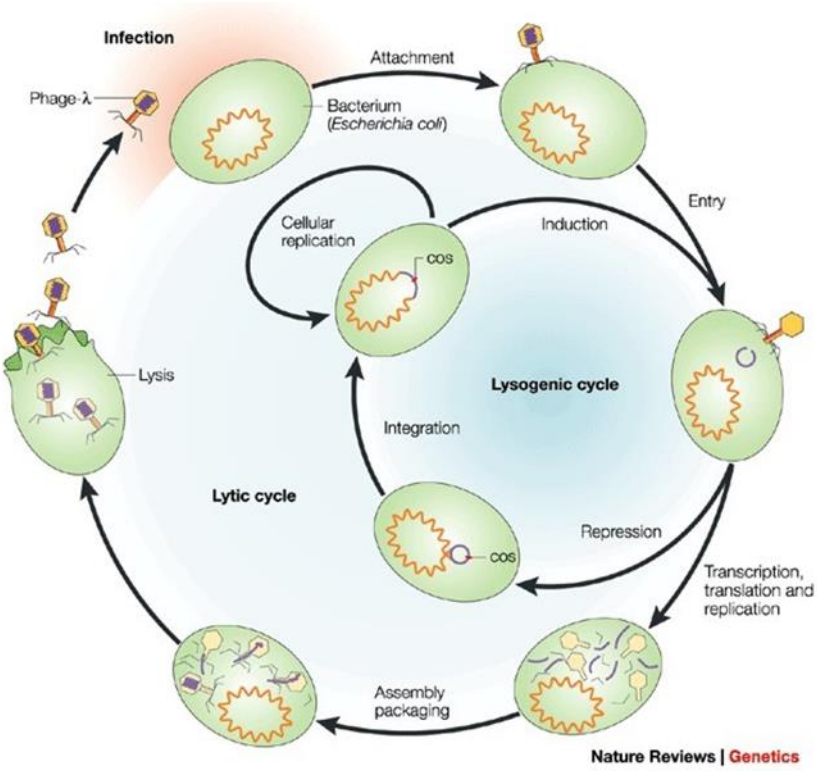


Figure 3 – Lifecycle of temperate phage coliphage-λ, illustrating both lytic and lysogenic cycle. **cos -** cohesive site, point of closure of the ends of the linear phage DNA that then circularises before integrating itself into the bacterial genome. Figure taken from Campbell, A. (2003). The future of bacteriophage biology, *Nature Reviews Genetics,* 4(6), pp.471-477.

Phages may also exhibit alternative lifecycles that are not as well characterized (Figure 4). (1) Pseudolysogeny, defined by Łoś and Węgrzyn (2012), is the state in which the genetic material of the phage is maintained within the host cell without integrating into the host DNA or inducing replication, effectively arresting further bacteriophage development and existing as an episome (Figure 4A) (Łoś and Węgrzyn, 2012). (2) Carrier state, where the phage is attached to the host surface without infection, typically in a non-permissible host (Figure 4B) (Abedon, 2009). (3) Chronic infection, most observed in filamentous phages, is the release of phage progeny to the surrounding environment without host lysis, resulting in continuous phage assembly, reproduction, and release from the bacterium (Figure 4C) (Clokie, et. al., 2011). The prevalence of these cycles remains poorly understood.

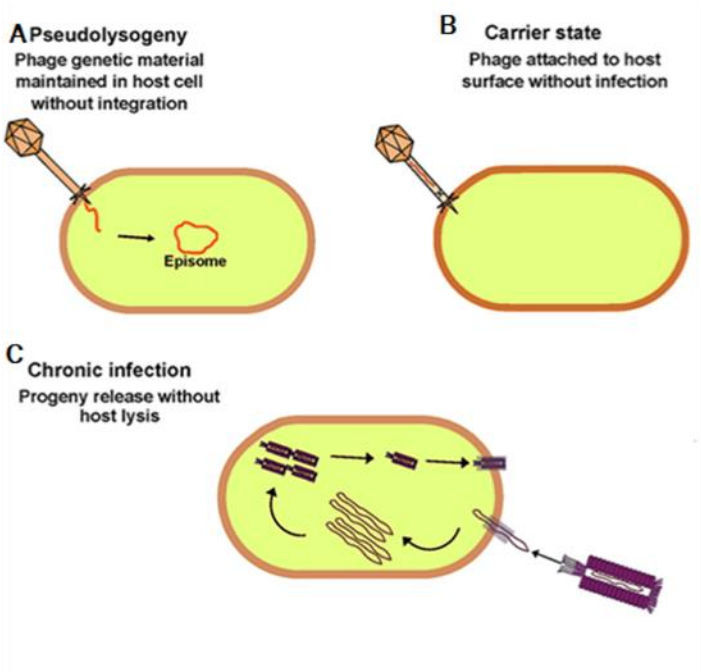


Figure 4 – Alternative lifecycles of bacteriophages, as opposed to the lytic and lysogenic cycle. **A –** Pseudolysogeny; **B –** carrier state; **C –** chronic infection. Figure was cropped and taken from Guerin, E. and Hill, C. (2020). Shining Light on Human Gut Bacteriophages, *Frontiers in Cellular and Infection Microbiology*, 10.

1.8 Resurgence of Phage Therapy, Current State, Challenges

Bacteriophage therapy has garnered more attention over the last two decades as an alternative treatment to combat bacterial infections and antimicrobial resistance (AMR). It has been well documented that bacteria are continuing to grow resistant to antibiotics and other drugs being develop, at a rate that we are not able to keep up with or get ahead of and, as of 2019, have been classified as one of the ten global healthcare issues humanity are currently facing (World Health Organization, 2019). It was estimated that in 2019, AMR contributed to 4.95 million deaths and could be directly attributed to 1.27 million, highlighting the severity of the issue (Antimicrobial Resistance Collaborators) As previously described, phage therapy encountered many issues following the discovery of phages, particularly due to their narrow host range, purification challenges at the time, poor pharmacokinetic reports, and the evolution of bacterial resistance to phages – the coupling of this with the discovery of antibiotics shelved what was thought to be promising potential (Smith, 1924; Krestownikowa and Gubin, 1925; Hadley, 1928; Cowie and Hicks, 1932; Luria and Delbrück, 1943; Fleming, 1928). The era of antibiotics began but by the 70s and 80s, it was clear that antibiotics were not sustainable, especially as they were being misused at the time. Smith and Huggins improved optimism of phage therapy during the 80s through a series of studies addressing many earlier hurdles. Namely, they proved that *in vitro* results could translate to *in vivo*; proved the superior efficacy of phages compared to antibiotics; combatted phage-resistant mutants with a multi-phage approach and used calcium carbonate as a precursor to phage treatment to prevent stomach acid from killing the phage (as was thought at the time) (Smith and Huggins, 1982, 1983, 1987). The combination of addressing the previous criticisms combined with developing high-throughput screening, genetic sequencing and increasing antibiotic resistance made phage therapy see a dramatic resurgence. This was particularly in the West, as Eastern Europe had already adopted phage therapy since World War II and had extensive evidence of using phages to treat antibiotic-resistant infections caused by *Staphylococcus, Pseudomonas, Klebsiella*, and *E. coli* amongst other infectious bacteria (Weber-

Dabrowska, et. al., 2000). In the West, phage therapy began being trialled in various animal models, seeing success in reduced mortality, and treating antibiotic resistance in mice, hamsters and other animals, even using phage cocktails and seeing success (Biswas, et. al., 2002; Soothill, 1992; Wang, et. al., 2006; Watanabe, et. al., 2007). The positive results of these studies combined with the need to develop alternative therapies to antibiotics showed a positive upside for phage therapy. Considering these positive results, there are still more hurdles to climb before phage therapy can become fully realised. Host specificity towards species and strain can be both a pro and a con; host specificity ensures that only targeted bacteria are being killed, and in environments like the gut microbiota that contain an abundance of other bacterial species, this prevents potentially beneficial bacteria from being killed. Conversely, infections that can be colonized by multiple strains of bacteria, e.g., burn wounds, diabetic foot ulcer infections, will not be as effectively treated by phage therapy and a phage cocktail of multiple phages effective against a range of bacteria would be an alternative, but this is also limited by difficulties of large-scale production and distribution (Servick, 2016). Adding to the issue, oral administration phage is generally considered to be the optimal route for infections within the body (McCallin, et. al., 2013; Merabishvili, et. al., 2009; Sarker, et. al., 2016), however there exists the possibility of the phage translocation across the intestinal epithelium, where phage may end up within circulating blood (Górski, et. al., 2006.) The exact effect of this on the body is unknown as some data has demonstrated that translocation of phage could be of benefit to the host via downregulation of immune response (Górski, et. al., 2006), while other studies have shown an innate host immune response that aimed to remove phage from the body in mice (Hodyra-Stefaniak, 2015 Park, et. al., 2014). It has also been suggested that phage translocation could affect immunocompromised patients, though overall there is no consensus on what the effects are within an individual. Currently, there are no approved phage therapy products for human use within the US or the EU; they are classified as drugs within the US and medicinal products within the EU but because no explicit framework or guidelines exist to oversee and regulate the use of phages, they are only reserved as a last resort for special cases where no other treatment would be deemed effective (Naureen, et. al., 2020; Verbeken, et. al., 2012). Phage therapy has made considerable strides since its inception and has proven to be viable as an alternative treatment for bacterial infections and AMR. However, much remains to be discovered about phages and the regulatory framework for the sustainable use of phages needs to be implemented before therapy becomes commercially available.

2. Aims

The aim of the project is to discover novel phages targeting *E. cloacae* and *R. gnavus* and investigate the use of these phages to treat bacterial infection and modulate the gut microbiota. For phages targeting *E. cloacae,* treating associated infections by killing the strain responsible and as an alternative to antibiotics would be the ultimate aim should phage therapy with these phages happen. With phages targeting *R. gnavus,* in patients with associated inflammation and/or IBD, modulation of the *R. gnavus* population by administering phages is the goal. The gut should not be made sterile of *R. gnavus* as it is still an important member of the gut microbiota. To meet this aim, five distinct scientific objectives were devised:

1. **Learn phage-hunting procedures and techniques and isolate novel phages for *Enterobacter cloacae*.**
2. **Sequence novel phage genomes and perform functional genome assembly and annotation.**
3. **Characterise novel phages via host-range analysis.**
4. **Develop plaque assay procedure to discover phage/s for *Ruminococcus gnavus.***
5. **If (4) is successful, repeat (2) and (3) novel for *Ruminococcus gnavus* phage/s.**

3. Materials & Methods

3.1 *Enterobacter cloacae* Culturing and Conditions

Enterobacter cloacae (*E. cloacae*) strain NC10005 was used as a target host. All experimental methods were either performed in a Class II ducted microbiological safety cabinet (MSC) or at a work bench with a Bunsen burner alight. Cultures were grown on Lysogeny Broth (LB) (Oxoid LTD, Basingstoke, UK) plates with 1.5% agar by streaking from frozen stock on to plates and incubating overnight at 37 °C. When used for enrichment and assays, colonies were sub-cultured in 15 ml of LB medium and incubated at 37 °C and 200 RPM for 3 hours in an incubator shaker.

3.2 *Ruminococcus gnavus* Culturing and Conditions

Ruminococcus gnavus strains (*R. gnavus*) ATCC 29149, E1, ATCC 35913, CC55_001C and Finegold were used as target hosts. Cultures were grown from prepared pre-reduced 10 ml glycerol stock, prepared by Emmanuelle Crost, made of 2% *R. gnavus* subculture in brain heart infusion broth supplemented with yeast extract and hemin [BHI-YH; BHI (Oxoid LTD, Basingstoke, UK) supplemented with 5 g/L of BactoTM yeast extract (Becton, Dickinson and Company, Sparks, MD) and 5 mg/L of hemin (Sigma-Aldrich)] and 80% glycerol, defrosted from storage at -80 °C. Cultures were propagated using anaerobe basal broth (ABB; Oxoid LTD, Basingstoke, UK) and grown in 37°C in strict anaerobic conditions (Concept Dual Chamber1000 anaerobic chamber, Baker Ruskinn), experimental methods were either performed within the anaerobic cabinet or in a Class II ducted MSC. 400 µl of glycerol stock was transferred to 10 ml anaerobic ABB broth and incubated overnight at 37 °C to grow a starter culture. Following incubation, 200 µl of starter culture was inoculated in fresh 10 ml anaerobic ABB broth to create a 2% subculture following further 6 – 8-hour incubation at 37 °C.

3.3 *R. gnavus* Media Comparison

200 µl starter culture of *R. gnavus* 29149 was inoculated in 10 ml of ABB and 10 ml of BHI-YH and grown for 8 hours to compare which media would be optimal for creating a sufficient bacterial lawn for phage enrichment and isolation. Optical density (OD) at 600 nm was recorded following inoculation and subsequently hourly after 3 hours of incubation (t=3) using 3 ml of subculture. At t=4 and onwards, 500 µl of subculture was transferred to a 15 ml centrifuge tube before 4 ml of respective melted and cooled (between 45 - 50 ˚C) 0.5% agar was added. Contents were transferred to respective 1.5% agar plates and swished to cover the plate uniformly. Plates were transferred to the anaerobic cabinet and incubated upside down at 37 ˚C overnight to form bacterial lawns. This was repeated for both media hourly. At t=4, 20 µl of subculture was added to 180 µl phosphate-buffered saline (PBS) pH 7.4 (ThermoFisher Scientific, UK) and a 10-fold dilution series (from 10^{-1} to 10^{-8}) was created. Two 5 µl spots of each dilution was spotted on to 1.5% agar plates of each media. Plates were then transferred to the anaerobic cabinet and incubated upside down at 37 ˚C overnight. This was repeated hourly. Following overnight incubation, the number of colony forming units was recorded.

3.4 Sample Collection

Water and wastewater samples were collected in 50 ml centrifuge tubes from various regions and used for isolation of phages. Potential *E. cloacae* phages were assessed using samples from River Wensum (Norwich, Norfolk) collected by me; Chester Wastewater Treatment Works (Chester, England), Five Fords WwTW (Wrexham, Wales) and Llangefni Sewage Works (Llangefni, Wales) samples were collected by collaborators Kata Farkas and Davey Jones at Bangor University. Samples were taken back to the Quadram Institute Biosciences (QIB, Norwich) and frozen at –20 °C. Potential

R. gnavus phages were assessed using pre-processed faecal samples taken from a study cohort of elderly patients with varying risks of mental decline collected by Oliver Charity stored at –20 °C. Wastewater and sewage samples collected by collaborators at Bangor University stored at 4 °C were also used. After defrosting, a maximum of 25 ml of sample was transferred to a fresh 50 ml centrifuge tube and centrifuged at 3000 RPM for 10 minutes. Multiple samples originating from the same location but taken at different timepoints were pooled together for use. Supernatant was filtered through a 0.45 µm cellulose acetate syringe filter unit (Sartorius AG, Germany) into a fresh glass universal bottle and stored at 4 °C to be used for assessing potential phage presence.

3.5 Sample Enrichment – *E. cloacae*

Filtered samples were enriched to allow the bacteriophage to proliferate within the host culture. The enrichment was performed by adding 200 µl of liquid bacterial culture (exponential phase) to a 1.7 ml tube with 1 ml of filtered sample, allowing 15 minutes for an adsorption period to initiate the infection process. The total contents of the 1.7 ml tube were transferred to 15 ml of LB broth and incubated in an incubator shaker at 37 °C and 200 RPM overnight. Enriched samples (enrichments) were then transferred to a 50 ml centrifuge tube and centrifuged at 3000 RPM for 10 minutes to pellet bacterial cells and debris. The supernatant was filtered through 0.45 µm cellulose acetate filter unit (Sartorius, Germany) into a glass universal bottle and stored at 4 °C.

3.6 Sample Enrichment – *R. gnavus*

400 µl of processed fecal samples were added to 100 µl of 2% *R. gnavus* subculture of each strain in a 1.7 ml tube and allowed to adsorb for 15 minutes. This was repeated for processed wastewater and sewage samples instead using 1 ml of sample and 500 µl of *R. gnavus* subculture. Contents were transferred to 10 ml of anaerobic ABB and incubated at 37 °C and 200 RPM overnight. Enrichments were transferred to a 15 ml centrifuge tube and centrifuged at 3000 RPM for 10 minutes. Supernatant was filtered through 0.45 µm filter unit into a glass universal bottle and stored at 4 °C to be used for testing bacteriophage presence.

3.7 Bacteriophage Isolation: Double Agar Method – *E. cloacae*

200 µl of liquid culture was transferred to a 15 ml centrifuge tube before 4 ml of melted and cooled (to 45 - 50 °C) 0.5% LB agar was added. Contents of tube were flicked several times before being poured on to a 1.5% LB agar plate and swished around gently to cover the plate uniformly. Plates were left to dry, allowing the overlay to solidify. A 10 µl "spot" of filtered enrichment/s was pipetted on top and allowed to adsorb/dry for at least 15 minutes. Plates were then incubated upside down overnight at 37 °C to form bacterial lawns and checked the next day for any sign of a clearance zone to indicate bacterial lysis. An alternative screening method performed for samples was the addition of 100 µl filtered enrichment to the centrifuge tube, along with bacteria, and allowing an adsorption period of a few minutes before the addition of melted agar, which allowed for the screening for individual plaques (i.e., a lysis zone originating from a single phage particle). Contents were then transferred directly to the plate, left to solidify, and incubated overnight in the same conditions.

3.8 Bacteriophage Isolation: Double Agar Method - *R. gnavus*

1.5% ABB agar plates were pre-reduced anaerobically for 24 hours in the anaerobic cabinet prior to use. 500 µl of *R. gnavus* 2% subculture was transferred, within the anaerobic cabinet, to a 15 ml centrifuge tube. 4 ml of melted and cooled (to 45 - 50 °C) 0.5% ABB agar was added to the tube and contents were flicked several times before being poured on to a 1.5% ABB agar plate and swished around gently to cover the plate uniformly. Plate was left to dry, allowing overlay to solidify. A 10 µl "spot" of filtered enrichment/s was pipetted on top and allowed to adsorb/dry for at least 15 minutes. Plates were then transferred back to the anaerobic cabinet, incubated upside down overnight at 37

°C for lawn formation and checked the next day for any sign of a clearance zone to indicate bacterial lysis.

3.9 Bacteriophage Single Plaque Purification

For storing bacteriophage samples, bacteriophage buffer was prepared using (per litre); 10 mM TRIS at pH 7.5, 10 mM $MgSO_4$, 68 mM NaCl and 0.1 mM of $CaCl_2.2H_2O$. Buffer was then autoclaved and stored at room temperature. Upon observing clearing and plaque formation, a portion of the overlay (a "plug") was removed (either using a pipette tip, toothpick, or plastic inoculating needle) and transferred to 500 µl of bacteriophage buffer in a 1.7 ml tube. The tube was then briefly vortexed and centrifuged at 9000 RPM for 5 minutes. Supernatant was removed and transferred to a new tube and a 10-fold serial dilution was created (from 10^{-1} to 10^{-8}) in bacteriophage buffer with 10 µl of each dilution being spotted onto an agar overlay plate with bacterial lawn, prepared as previously stated (3.5). The dilution series was repeated at least 3 times until consistent single plaque sizes and morphology was clearly observed.

3.10 Bacteriophage Stock

Upon completion of the dilution series, a plug was excised from the overlay at the lowest dilution (or one with most visible plaques), resuspended, vortexed, and centrifuged as previously described (3.5). Supernatant was transferred to a 1.7 ml tube where 100 – 150 µl was spread on to multiple bacterial lawn overlay plates. Plates were then incubated until clearing was observed (between 3 hours and overnight). Bacteriophage buffer was aliquoted as 3 ml of buffer per plate to a 50 ml centrifuge tube. Once clearing was observed, plates were scraped off and added to bacteriophage buffer tube before being vortexed and centrifuged at 4000 RPM for 5 minutes. Supernatant was filtered through a 0.45 µm filter unit into a glass universal bottle. Bacteriophage stock can then be used for further downstream analysis. The concentration of plaque forming units (PFUs) was carried out to determine phage titre. To calculate the PFU, the number of plaques observed at the lowest serial dilution was counted and recorded as per 10 µl then multiplied by 100 to calculate the number of plaques per 1000 µl (1 ml). The scale of the lowest dilution factor was reversed and multiplied by the number of plaques per 1 ml to get the final number of PFUs.

3.11 Host-Range Analysis & Phage Characterisation – *E. cloacae*

Host range analysis of *E. cloacae* phages was determined by spotting 10 µl of phage lysate onto bacterial lawn of *Enterobacter ludwigii* strains HVP30 and HVP31, a clinical isolate of *E. cloacae* from a patient at Great Ormond Street Hospital (code GOSH2), *Escherichia coli* es2833 and *E. cloacae* strains PO27E, PO95LP and PO69D using the double agar overlay method (3.5). Clearing or plaque formation were assessed following overnight incubation at 37°C. Upon identifying successful zones of clearing, host-range was confirmed using follow-up spotting of a 10-fold dilution series, as followed using the single plaque purification method (3.7). Plaque formation was assessed and PFU per ml was calculated. A bacterial strain was confirmed as host for a phage isolate when individual plaques were observed. Efficiency of plaquing (EOP) was calculated using PFU per ml of each phage against the respective strain and calculating a ratio relative to the original host, *E. cloaocae* NC10005.

3.12 Bacteriophage DNA Extraction

DNA was extracted from purified phage stock subjected to DNase I (1 µg/µl) and RNase A (1 µg/µl) (QIAGEN, Germany) using AllPrep PowerViral DNA/RNA Kit (QIAGEN, Germany) according to the manufacturer's protocol. Where sufficient DNA was not extracted, DNA extraction was repeated using the Maxwell RSC Instrument (Promega, United Kingdom) instead, according to the manufacturer's protocol. Extracted DNA was stored in buffer in 2 ml collection tubes.

3.13 DNA Quantification and Sequencing

For DNA quantification, Qubit HS dsDNA kit, Qubit assay tubes and Qubit 3.0 fluorometer were used (ThermoFisherScientific, UK). Working solution was composed of 1 x *n* µl Qubit reagent diluted in 199 x *n* µl Qubit buffer where *n* equals the combined number of standards, samples, and an additional assay tube (to account for pipette errors), in a centrifuge tube. High and low standard tubes were prepared by adding 190 µl of working solution to 10 µl pre-diluted DNA standards from the kit. Sample tubes were prepared by adding 10 µl of extracted DNA sample to 190 µl working solution. All tubes were then vortexed for 3 seconds before incubation at room temperature for 2 minutes. Standards were read in Qubit 3.0 fluorometer, and a standard curve was generated, referring to the Qubit user manual when necessary. Samples were read in the fluorometer according to the user manual and concentrations were noted. DNA was then sent for sequencing using either Illumina or Nanopore (QIB Core Sequencing Services, Norwich) and uploaded to the QIB internal platform for sequence data management Integrated Rapid Infectious Disease Analysis (IRIDA, Canada) as forward and reverse reads. Reads were imported to the Galaxy web platform (Afgan, et. al., 2018) for genome assembly and analysis.

3.14 Genomic Assembly and Analysis

Sequence data was pre-processed using fastp (Chen, et. al., 2018) to prepare them for assembly. Reads were then assembled using Shovill (Seeman, 2017) – a faster SPAdes (Bankevich, et. al., 2012) assembly using paired end reads – generating a file of contigs. Contigs were visualised using Bandage (Wick, et. al., 2022) to determine which contigs had assembled into the phage genome. When the contig representing the phage genome was determined, it was run through the nucleotide basic local alignment search tool (BLAST) (Altschul, et. al., 1990), to search for regions of similarity to other uploaded nucleotide sequences, identifying any significant sequence homology to other phages. Criteria of >95% sequence identity and the whole genome length were used to determine if any sequences with significant alignments belonging to phages in the BLAST database could be within the same species as my isolated phage/s. Genbank (.gbk) files of phages meeting said thresholds were downloaded for genome comparison. An alternative method genome assembly was adopted, using the method outlined by Elek, et. al., 2023 – the HYPPA workflow (Elek, et. al., 2023) – as a comparison to the initial assembly and for validation of assembled phages. Prokaryotic genome annotation tool Prokka (Seeman, 2014) was used to annotate the assembled contigs of isolated phages, creating genbank files (amidst other outputs) containing coding regions and their predicted product. Where this had failed, Pharokka (Bouras, et. al., 2023) was instead used for annotation. Genbank files of isolated phages and phage relatives were ran through the programme Clinker on the Comparative Gene Cluster Analysis Toolbox (CAGECAT) webserver to search for and visualise homologous gene clusters (Gilchrist and Chooi, 2021). To analyse intergenomic similarity to phage relatives, Virus Intergenomic Distance Calculator (VIRIDIC; Moraru, et. Al., 2020) was used to produce pairwise intergenomic similarities amongst viruses. For taxonomic analysis and classification, VIPTree was used to generate a "proteomic tree" of viral genome sequences based on genome-wide sequence similarities computed by tBLASTx and compared to the Virus-Host database (Mihara, et. al., 2016). When evaluating the potential of a prophage in MO3, Phage Search Tool Enhanced Release (PHASTER; Arndt, et. al., 2016) was used to identify and annotates prophage sequences, assigning a score based on parameters that indicate the likelihood of a region belonging to a prophage.

4. Results

4.1 Isolation + Characterisation of novel *E. cloacae* phages

To test for phages against *Enterobacter cloacae* strain NC10005 (hereafter also referred to as *E. cloacae*), various wastewater samples were used. Samples of wastewater from River Wensum, Norfolk; River Thames, London; Lake Nicholson, London; and University of East Anglia Broad, Norfolk were collected in 50 ml centrifuge tubes by me. Samples collected by collaborators at Bangor University, Wales were also tested and originated from Kinmel Bay, Llandudno Junction, Holyhead, Flint, Treborth, Chester, Wrexham, and Llangefni taken at different timepoints. Samples of the same origin but different timepoints were pooled together to maximise efficiency of screening for phages. Phage enrichments for *E. cloacae* strain NC10005 were made as described in Materials & Methods (section 3.5) for 12 samples; 8 from wastewater treatment plants and 4 river samples (Table 1). For 4 wastewater treatment plants, multiple samples were collected at different timepoints by collaborators and aliquots of each were combined and enriched with *E. cloacae*. Aliquots of the filtered phage enrichments were spotted on to lawns of the enrichment strain and tested for clearance (Figure 5). Clearance was observed in 6 out of 12 enrichments (Table 1) and these zones of clearing were used for a dilution series to observe the formation of single plaques, indicating the presence of a productively infecting bacteriophage.

Table 1 – Samples tested for phages against *E. cloacae* NC10005, including origin and timepoints used. Observation of clearing and subsequent formation of single plaques is noted. *N* = no, *Y* = yes.

Sample Origin	**Sample Location**	**Timepoints Used**	**Clearing Observed?**	**Single Plaques Formed?**
Norfolk, England	River Wensum	07/02/2022	N	N
Kinmel Bay, Wales	Kinmel Bay Waste water plant	11/11/2021	Y	N
Llandudno Junction, Wales	Ganol Waste water plant	11/11/2021	Y	N
Flint, Wales	Flint Waste water plant	11/11/2021	Y	N
Chester, England	Chester Wastewater Treatment Works	11/11/2021; 18/11/2021; 22/11/2021	Y	Y
Wrexham, Wales	Five Fords WwTW	12/11/2021; 18/11/2021; 22/11/2021	Y	Y
Llangefni, Wales	Llangefni Sewage Works	11/11/2021; 12/11/2021; 18/11/2021; 19/11/2021; 22/11/2021	Y	Y
Bangor, Wales	Treborth Waste water plant	11/11/2021	N	N
Holyhead, Wales	Holyhead Waste water plant	11/11/2021; 12/11/2021; 18/11/2021	N	N
London, England	Lake Nicholson	23/04/2022	N	N
London, England	River Thames	23/04/2022	N	N
Norfolk, England	University of East Anglia Broad	26/04/2022	N	N

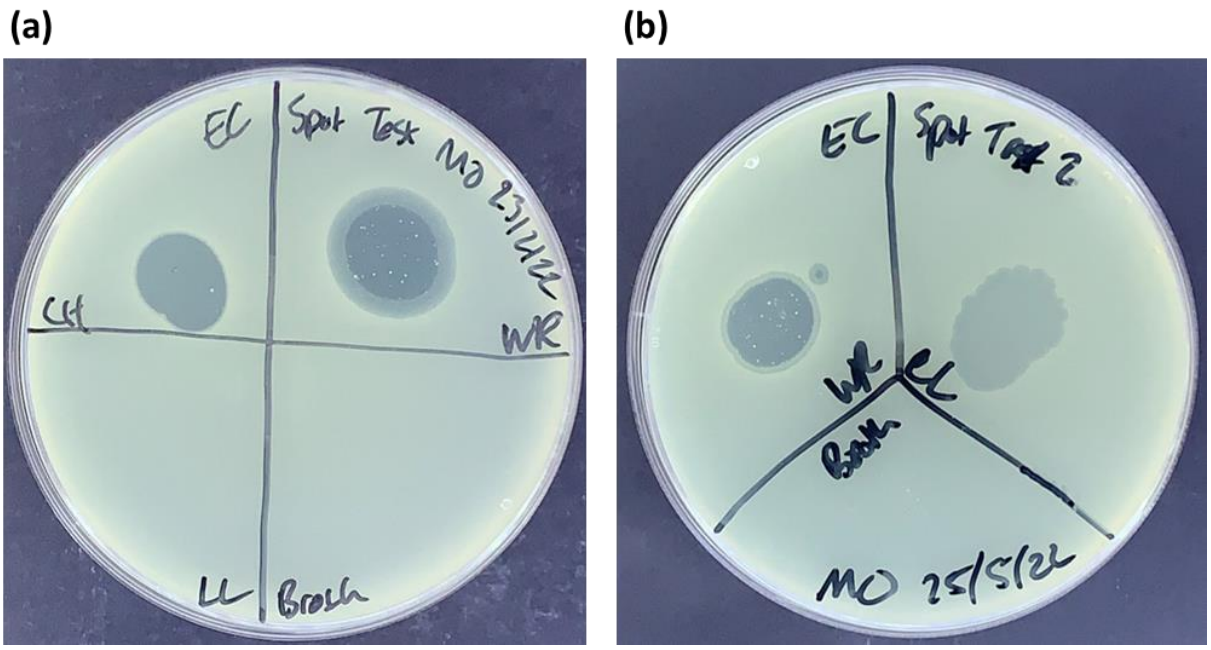


Figure 5 – Spot test of wastewater samples from Chester (CH), Wrexham (WR) and Llangefni (LL) (and LB broth as negative control) on lawn of *E. cloacae* NC10005 (EC). Samples were pooled together from the same treatment plant at different timepoints, hence there is a distinction between the LL samples tested in **(a)** and **(b)**. The same WR sample in **(a)** was used as a positive control in **(b)**. **(a)** Zones of clearing can be observed for CH and WR, WR also displays a "halo" – a turbid, outer ring surrounding its zone of clearing. **(b)** Clearing can be observed for LL, with a puffy morphology and is slightly opaquer compared to the positive control of WR.

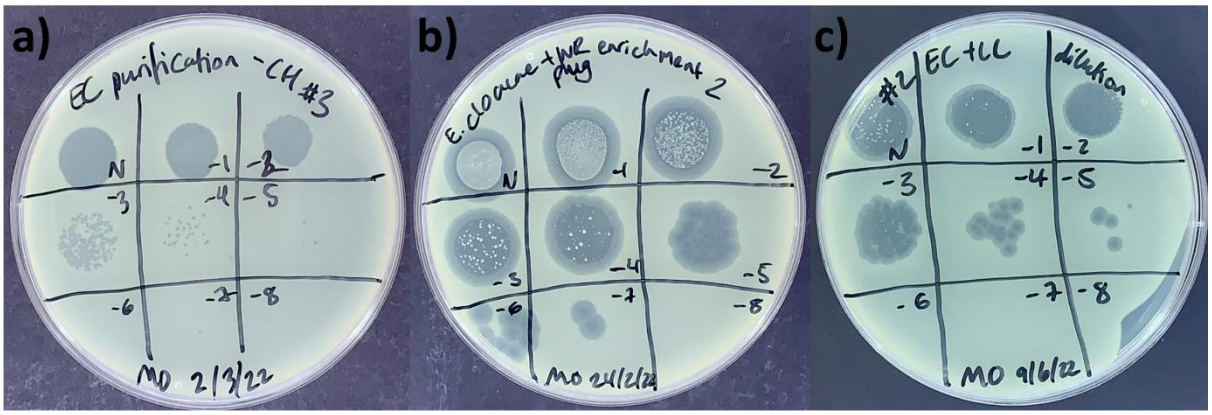


Figure 6 – Single plaque formations of wastewater samples from Chester (**a**), Wrexham (**b**), and Llangefni (**c**).

Single plaques were observed for samples originating from Chester, Wrexham and Llangefni. Several zones of clearing did not produce single plaques. Following a minimum of three rounds of single plaque purification, three enrichments produced plaques with distinct plaque morphologies (Figure). These phages were named MO1 (6a), MO2 (6b) and MO3 (6c) respectively upon confirming phage presence. Clearing/plaque morphology differs between the type of phage and can be used to make a distinction between different phage isolates. However, it is important to note that even different phages can produce plaques with identical morphology, whereas the same phage can have multiple plaque morphologies. MO1 (Figure 6a) exhibits a circular, transparent zone of little-to-no bacterial growth; MO2 (Figure 6b) demonstrates a turbid, outer ring in addition to this and MO3 (Figure 6c) is slightly more opaque opacity with a "puffy" morphology. After the formation of single plaques was confirmed, plaques were amplified into stocks to be stored and for downstream analysis. Phage titre was calculated diluting an aliquot of phage stock and calculating the number of plaque forming units per ml (PFU/ml).

Phage host-range was tested by spotting a dilution series from an aliquot phage stock on overlays of bacteria available in the Adriaenssens lab, including an antibiotic-resistant clinical isolate of *E. cloacae*, other *Enterobacter* species, additional *E. cloacae* strains and a clinical *Escherichia coli* (*E. Coli)* isolate. These bacteria were available at the time of performing the host range analysis and were chosen; clinical isolates were especially relevant for attempting to observe effectiveness of phages on bacteria isolated from patients as a precursor to eventual phage therapy; additional *E. cloacae* strains and *Enterobacter* species were chosen to evaluate the host range within other strains and the genus, and *E. coli* was chosen to attempt to expand the host range to another member of the gut flora that is also within the same Enterobacteriaceae family. Specifically, a clinical isolate of a resistant *E. cloacae* strain from a patient at Great Ormond Street Hospital (code GOSH2), and *Enterobacter ludwigii (E. ludwigii)* strains HVP30 and HVP31. The exact resistance profile of the *E. cloacae* GOSH2 was kept classified by the hospital. Additional host-range was tested by Dr. Hannah Pye using the same method against a clinical *Escherichia coli (E. coli)* isolate es2833 and *E. cloacae* strains PO27E, PO95LP and PO69D isolated from infant faecal samples as well as original strains tested by me. Genome sequences were only available for *E. cloacae* host strain NC10005.

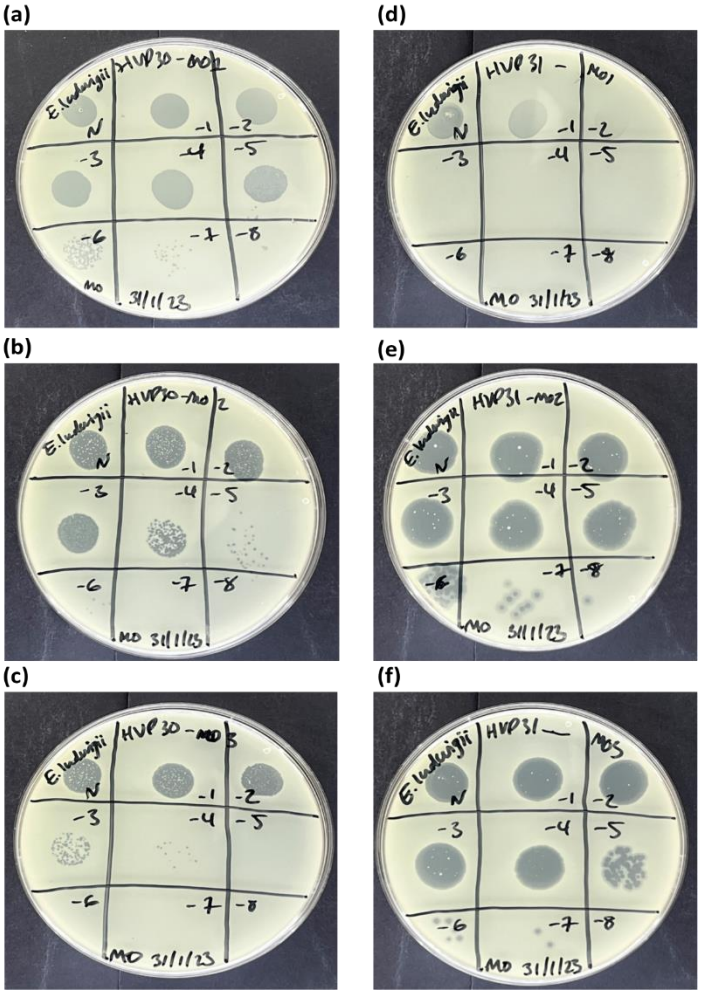


Figure 7 – Host-range test of phages MO1, MO2 and MO3 against *E. ludwigii* HVP30 and HVP31. Plaques are observed at varying dilutions for all phages against both strains (**a, b, c , d, e**) except MO1 against HVP31 (**f**), which only produces clearing.

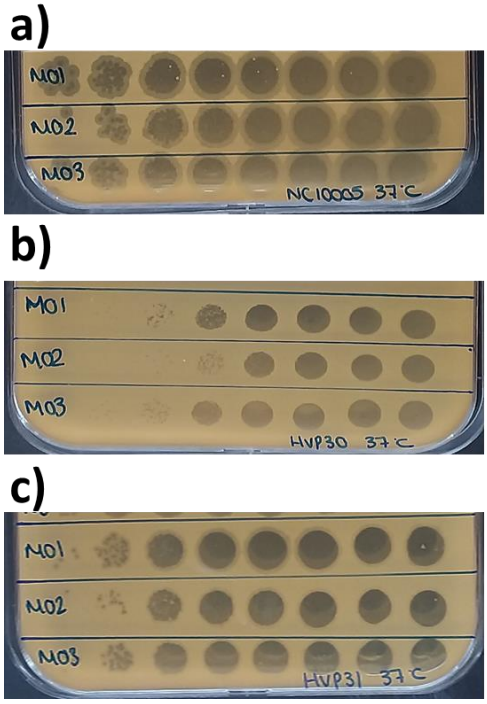


Figure 8 – Host-range tests of MO1, MO2 and MO3 against *E. cloacae* NC10005, repeated by Dr. Hannah Pye. Serial dilution series of each phage were prepared, as previously described, by diluting an aliquot of phage stock in phage buffer.

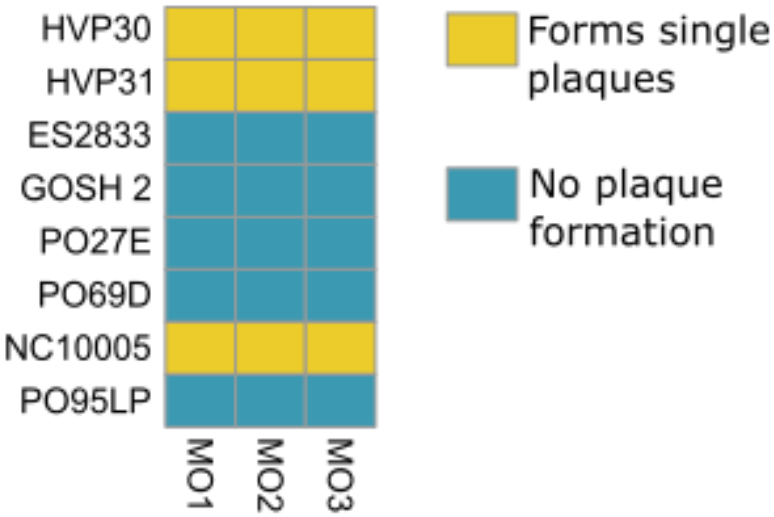


Figure 9 – Heat map of host-range analysis of phages MO1, MO2 and MO3. *Enterobacter ludwigii* strains HVP30 and HVP31 and *E. cloacae* NC10005 were tested by me. *E. cloacae* strains PO27E, PO69D, PO95LP, clinical isolate GOSH2 and *Escherichia coli* strain ES2833 were tested by Dr. Hannah Pye alongside original tested strains.

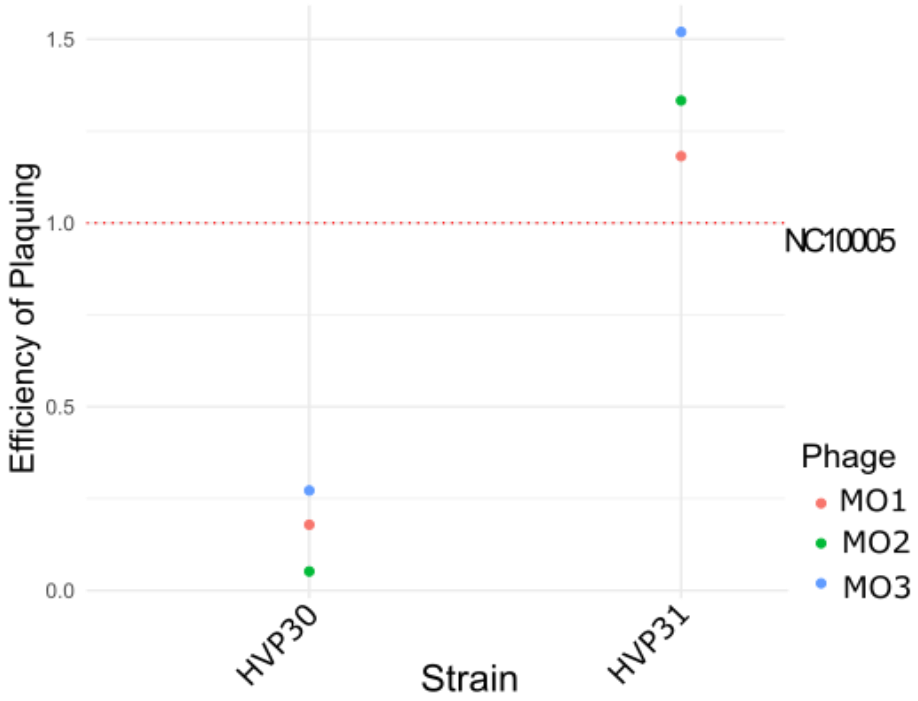


Figure 10 – Efficiency of plaquing (EOP) for phages MO1, MO2 and MO3 against *E. ludwigii* strains HVP30 and HVP31 in comparison to *E. cloacae* strain NC10005 (red dotted line).

Host-range was tested as previously described using the double agar method with each bacterium against phages MO1, MO2 and MO3. In total, 8 different bacterial hosts were tested and 3 were able to be successfully infected and produce single plaques; *E. cloacae* NC10005 and *E. ludwigii* strains HVP30 and HVP31 (Figures 7, 8, 9). These strains were tested twice, once by me and another time by Dr. Pye. Initially, single plaque formation was only observed for MO2 and MO3 on both *E. ludwigii* strains (Figure 7b, c, e, f). MO1 produced single plaques on *E. ludwigii* HVP30 but only a zone of clearance against *E. ludwigii* HVP31 (Figure 7a, d). Upon testing by Dr. Pye, all 3 phages were now able to produce single plaques against both *E. ludwigii* strains (Figure 8). Efficiency of plaquing (EOP) relative to the initial host *E. cloacae* NC10005 was calculated (Figure 10). All 3 phages had lower EOP values for *E. ludwigii* HVP30, between approximately 0.1 and just over 0.25. In contrast, HVP31 had higher EOP values for all phages, between 1.2 and 1.6.

DNA extraction of isolated phages was carried out for genome sequencing. In addition to Illumina, Nanopore sequencing was also done to be used for an alternative phage genome assembly method using both types of read data. Extracted DNA was then quantified as previously described and submitted internally for Nanopore sequencing.

Table 2 – Phages isolated against *E. cloacae* NC10005, describing sample origin, plaque morphology, phage titre and DNA concentrations.

Phage	Sample Origin	Plaque Morphology	Titre (PFU/ml)	DNA Concentration (ng/µL)	
				Illumina	Nanopore
MO1	Chester, Wales	Small, clear plaques	3×10^9	5.56	26.6
MO2	Wrexham, Wales	Large clear plaques, thick turbid outer ring	8×10^{10}	0.5	28.8
MO3	Llangefni, Wales	Large clear plaques, thin turbid outer ring	2×10^9	0.4	60.6

4.2 *E. cloacae* phage Genome Assembly – Shovill

Phages were submitted internally for Illumina sequencing (QIB Core Sequencing Services, Norwich). Initial genome assembly and annotation was carried out using Illumina short-read data predominantly on the Galaxy web platform (Afgan, et. al., 2018). MO1, MO2 and MO3 had 1704320, 3790344 and 985956 reads available per sample, respectively. Reads were pre-processed using fastp (Chen, et. al., 2018), including filtering out too low quality and too short reads, trimming adapter sequences, cutting low quality bases per read and trimming reads in the front and tail (amongst other features) (Table 3). After quality control, 96% of reads were retained of MO1 and MO2 and 97% for MO3, with no difference in mean read lengths.

Table 3 – Fastp pre-processing of phages for mean length of read, total reads, total bases, quality score (Q20 and Q30 bases) and guanine-cytosine (GC) content.

	Parameter	**Before filtering**	**After filtering**
MO1	Mean length (bp)	137.0	137.0
	Total reads (millions)	1.704	1.637
	Total bases (millions)	235.227	225.486
	Q20 bases	215.182	209.007
	Q30 bases	205.503	200.050
	GC Content	43.212	43.085
MO2	Mean length (bp)	147.0	147.0
	Total reads (millions)	3.790	3.645
	Total bases (millions)	559.368	537.668
	Q20 bases	489.765	479.323
	Q30 bases	459.378	450.847
	GC Content	47.980	47.802
MO3	Mean length (bp)	145.0	145.0
	Total reads (millions)	0.986	953.064
	Total bases (millions)	143.949	138.964
	Q20 bases	131.945	128.668
	Q30 bases	123.434	120.626
	GC Content	48.235	48.229

All phages had a decrease in parameters following pre-processing filtering. Mean length of reads varied between phages, with MO2 having the largest (147 bp) and MO1 having the shortest (137 bp).

Sequence data was then assembled using Shovill (Seeman, 2017) – a faster SPAdes (Bankevich, et. al., 2012) assembly using paired end reads. The total length of the assemblies of MO1, MO2 and MO3 were 633,770, 50,130 and 170,917 base pairs (bp), assembled into 87, 1 and 7 contigs respectively. Contig assemblies were visualised using Bandage (Wick, et. al., 2015) to generate *de novo* assembly graphs to determine which contigs could contain parts of the phage genome.

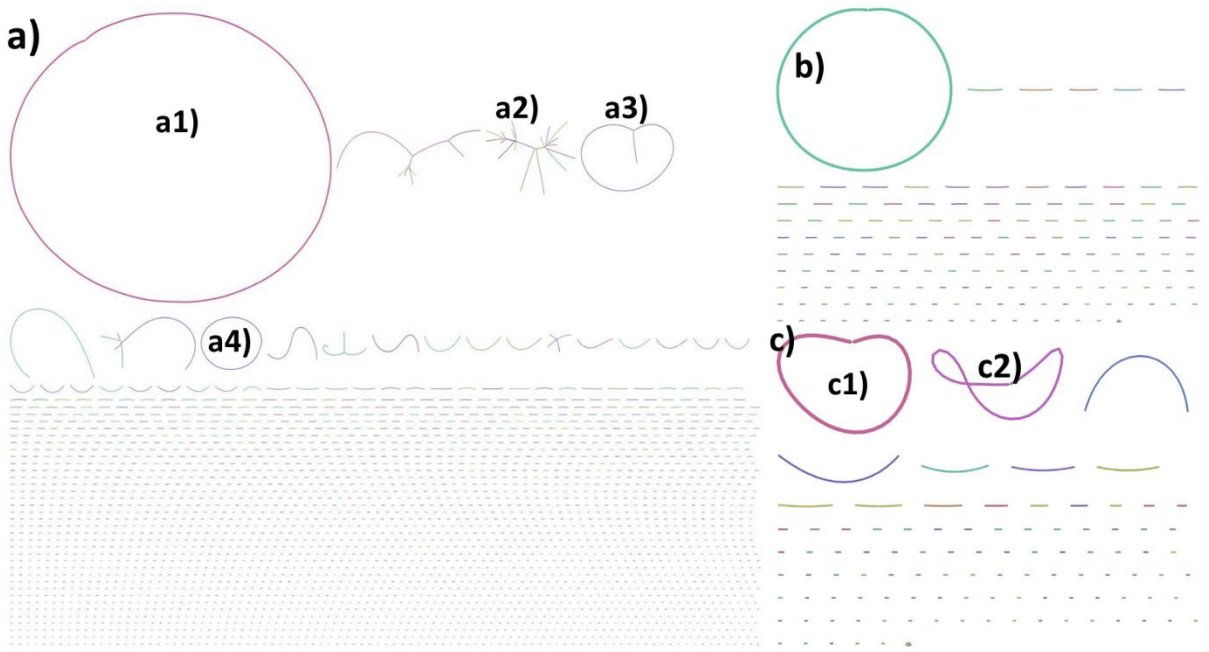


Figure 11 – Visualisation of d*e novo* assembly graph of phage MO1 (**a**), MO2 (**b**) and MO3 (**c**) assembled into 87, 1 and 7 contigs respectively. Contigs refer to the overlapping DNA sequences that can be used to reconstruct the full assembled genome. It should be noted that graphs show full genome assembly, including contigs and DNA fragments that have not been assembled into contigs. **(a)** MO1 phage genome assembly, into 87 contigs. (**a1**) Ideal phage assembly, of MO1 into a single contig with no other branching lines. Shorter lines branched together represent shorter contigs and tend to show poor assembly or contamination, as in **(a2)**. **(a3)** and **(a4)** could also be parts of the phage genome that assembled into different contigs. Other single contigs may also be contamination and require further investigation. (**b**) MO2 phage genome assembly into a single circular contig. (**c**) MO3 phage genome assembly into 7 contigs. (**c1**) and (**c2**) have assembled into separate singular, circular contigs.

Amongst the 87 assembled contigs for MO1 (Figure 11), one had formed a single circular main contig (Figure 11a) alongside other smaller ones (11a3, 11a4). The branching of shorter lines together (11a2) tends to show poor assembly, the presence of repeat regions that cannot be resolved by the assembly algorithm or possible contamination and are treated as such. MO2 (Figure 11b) assembled into a singular, circular contig that is taken as representing the whole phage genome. MO3 (Figure 11c) assembled into 7 contigs of which 2 assembled into single, circular contigs separate from each other (Figure 11c1, c2) suggesting the presence of two separate complete phage genomes, both being parts of a single phage genome or possible contamination. To distinguish between the two, they are hereafter referred to as MO3.1 (Figure 11c1) and MO3.2 (Figure 11c2), the former being used as the "main" phage genome of MO3 for now. This finding will be discussed in more depth in section 4.4 following genome annotation.

Following assembly and contig visualization using Bandage it was decided that where multiple contigs had been assembled, for MO1 and MO3, to use the first contigs as representing most, if not all, of the phage genomes. This was done while bearing in mind that additional contigs could contain aspects of the phage genome but would require further investigation. These contigs of MO1, MO2 and MO3 were 172,146 bp, 50,130 bp and 46,838 bp respectively.

4.3 *E. cloacae* phage Genome Assembly – HYPPA Workflow

To validate the assemblies with Shovill and attempt a more complete and comprehensive genome assembly, Oxford Nanopore Technology long-read assembly data was used in combination with Illumina short-read sequencing in a hybrid and poly-polish phage assembly workflow (referred to as "HYPPA" from here onwards) outlined by Elek and colleagues (Elek, et. al., 2023). The workflow was as

follows: short-read data was pre-processed using fastp (Chen, et. al., 2018). Long-read data was quality checked with NanoStat (De Coster, et. al., 2018; Table 4).

Table 4 – NanoStat quality check of long-read sequencing data for MO1, MO2 and MO3.

NanoStat Parameters	Phage		
	MO1	MO2	MO3
Mean read length	743.6	397.5	533.0
Mean read quality	13.2	13.4	13.4
Median read length	282.0	232.0	273.0
Median read quality	12.9	13.1	13.2
Number of reads	1064189.0	2243921.0	2219110.0
Read length N50	133.0	359.0	604.0
Total bases	791382194	892059546.0	1182844430.0
Quality Cutoffs	**Percentage and megabases of reads above quality cutoffs**		
>Q5	1064189 (100.0%) 791.4 Mb	2243921 (100.0%) 892.1 Mb	2219110 (100.0%) 1182.8 Mb
>Q7	1064189 (100.0%) 791.4 Mb	2243921 (100.0%) 892.1 Mb	2219110 (100.0%) 1182.8 Mb
>Q10	1064189 (100.0%) 791.4 Mb	2243886 (100.0%) 892.0 Mb	2219079 (100.0%) 1182.8 Mb
>Q12	721191 (67.8%) 634.6 Mb	1578586 (70.3%) 687.3 Mb	1579607 (71.2%) 938.4 Mb
>Q15	193612 (18.2%) 242.1 Mb	484311 (21.6%) 253.6 Mb	476031 (21.5%) 356.4 Mb
Top 5 highest mean basecall quality scores and their read lengths			
1:	26.3 (233)	27.0 (177)	27.1 (170)
2:	25.3 (247)	26.9 (387)	26.1 (217)
3:	25.2 (300)	26.9 (387)	26.0 (219)
4:	24.5 (328)	26.7 (208)	26.0 (185)
5:	24.4 (197)	26.3 (212)	25.7 (147)
Top 5 longest reads and their mean basecall quality score			
1:	77101 (15.1)	104064 (13.9)	92405 (12.8)
2:	53359 (13.8)	73795 (12.6)	61240 (13.9)
3:	53092 (12.2)	72315 (13.6)	59624 (14.7)
4:	53084 (14.7)	61795 (11.9)	58732 (10.4)
5:	52858 (15.1)	55730 (11.3)	55936 (14.0)

Genome assemblies were then carried out with Flye (Kolmogorov, et. al., 2019), with Canu (Koren, et. al., 2017) being used as an alternative assembler where Flye was unable to generate a high-quality assembly. Trimming and error correction were performed as part of the default settings of both assemblers. Additionally, Flye performed an iteration of long-read polishing by default. Phages were assembled without trimming adapter/barcode sequences for long reads.

Table 5 – Results of Flye genome assembly for MO1.

MO1			
Sequence Name	**Length (bp)**	**Mean Coverage (Reads)**	**Circular?**
Contig_1	50013	10861	Y

Table 6 – Results of Flye genome assembly for MO2.

MO2			
Sequence Name	**Length (bp)**	**Mean Coverage (Reads)**	**Circular?**
Contig_3	35984	9192	N
Contig_20	5902	447	N
Contig_10	4820	12426	N
Contig_6	3948	14522	N
Contig_5	3257	3221	N
Contig_21	1143	481	N
Contig_9	742	13507	N
Contig_8	741	13774	N
Contig_4	508	14646	N
Total Length (bp)	57045	-	-

Table 7 – Results of Canu genome assembly for MO3.

MO3			
Sequence Name	**Length (bp)**	**Mean Coverage (Reads)**	**Circular?**
Tig00000001	58695	127	N

Flye succeeded in assembling genomes for MO1 and MO2 but was unable to do so for MO3, where Canu was used instead. MO1 and MO3 both assembled into single contigs numbering 50013 and 58696 bp respectively. MO2 assembled into 9 single contigs totalling 57045 bp. MO1 was the only phage that had a predicted circular genome. Several iterations of short-read and long-read polishing were performed on assemblies in a specific order. Firstly, two iterations of long-read polishing were performed using Medaka (Wright and Wykes, 2020) with default settings, the previously polished data was then used as the input for the next round of polishing.

Table 8 – Long-read polishing iterations done with Medaka and results after.

Phage	**Input**		**Medaka #1**		**Medaka #2**	
	# of contigs	**Total length (bp)**	**# of contigs**	**Total length (bp)**	**# of contigs**	**Total length (bp)**
MO1	1	50013	1	50013	1	50013
MO2	9	57045	9	57045	9	57045
MO3	1	58695	1	58695	1	58695

Following two iterations of long-read polishing with Medaka, the number of contigs and the total length remained the same for all three phages (Table 8). Next, one iteration of short-read polishing was performed using Polypolish (Wick and Holt, 2022, Table 9) with default settings using pre-processed short-read data (from fastp) and the second iteration of long-read polishing from Medaka (Table 8).

Table 9 – First short-read polishing iterations done with Polypolish.

Phage	Input		Polypolish	
	# of contigs	Total length (bp)	# of contigs	Total length (bp)
MO1	1	50013	1	50013
MO2	9	57045	9	57045
MO3	1	58695	1	58695

After polishing with Polypolish, all three phages remained unchanged in the number of contigs and total length (Table 9). As the final step of the workflow, a second iteration of short-read polishing was performed using POLCA (Zimin and Salzberg, 2020) with default settings and the first iteration of short-read polishing data from Polypolish (Table 10).

Table 10 – Second short-read polishing iterations done with POLCA.

Phage	Input		POLCA (Final Polishing)		
	# of contigs	Total length (bp)	# of contigs	Total length (bp)	Consensus Quality
MO1	1	50013	x	x	x
MO2	9	57045	9	57045	99.9965
MO3	1	58695	1	58695	100

MO1 was unable to undergo a second iteration of short-read polishing and generate a consensus quality score (Table 10), thus the Polypolish polishing step was taken as the final iteration of the assembled genome with 1 contig at 50013 bp. MO2 and MO3 remained the same after going through polishing and were assigned quality scores of 99.9965 and 100 respectively. These were taken as the final assembled genomes from the HYPPA workflow and were as follows: MO2, 9 contigs at 57045 bp total and MO3, 1 contig at 58695 bp. To validate the assemblies, they were compared to the genome assemblies completed using Shovill and Illumina short reads (Table 11).

Table 11 – Comparison of Shovill and HYPPA assemblies for phages MO1, MO2 and MO3. * - the first contigs were taken as representing most of the genome of the respective phages and used for further downstream analysis.

	Genome length (bp)		No. of contigs assembled	
	Shovill	**HYPPA**	**Shovill**	**HYPPA**
MO1	172146	50013	87*	1
MO2	50130	57053	1	9
MO3	46838	58695	7*	1

Assembled genome lengths of total contigs were bigger when assembled with the HYPPA workflow for MO2 and MO3 while smaller for MO1 (Table 11). MO1 has the largest size difference between the two assemblies, with approximately 120kb fewer base pairs from the HYPPA workflow. All 3 phages assembled into a different number of contigs, with MO1 having the greatest difference between 87 from the Shovill assembly and 1 from the HYPPA assembly and MO2 being the only phage to increase

in the number of contigs. Phage contigs from both assemblies were run through the nucleotide basic local alignment search tool (BLASTN) (Altschul, et. al., 1990) against each other to identify similarities.

Table 12 – BLASTN comparison of assembled genomes from Shovill and HYPPA workflow. QC – query cover, the percentage of the query that matches the result. Note, the underlined assemblies were made the subjects of the comparison, and their respective alternative assembly was the query when input.

		MO1 (Shovill)	**MO2 (Shovill)**	**MO3.1 (Shovill)**
MO1 (HYPPA)		No significant similarity found		
MO2 (HYPPA)	Contig 5		100% QC / 100% Identity	
	Contig 21		No significant similarity	
	Contig 3		100% QC / 100% Identity	
	Contig 6		100% QC / 100% Identity	
	Contig 4		100% QC / 100% Identity	
	Contig 10		100% QC / 100% Identity	
	Contig 8		100% QC / 100% Identity	
	Contig 20		No significant similarity	
	Contig 9		100% QC / 100% Identity	
MO3 (HYPPA)				88% QC / 100% identity

BLASTN results determined that MO1 assembled with HYPPA had no significant similarity to the Shovill assembly (Table 23). MO3.1 was found that 88% of its assembled genome was 100% identical to MO3 assembled via HYPPA. Of the 9 contigs assembled via HYPPA for MO2, all but two (Contigs 21 and 20) had 100% query cover that was 100% identical. Contigs 21 and 20 reported no significant similarity to MO2 assembled via Shovill. As a draft genome the first and largest contig of MO2 was taken forward as representing the phage genome – contig 3 at 35984 bp. This was done while knowing that MO2 would require a full genome assembly of all contigs but contig 3 was to be used in the meantime. All phages were run through BLASTN to identify closest relatives of each phage and for comparison based on nucleotide similarity.

Table 13 – First 7 BLASTN results of phages MO1, MO2 and MO3 assembled with HYPPA. Only results of complete phage genomes were considered.

MO1 - HYPPA				
Phage name	**Query cover (%)**	**% Identity**	**Length (bp)**	**Accession Number**
Escherichia phage Isaaklselin	87	94.78	49.9k	MZ501077.1
Klebsiella phage vB_KaS-Gatomon	86	92.83	49.7k	LR881110.1
Klebsiella phage vB_KaS-Ahsoka	86	92.83	49.7k	LR881108.1
Klebsiella phage vB_KppS-Samwise	86	92.83	49.8k	OY639338.1
Escherichia phage JohannLBurckhardt	86	92.83	49.8k	MZ501085.1
Escherichia phage Henu7	85	91.17	49.5k	NC_054894.1
Enterobacter phage N5822	81	91.21	78.1k	MW032452.1
MO2 - HYPPA				
Escherichia phage Isaaklselin	87	95.53	49.9k	MZ501077.1
Klebsiella phage vB_KppS-Samwise	84	92.84	49.7k	OY639338.1
Klebsiella phage vB_KaS-Gatomon	84	92.84	49.7k	LR881110.1
Escherichia phage JohannLBurckhardt	84	92.84	49.7k	MZ501085.1
Escherichia phage Henu7	84	91.17	49.5k	NC_054894.1
Enterobacter phage N5822	78	90.80	78.1k	MW032452.1
Escherichia phage BEC3	31	78.32	51031	MN158217.1
MO3 - HYPPA				
Escherichia phage Isaaklselin	85	94.70	49.9k	MZ501077.1
Escherichia phage JohannLBurckhardt	84	92.90	49.7k	MZ501085.1
Klebsiella phage vB_KaS-Gatomon	86	92.85	49.7k	LR881110.1
Klebsiella phage vB_KaS-Ahsoka	86	92.85	49.7k	LR881108.1
Klebsiella phage vB_KppS-Samwise	86	9285	49.7k	OY639338.1
Escherichia phage Henu7	83	91.16	49.5k	NC_054894.1
Enterobacter phage N5822	80	91.21	78.1k	MW032452.1

MO1 and MO3 returned the same 7 BLASTN results for phage relatives with highly similar query cover and identity (Table 13_. Of these results only one, *Enterobacter* phage N5822, belonged to the same *Enterobacter* genus as the phages, with other relatives spanning *Klebsiella* and *Escherichia*. For the same relatives, MO1 and MO3 had very similar percentages query cover, differing by a few percent. The same can be said for the percentage identity, where the difference was even more marginal and differing by .05%. MO2 also returned similar relatives at a similar percentage query cover and identity compared to MO1 and MO3. All phages appeared to have similar phage relatives – to determine genomic similarities, the Virus Intergenomic Distance Calculator (VIRIDIC; Moraru, et. al., 2020) was used to produce pairwise intergenomic similarities amongst viruses. The phage results with significant similar alignments to each isolated phage (Table 13) were used alongside additional relatives from each phage. In total, 20 relatively similar phages were compared amongst each other and to MO1, MO2 and MO3.

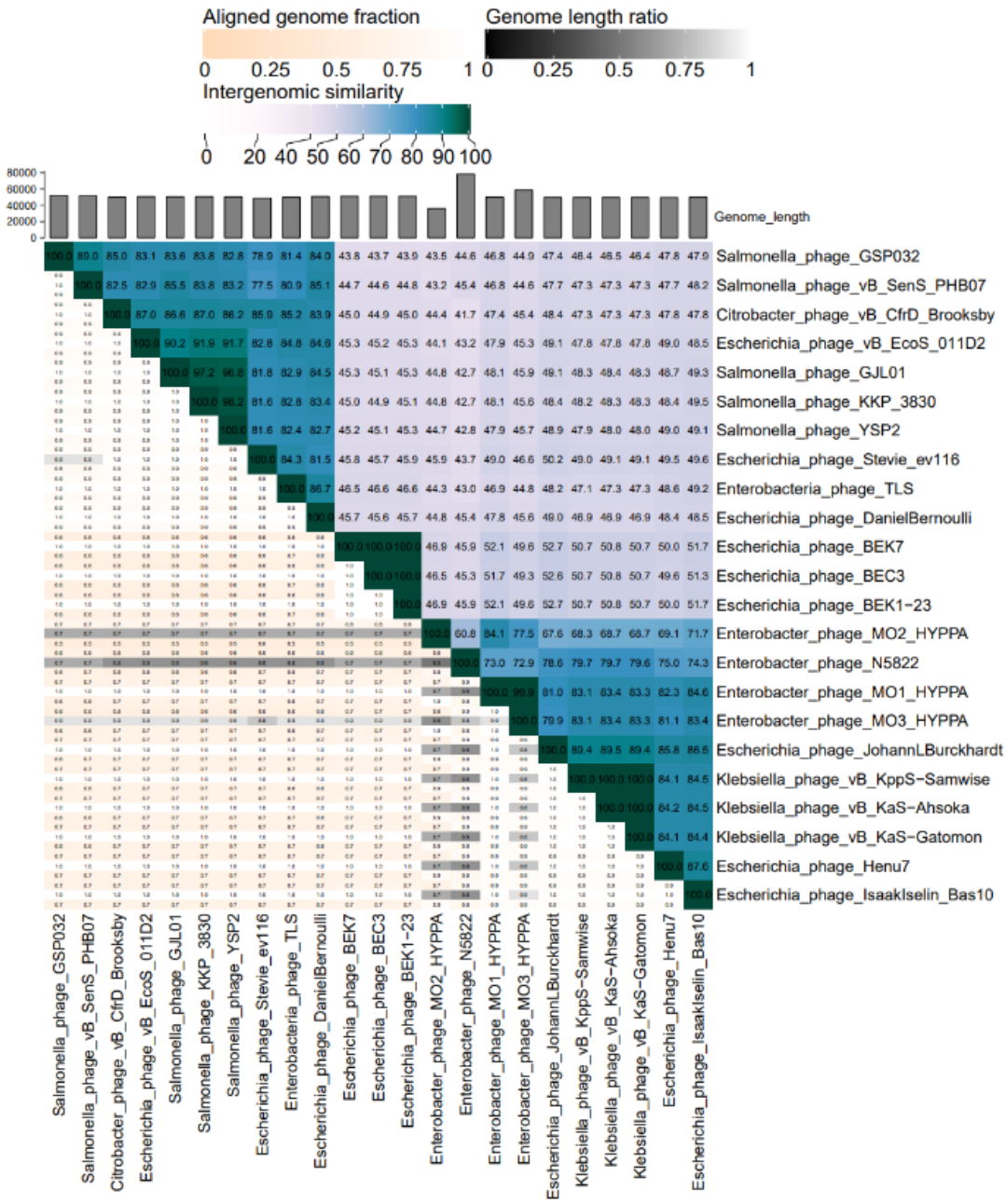


Figure 12 - Nucleotide-based intergenomic similarities of MO1, MO2 and MO3 assembled with HYPPA with a selection of related phages (based on query cover % and identity %) using VIRIDIC. A heatmap of the intergenomic similarity values was generated and given as percentages (right half, blue green heatmap). Each genome pair is represented by three values (left half), where the top and bottom represent the aligned genome fraction for the genome in the row and column, respectively, where darker colour indicates that a lower fraction of the genome was aligned. The middle value (grey scale) represents the genome length ratio for each genome pair, where darker colour indicates increasing distance between phages.

MO1, MO2 and MO3 clustered together alongside the previously identified phage relatives (Figure 12, Table 13). MO1 was found to have 99% intergenomic similarity with MO3, heavily suggesting that the two may be the same phage. MO2 and MO3 reported 77.5% identity to each other. All 3 phages clustered with ~80% intergenomic similarity to their closest relatives (bottom right corner) spanning the *Escherichia, Klebsiella* and *Enterobacter* genera. 40 – 50% intergenomic similarity was found with relatives of phages predominantly of *Salmonella* as well as a *Citrobacter* phage and additional *Escherichia* phages.

Based on the outputs of both the Illumina and HYPPA assemblies, the hybrid and poly-polish approach of HYPPA, and the comparison between the two it was decided that this would be the complete method for genome assembly. As MO1's HYPPA assembly was inconsistent with the Shovill assembly and near identical to MO3's HYPPA assembly it was decided that MO1 assembled with Shovill would be taken forward as the assembled genome for MO1 instead (Table 14). MO2 and MO3 assembled with HYPPA were taken forward as the assembled genomes. MO1 and MO2 are both draft genomes, requiring further complete genome assembly with additional assembled contigs.

Table 14 – Final phage assembly results of MO1, MO2 and MO3. * - MO1 and MO2 are draft genomes, both assembled into multiple contigs that could be considered as making up parts of the phage genome and require complete genome assembly. The first and largest contigs of each were taken as representative of their phage genomes.

Phage	Assembly	Total number of contigs assembled	Length of genome (bp)
MO1	Shovill	87*	172146
MO2	HYPPA	9*	35985
MO3	HYPPA	1	58695

4.4 *E. cloacae* phage Genome Annotation

The assembled contigs of MO1, MO2 and MO3 were initially annotated with prokaryotic genome annotation tool Prokka (Seeman, 2014). Prokka performs gene calling with start and stop codons by using Prodigal (Hyatt, 2010), identifying coordinates of candidate genes. These are then compared to the UniProt database (The UniProt Consortium, 2021) and assigned functions. In the event no matches can be found, genes are labelled as 'hypothetical proteins'. Prokka produces various output files, including genbank (.gbk) files containing sequences and annotations. Annotations with Prokka resulted in many protein products being annotated as 'hypothetical'. To resolve this, the bacteriophage-based prokaryotic genome annotation tool Pharokka (Bouras, et. al., 2023) was used to improve genome annotations (Table 15).

#

Table 15 – Comparison of annotations of phages MO1, MO2 and MO3 using Prokka and Pharokka. Coding sequences (CDS) are assigned coordinates and then compared to a reference database to be assigned functions. CDS that can't be matched within the database are labelled as hypothetical or unknown proteins. The % of the genome that remains hypothetical or unknown is noted.

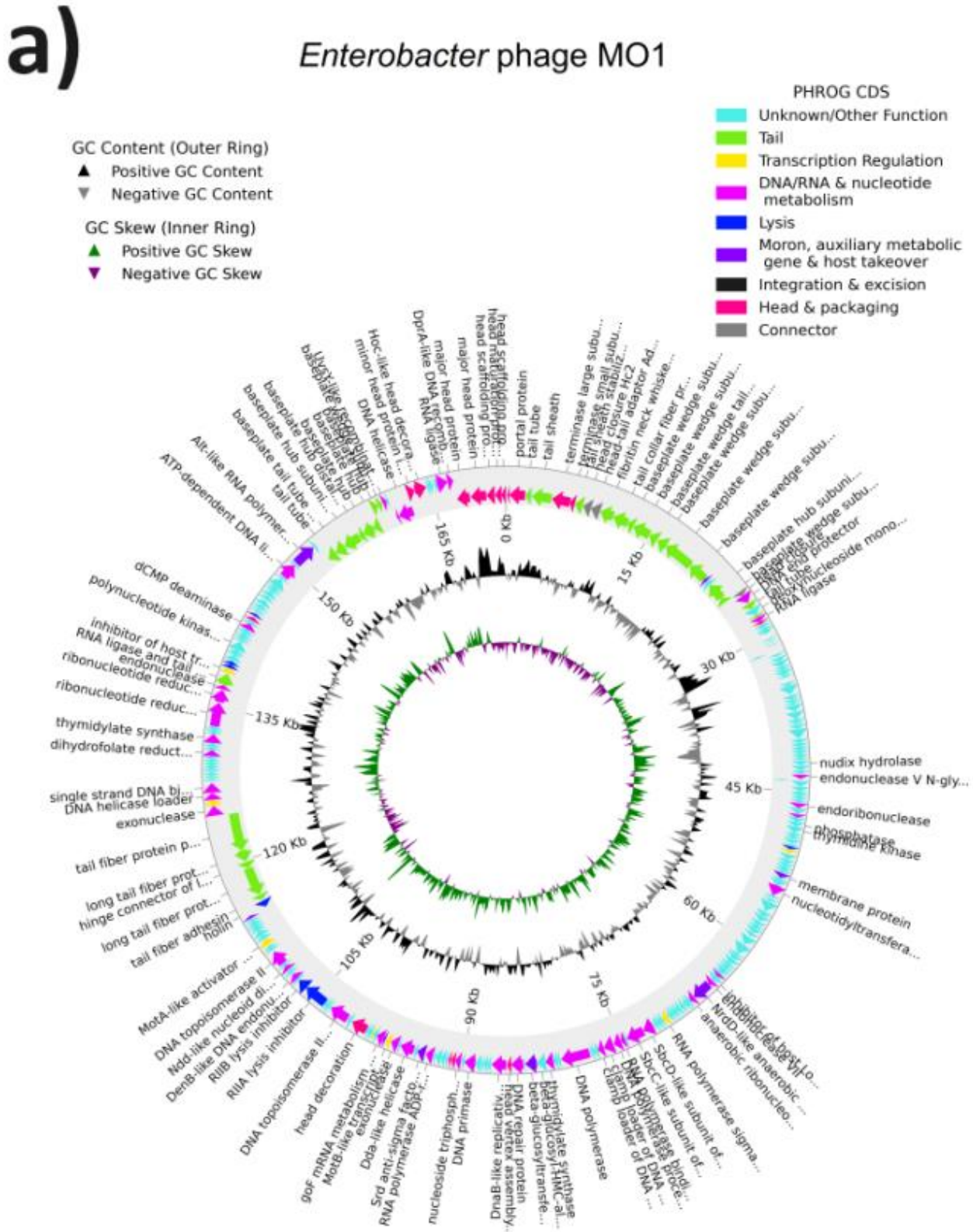
MO1		
	Prokka v1.14.5	**Pharokka v1.3.2**
CDS	289	305
Annotated functional CDS	19	129
Hypothetical/Unknown CDS	270	176
Hypothetical/Unknown % of genome	93.4	57.7
MO2		
	Prokka v1.14.5	**Pharokka v1.3.2**
CDS	50	59
Annotated functional CDS	3	28
Hypothetical/Unknown CDS	47	31
Hypothetical/Unknown % of genome	94.0	52.5
MO3		
	Prokka v1.14.5	**Pharokka v1.3.2**
CDS	116	130
Annotated functional CDS	2	36

Hypothetical/Unknown CDS	114	94
Hypothetical/Unknown % of genome	98.3	72.3

Using Pharokka resulted in the labelling of more CDS, more annotated CDS and less hypothetical CDS than compared to using Prokka for all phages, demonstrated by a lower % of the genome being hypothetical when annotated with Pharokka (Table 15) The most notable difference is seen with MO1 – where Prokka initially produced 289 CDS, 19 annotated and 270 hypothetical, Pharokka produced 305 CDS, 129 annotated and 176 hypothetical, going from 93.4% of the genome being hypothetical to 57.7%.

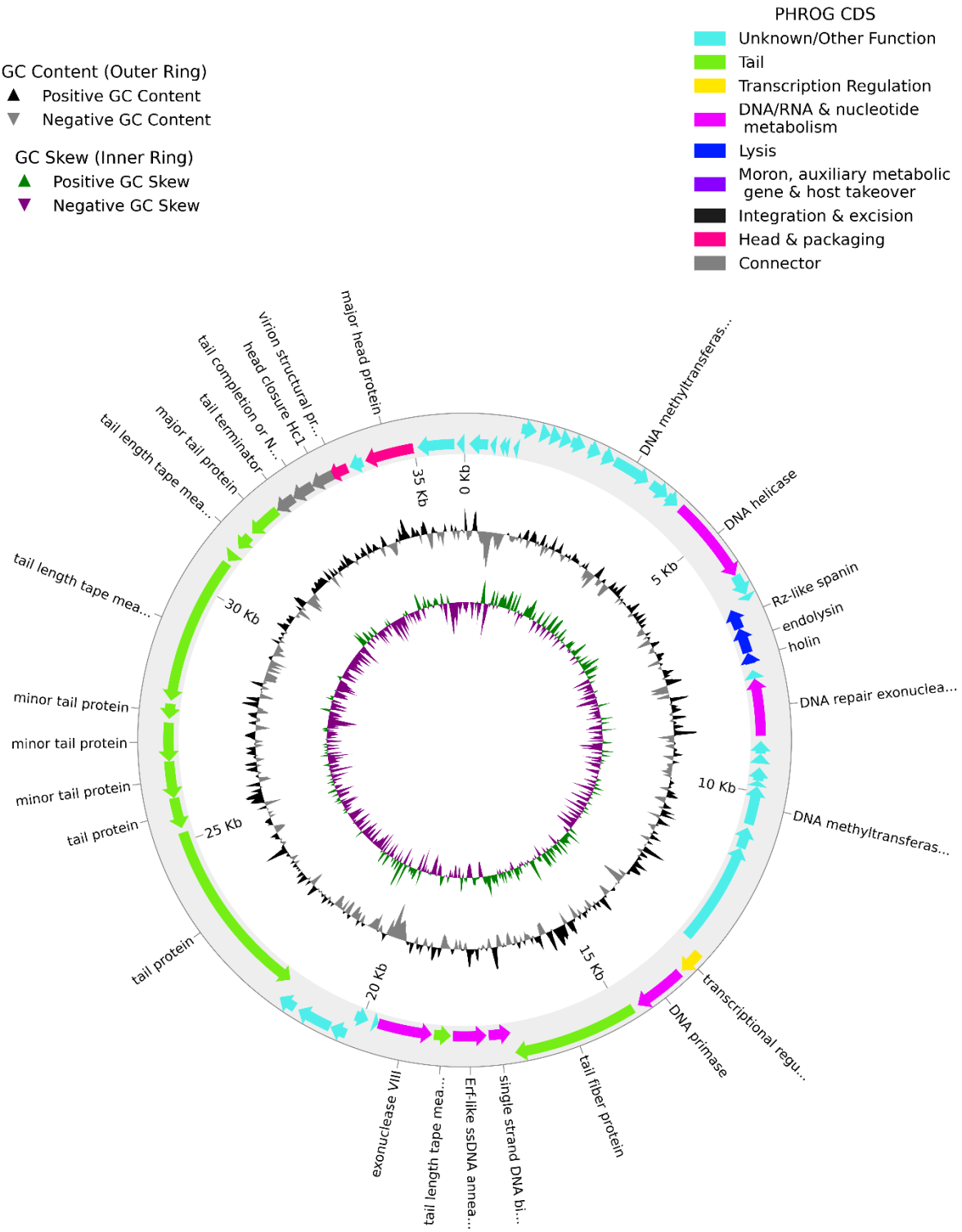
Table 16 – Functional group descriptions of phages MO1, MO2 and MO3 with Pharokka. Pharokka uses PHANOTATE (McNair, et. al., 2019), a gene prediction program tailored specifically to bacteriophages. Following gene prediction, functional annotations are assigned by matching each predicted CDS to the following databases: the Prokaryotic Virus Remote Homologous Groups (PHROGs) database, the Comprehensive Antibiotic Resistance Database (CARD) and the Virulence Factors of Pathogenic Bacteria (VFDB) database.

Function description	MO1	MO2	MO3
CDS	305	59	130
Connector	3	3	3
DNA, RNA & nucleotide metabolism	44	6	6
Head & packaging	17	2	6
Integration & excision	0	0	0
Lysis	7	3	3
Moron, auxiliary metabolic gene, and host takeover	8	0	0
Other	13	2	6
Tail	29	11	11
Transcription regulation	8	1	1
Unknown function	176	31	94
tRNAs	19	0	0
Virulence factors	0	0	0
Antimicrobial resistance genes	0	0	0


a)
Enterobacter phage MO1
PHROG CDS
Unknown/Other Function
Tail
Transcription Regulation
DNA/RNA & nucleotide metabolism
Lysis
Moron, auxiliary metabolic gene & host takeover
Integration & excision
Head & packaging
Connector
GC Content (Outer Ring)
Positive GC Content
Negative GC Content
GC Skew (Inner Ring)
Positive GC Skew
Negative GC Skew
0 Kb
15 Kb
30 Kb
45 Kb
60 Kb
75 Kb
90 Kb
105 Kb
120 Kb
135 Kb
150 Kb
165 Kb
nudix hydrolase
endonuclease V N-gly...
endoribonuclease
phosphatase
thymidine kinase
membrane protein
nucleotidyltransfera...
RNA polymerase sigma...
SbcD-like subunit of...
DNA polymerase
DNA repair protein
DNA primase
nucleoside triphosph...
Dda-like helicase
DNA topoisomerase II
MotA-like activator
hollin
tail fiber adhesin
long tail fiber prot...
hinge connector of l...
tail fiber protein p...
exonuclease
DNA helicase loader
single strand DNA bi...
dihydrofolate reduct...
thymidylate synthase
ribonucleotide reduc...
endonuclease
RNA ligase and tail
inhibitor of host tr...
polynucleotide kinas...
dCMP deaminase
ATP-dependent DNA li...
Alt-like RNA polymer...
baseplate tail tube ...
baseplate hub subuni...
baseplate hub distal...
minor head protein
Hoc-like head decora...
DNA helicase
DprA-like DNA recomb...
RNA ligase
major head protein
head scaffolding pro...
portal protein
tail tube
tail sheath
terminase large subu...
terminase small subu...
head closure HC2
head-tail adaptor Ad...
fibritin neck whiske...
tail collar fiber pr...
baseplate wedge subu...
baseplate wedge tail...
baseplate wedge subu...
baseplate wedge subu...
baseplate hub subuni...
baseplate wedge subu...
head closure
tail tube protector
oxynucleoside mono...
RNA ligase

b)

Enterobacter phage MO2



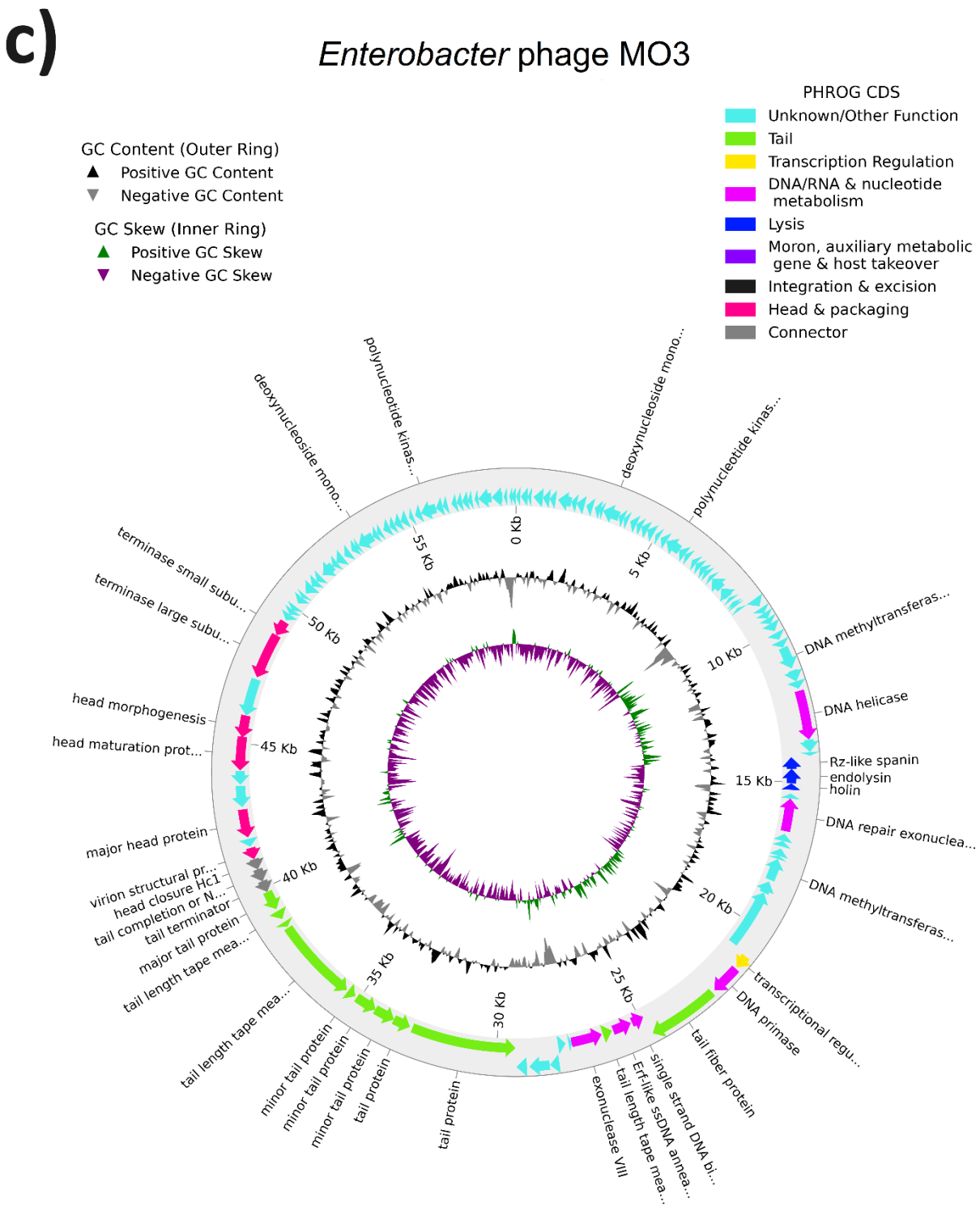


Figure 13 – Circular genome plots of phages MO1 **(a)**, MO2 **(b)** and MO3 **(c)**. CDS are grouped by function and coloured accordingly (key, top right). Note: MO1 has been fully annotated but only 80% of annotations are represented in **(a)**. The "unknown function" and "other function" categories were plotted together. Plots were created using Pharokka (Bouras, et. al., 2023).

Annotations with Pharokka revealed descriptions of functions for all phages (Table 16; Figure 13). All phages were predicted the same number of connector genes, despite the differences in genome size. No phages had predicted genes for integration & excision, virulence factors or antimicrobial resistance.

MO1 was the only phage to have predicted genes for transfer RNAs (tRNAs) and auxiliary metabolic genes that may assist in host takeover. Of the genes annotated, most within MO1 were predicted to be associated with DNA, RNA & nucleotide metabolism. For MO2 and MO3, most annotated genes were predicted as being associated with the tail.

In initial assembly of MO3 with Shovill, two separate contigs had been assembled and circularized (Figure 11c) implying the possibility of 2 separate complete phage genomes or both making up the entire phage genome of MO3. MO3.1 (Figure 11c1) was initially taken as making up the phage genome of MO3 and MO3.2 (Figure 11c2) was annotated with Pharokka (Bouras, et. al., 2023) to investigate if this could also be part of the phage genome.

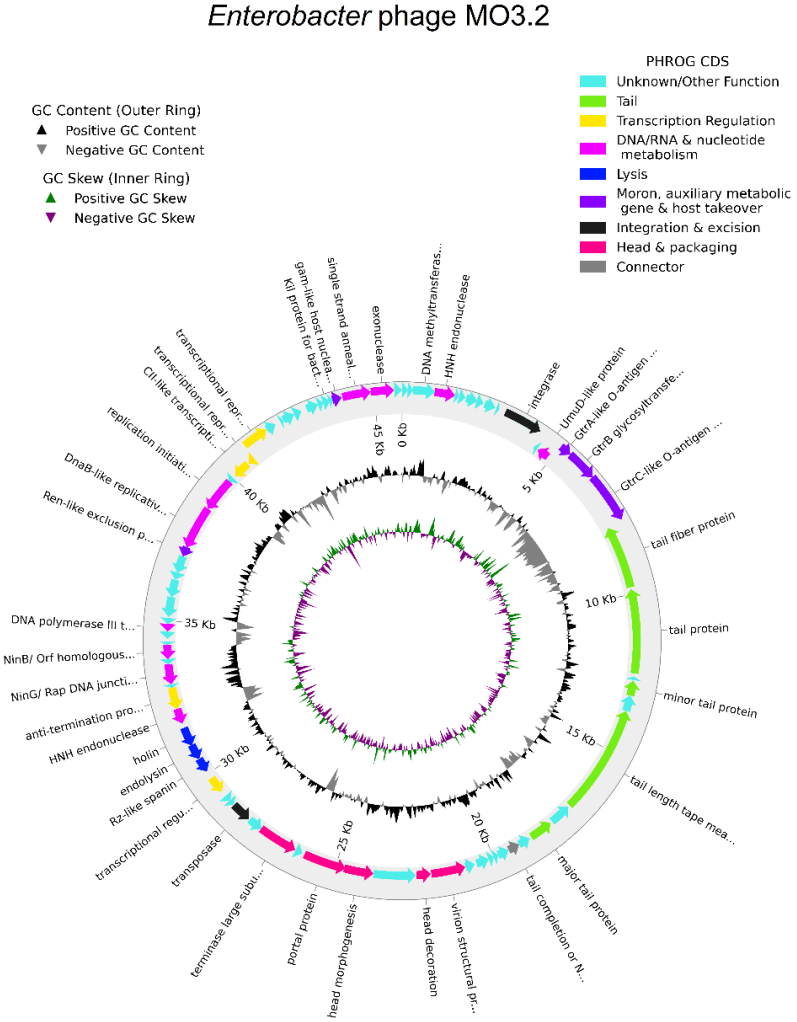


Figure 14 – Circular genome plot of MO3.2. Plot was created using Pharokka (Bouras, et. al., 2023).

MO3.2 contains many expected and similar genes compared to the other annotated phages, including tail proteins, portal proteins and transcription regulation (Figure 14). The presence of a predicted integrase gene however, normally present in temperate phages for host genome integration, suggested the strong likelihood of MO3.2 being a prophage that had been induced during the isolation process. To investigate this further, several tools for predicting prophages and prophage regions were used. Phage Search Tool Enhanced Release (PHASTER; Arndt, et. al., 2016) identifies and annotates prophage sequences within bacterial genomes and plasmids, assigning a score based on parameters that indicate the likelihood of a region belonging to a prophage.

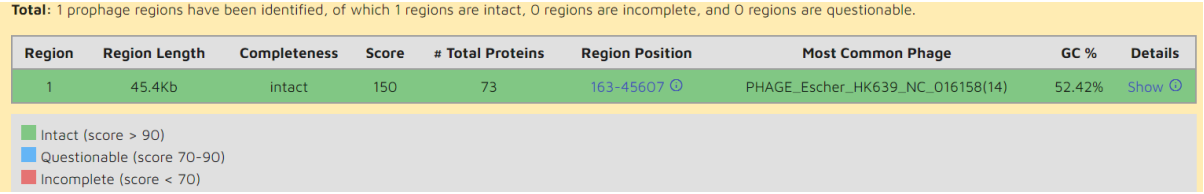


Figure 15 – PHASTER results for MO3.2, predicting the likelihood of the region being a prophage. PHASTER (Arndt, et. al., 2016) uses a custom collection of phage and prophage protein sequences from the NCBI phage database and a prophage database (Srividhya, et. al., 2006) to identify potential prophage -associated protein sequences. A completeness score is assigned from 0 – 150 to the selected region based on two scenarios: i) the clustering/region that only contains genes/proteins of a known phage and ii) if >50% of the genes/proteins in the region are related to a known phage. The presence of prophage-associated genes, e.g., integrase, increases the score and the likelihood the region is a prophage.

Table 17 – BLASTN results of *Enterobacter cloacae* NC10005 (subject) against MO3.2 (query).

	MO3.2
***Enterobacter cloacae* NC10005**	Identity = 41565/41565 bp (100%), Gaps = 0/41565 (0%)

BLASTN results returned that 41.5k bp of MO3.2 were 100% identical to *E. cloacae* NC10005, confirming that the prophage region had originated from the bacterium and induced during the isolation process. As such, MO3.2 was not considered in the final genome assembly of MO3 and considered to be a prophage.

4.5 Comparative Genomics of *E. cloacae* phages

MO1 was run through BLASTN (Altschul, et. al., 1990), to search for nucleotide regions of similarity to publicly available virus genomes, identifying any significant sequence homology to other phages. The entire genome length and >95% sequence identity was used to determine if any sequences with specific alignments belonging only to complete assembled phage genomes could be within the same species. This was previously done for MO2 and MO3 (Table 13).

Table 18 – First 7 BLASTN results of MO1. Only results of complete phage genomes were considered.

MO1				
Phage name	**Query cover (%)**	**% Identity**	**Length (bp)**	**Accession Number**
Enterobacter phage vB_VIPECLOM01	97.00	99.30	171,903	OQ721912.1
Enterobacter phage vB_VIPECLUM02	97.00	99.29	172,129	OQ721913.1
Enterobacter phage pg7	95.00	99.14	172,376	NC_023561.1
Enterobacter phage fGh-Ecl01	98.00	99.14	171,663	ON212265.1
Enterobacter phage fGh-Ecl04	97.00	98.53	172,419	ON212267.1
Enterobacter phage vB-EclM_KMB19	95.00	95.53	172,697	OL828290.1
Klebsiella phage vB_KpnM_VPA32	94.00	96.29	172,331	OP558005.1

BLASTN results returned several *Enterobacter* phages of a high query cover and % identity to MO1 in addition to a *Klebsiella* phage (Table 18). *Enterobacter* phages vB_VIPECLOM01, vB_VIPECLUM02, PG7

and fGh-Ecl01 were all >99% identical to MO1 with >95% query cover. The lengths of all results are comparable to MO1 (172,146 bp), within 1000 bp. Genbank files of the most closely related phage genomes were downloaded and compared using VIRIDIC (Moraru, et. al., 2020) for a comparative genome analysis alongside the previously found relatives of MO2 and MO3. In total, 24 relatives were compared to MO1, MO2 and MO3.

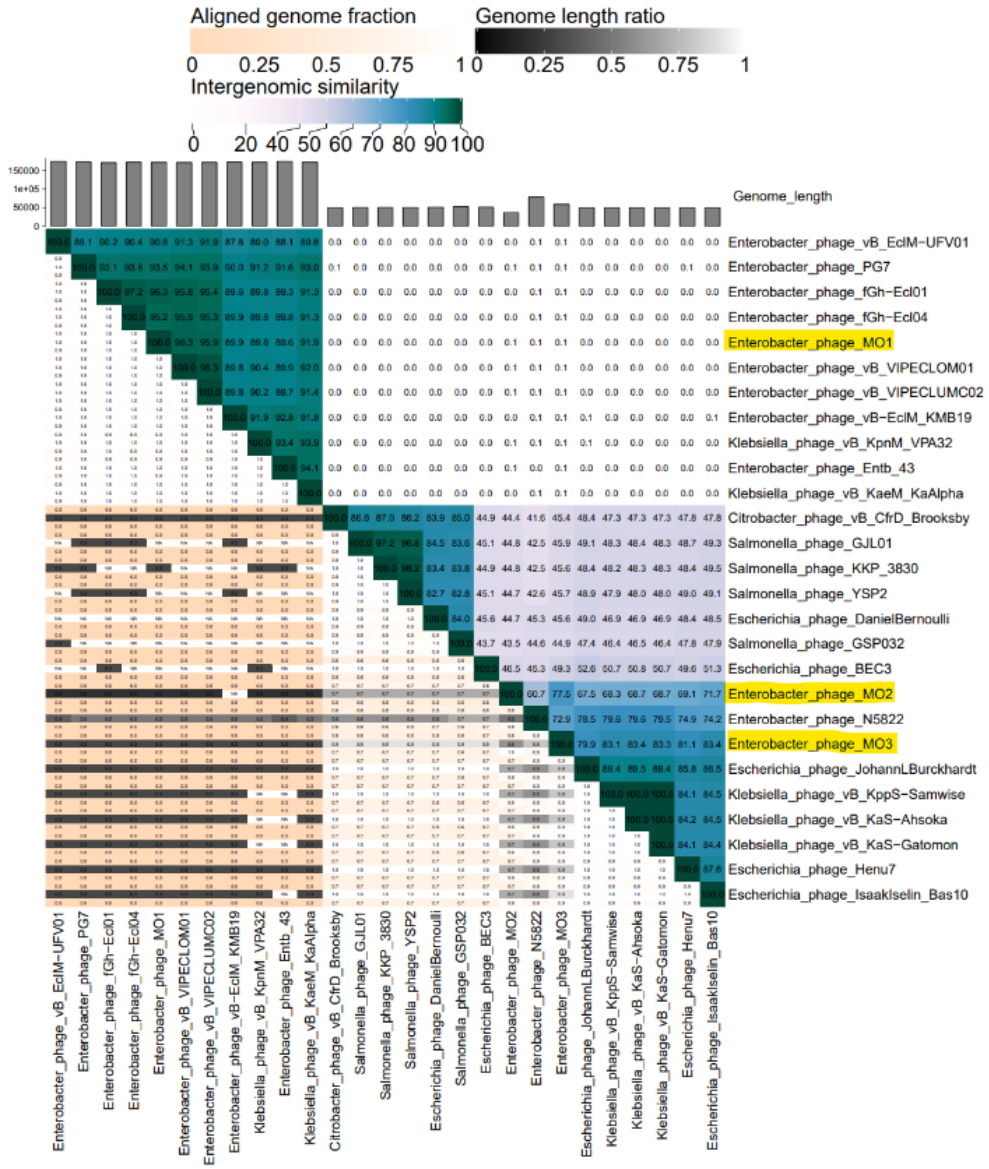


Figure 16- Nucleotide-based intergenomic similarities of phages MO1, MO2 and MO3 (highlighted in yellow) with a selection of related phages (based on query cover % and identity %) using VIRIDIC. A heatmap of the intergenomic similarity values was generated and given as percentages (right half, blue green heatmap). Each genome pair is represented by three values (left half), where the top and bottom represent the aligned genome fraction for the genome in the row and column, respectively, where a darker, tan colour indicates that a lower fraction of the genome was aligned. The middle value (grey scale) represents the genome length ratio for each genome pair, where darker colour indicates increasing distance between phages.

MO1 clustered with its relative distinctly from MO2 and MO3, of which both clustered together with their respective relatives (Figure 16). In comparison to each other, MO2 and MO3 had were 77.5% identical and both phages were ~0.1% identical to MO1, showing a stark difference. Relatives of MO2 and MO3 had a high aligned genome fraction (~0.7 - ~0.9). MO1 and its relatives (top half, left)

clustered together with >88% intergenomic similarity amongst themselves. Aligned genome fraction between relatives of MO1 and those of MO2 & MO3 was lower, indicated by darker, tan colour and darker greyscale. VIRIDIC results showed a distinct lack of intergenomic similarity between MO1 and MO2/MO3, suggesting the belonging to different genera. To investigate the taxonomy of MO1, MO2 and MO3 and their related relatives, VIPTree (Nishimura, et. al., 2017) was used to generate a "proteomic tree" of viral genome sequences based on genome-wide sequence similarities computed by tBLASTx and compared to the Virus-Host database (Mihara, et. al., 2016). All phages previously used in the VIRIDIC analysis (Figure 16) were uploaded to VIPTree (Figure 17).

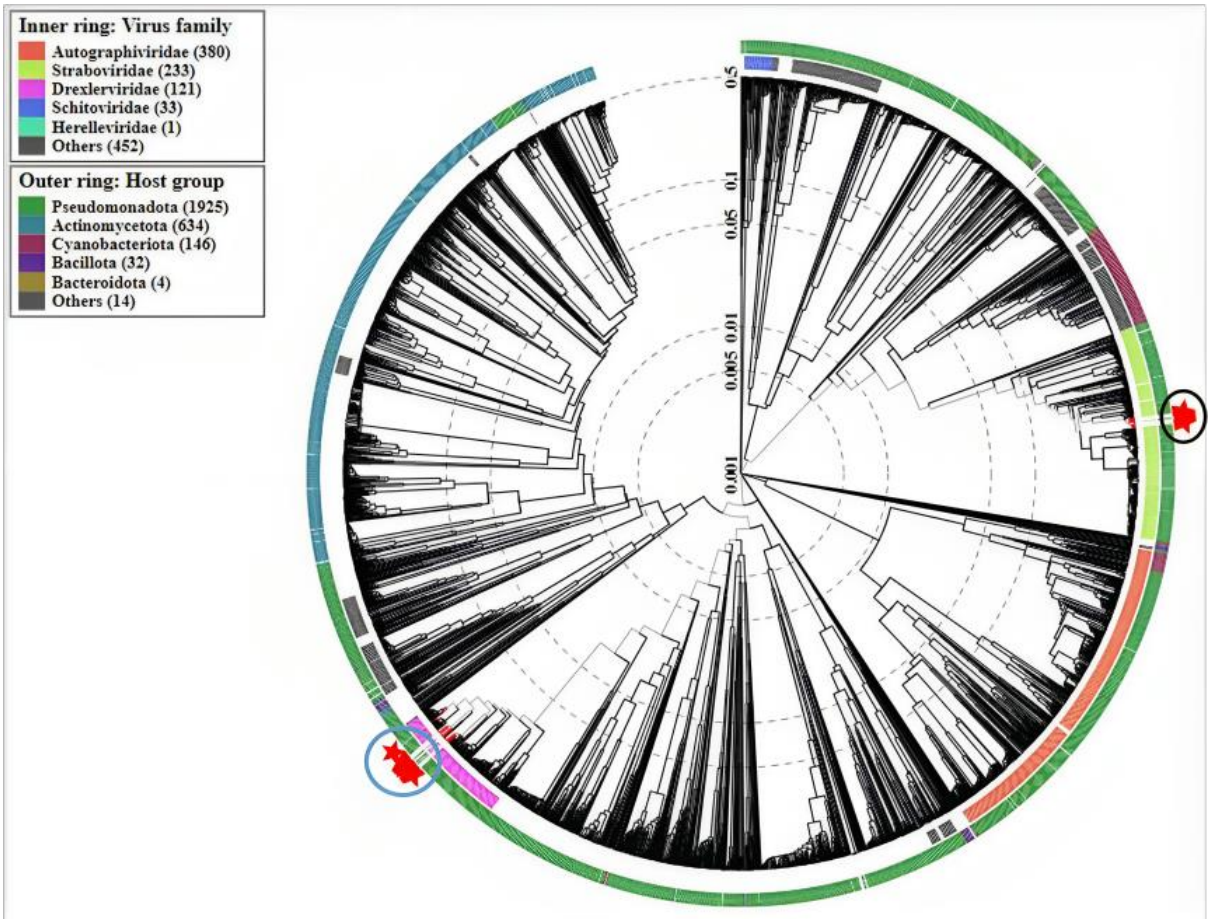


Figure 17 – Proteomic phylogenetic tree of MO1, MO2, MO3 and related phages alongside additional prokaryotic double-stranded (ds) DNA viruses created by VIPTree (Nishimura, et. al., 2017). Inner and outer rings display virus families and host group phyla respectively and are colour-coded according to key (top left). Red stars – uploaded viral sequences of MO1, MO2, MO3 and phage relatives. Blue circle – cluster of MO2, MO3 and associated phage relatives. Black circle – cluster of MO1 and associated phage relatives. Viral sequences are clustered and ordered according to a similarity score (inner numbers) from 0 – 1 based on % identity and query cover %.

The clustering of phages relative to their predicted taxonomical organisation can be observed in Figure 17. Most noticeably, two distinct clusters of phages were observed. The blue circle contains phages MO2, MO3 and their relatives while the black circle contains MO1 and its relatives. Both clusters belong to the same host group phyla, Pseudomonadota (previously named Proteobacteria, renamed in 2021). MO2, MO3 and their associated relatives are predicted as belonging to the *Drexlerviridae* family of viruses. In contrast, MO1 and its relatives appear as belonging to the *Straboviridae* family. This furthers the distinction between MO1 and MO2/MO3 from each other, evidenced by the differential predicted taxonomic classification. Beyond the predicted families, the taxonomic classification of the closest relatives was observed within the NCBI database to identify the exact subfamily, genus, and species (Table 19) (Schoch, et. al., 2020).

Table 19 – Taxonomic classification of 10 relatives of MO1, according to the NCBI Taxonomy Browser from family to species (Schoch, et. al., 2020).

Phage name	**Accession number**	**NCBI Taxonomy ID**	**Family**	**Subfamily**	**Genus**	**Species**
Enterobacter phage vB_VIPECLUMC01	OQ721913.1	3034281	*Straboviridae*	*Tevenvirinae*	*Karamvirus*	Unclassified
Enterobacter phage vB_VIPECLUMC02	OQ721913.1	3034282	*Straboviridae*	*Tevenvirinae*	*Karamvirus*	Unclassified
Enterobacter phage pg7	NC_023561.1	1455074	*Straboviridae*	*Tevenvirinae*	*Karamvirus*	*Karamvirus* pg7
Klebsiella phage vB_KaeM_KAlpha	MN013084.1	2591367	*Straboviridae*	*Tevenvirinae*	*Karamvirus*	Unclassified
Enterobacter phage fGh-Ecl01	ON212265.1	2935060	*Straboviridae*	*Tevenvirinae*	*Karamvirus*	Unclassified
Enterobacter phage fGh-Ecl04	ON212267.1	2935062	*Straboviridae*	*Tevenvirinae*	*Karamvirus*	Unclassified
Enterobacter phage vB-EclM_KMB19	OL828290.1	2917963	*Straboviridae*	*Tevenvirinae*	*Karamvirus*	Unclassified
Enterobacter phage vB_EclM-UFV01	ON454249.1	2945895	*Straboviridae*	*Tevenvirinae*	*Karamvirus*	Unclassified
Klebsiella phage vB_KpnM_VPA32	OP558005.1	3041494	*Straboviridae*	*Tevenvirinae*	*Karamvirus*	Unclassified
Enterobacter phage Entb_43	ON585039.1	2946013	*Straboviridae*	*Tevenvirinae*	*Karamvirus*	Unclassified *Karamvirus*

All chosen relatives of MO1 have been classified as belonging to the *Straboviridae* family (Table 19). With further definition, they were also referred to as belonging to the *Tevenvirinae* subfamily, *Karamvirus* genus and are unclassified *Karamvirus* species except for *Enterobacter* phage pg7 that has been classified. Given the clustering of MO1 with its relatives and their taxonomic classification, MO1 was tentatively classified from family to species as belonging to *Straboviridae, Tevenvirinae, Karamvirus* and unclassified *Karamvirus* species respectively. The taxonomy of MO2 and MO3's relatives were also looked at (Table 20).

Table 20 – Taxonomic classification of 14 relatives of MO2 and MO3, according to the NCBI Taxonomy Browser from family to species (Schoch, et. al., 2020). * - phage currently has no taxonomic classification between family and species level but the earliest taxonomy is class, of which KKP 3830 belongs to *Caudoviricetes*. **Bold** – phages that clustered with >60% intergenomic similarity to MO2 and MO3 in Figure 16.

Phage name	**Accession number**	**NCBI Taxonomy ID**	**Family**	**Subfamily**	**Genus**	**Species**
***Escherichia* phage Isaaklselin**	MZ501077.1	2851974	*Drexlerviridae*	*Tempevirinae*	*Henuseptimavirus*	Unclassified
***Klebsiella* phage vB_KppS-Samwise**	OY639338.1	2762815	*Drexlerviridae*	*Tempevirinae*	*Henuseptimavirus*	Unclassified
***Klebsiella* phage vB_Kas-Gatomon**	LR881110.1	2762813	*Drexlerviridae*	*Tempevirinae*	*Henuseptimavirus*	Unclassified
***Klebsiella* phage KaS-Ahsoka**	LR881108.1	2762814	*Drexlerviridae*	*Tempevirinae*	*Henuseptimavirus*	Unclassified
***Escherichia* phage JohannLBurckhardt**	MZ501085.1	2851975	*Drexlerviridae*	*Tempevirinae*	*Henuseptimavirus*	Unclassified
***Escherichia* phage Henu7**	NC_054894.1	2589652	*Drexlerviridae*	*Tempevirinae*	*Henuseptimavirus*	*Henuseptimavirus* henu7
***Enterobacter* phage N5822**	MW032452.1	2777323	*Drexlerviridae*	*Tempevirinae*	*Henuseptimavirus*	Unclassified
Escherichia phage BEC3	MN158217.1	2656624	*Drexlerviridae*	*Tempevirinae*	*Hanrivervirus*	Unclassified
Salmonella phage GSP032	ON855040.1	2962599	*Drexlerviridae*	*Tempevirinae*	*Tlsvirus*	Unclassified
Salmonella phage KKP 3830*	OQ674106.1	3027683	Unclassified	Unclassified	Unclassified	Unclassified *Caudoviricetes*
Escherichia phage DanielBernoulli	MZ501059.1	2851972	*Drexlerviridae*	*Tempevirinae*	*Tlsvirus*	Unclassified
Salmonella phage GJL01	KY657202.1	1965464	*Drexlerviridae*	*Tempevirinae*	*Tlsvirus*	*Tlsvirus* YSP2

Salmonella phage YSP2	NC_047898.1	2053686	*Drexlerviridae*	*Tempevirinae*	*Henuseptimavirus*	*Tlsvirus* YSP2
Citrobacter phage vB_CfrD_Brooksby	OL539443.1	2902661	*Drexlerviridae*	*Tempevirinae*	*Tlsvirus*	Unclassified

The taxonomy of the relatives for MO2 and MO3 (Table 20) was found to be broader compared to that of MO1 (Table 18) All relatives, aside from *Salmonella* phage KKP 3830, were classified as belong to the *Drexlerviridae* family and *Tempervirinae* subfamily. KKP 3830 was found to not be classified with any taxonomy from family to species but is within the class *Caudoviricetes* and, as such, can be said to be an unknown *Caudoviricetes* species for now. Most relatives were found to belong to the *Henuseptimavirus* genus and are unclassified *Henuseptimavirus* species. Of these, *Escherichia* phage Henu7 was the only one that was found to be species classified. Many of these phages (bold) were also found to cluster most closely to MO2 and MO3 with >60% intergenomic similarity (Figure 16) Phages that had a lower intergenomic similarity were found to belong to the genera *Hanrivervirus* or *Tlsvirus,* the latter of which contained two *Tlsvirus* YSP2 species. Based on the results of Figure 16 and Table 19, and the clustering of certain relatives, MO2 and MO3 were tentatively classified as belonging to the *Drexlerviridae* family, *Tempevirinae* subfamily, *Henuseptimavirus* genus and are unclassified *Henuseptimavirus* species. Given the genomic similarity between MO2 and MO3 further, clinker (Gilchrist, et. al., 2020) was used to search for and identify homologous gene clusters.

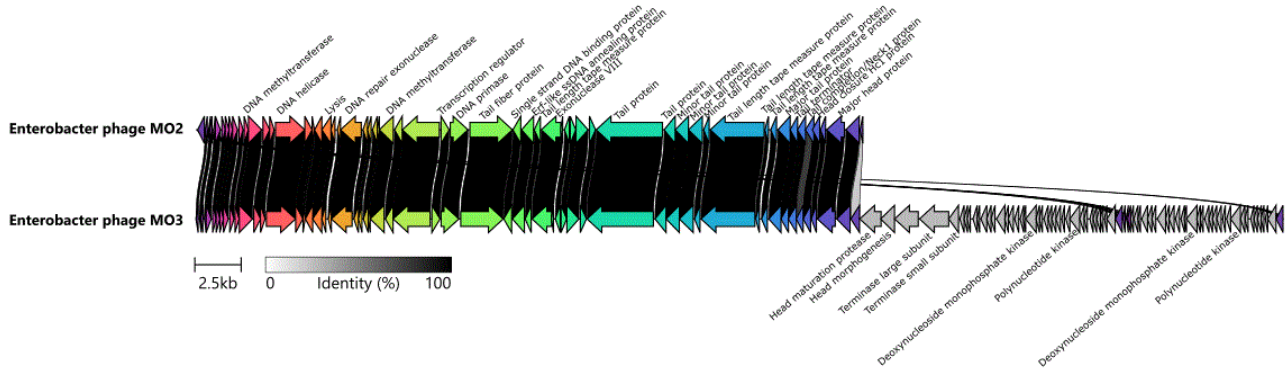


Figure 18 - Genome map and gene clustering for MO2 and MO3 created using clinker (Gilchrist, et. al., 2020). Arrows represent coding sequences and pairwise comparisons of gene similarities are indicated by percentage identity given as links in greyscale, with darker shading representing areas of higher similarity. Genes without any sequence similarity are indicated without links and are grey arrows. Hypothetical proteins were not annotated.

Genomic similarity and gene clustering between MO2 and MO3 was observed (Figure 18). MO2 shares all its genome with MO3 at ~100% identity, including the clustering of its genes. DNA-associated genes (helicase, methyltransferase) are clustered towards the start of the genome; genes encoding tail proteins towards the middle and head proteins towards the end. MO3 has a significant portion of its genome that it does not share with MO2, most likely due to MO2 being a draft genome with potential missing parts of its genome, and does not cluster to any genes, including additional genes associated with the formation of the phage head (maturation and morphogenesis). The presence of deoxynuceloside monophosphate kinases and polynucleotide kinases was also observed in MO3.

4.6 Isolation + Characterisation of novel *R. gnavus* phages

4.6.1 Development of *R. gnavus* lawn overlay assay

Before beginning to search for phages targeting *R. gnavus*, establishing a sufficient bacterial lawn for testing was required. Previously, there had been no recorded literature detailing a protocol for isolation of *R. gnavus* phages, including growth conditions of the bacteria or protocol to grow a lawn. Therefore, understanding the growth conditions was first required. *R. gnavus* ATCC 29149 was introduced to me by Dr Emmanuelle Crost and was originally grown from pre-reduced frozen 2% glycerol stocks in BHI-YH prepared by her, with a starter culture grown in anaerobic BHI-YH overnight from stock at 37 °C in strict anaerobic conditions. The culture was then inoculated into fresh BHI-YH and grown in the same conditions overnight to create a 2% subculture, culminating in a total 48-hour growth period as per the Juge lab protocols for growing *R. gnavus*. To test if *R. gnavus* could be grown on an agar plate, both the starter culture and subculture were streaked out onto 25 ml 1.5% BHI-YH agar plates that had been pre-reduced for 48 hours. Plates were then incubated for 72 hours at 37 °C in strict anaerobic conditions.

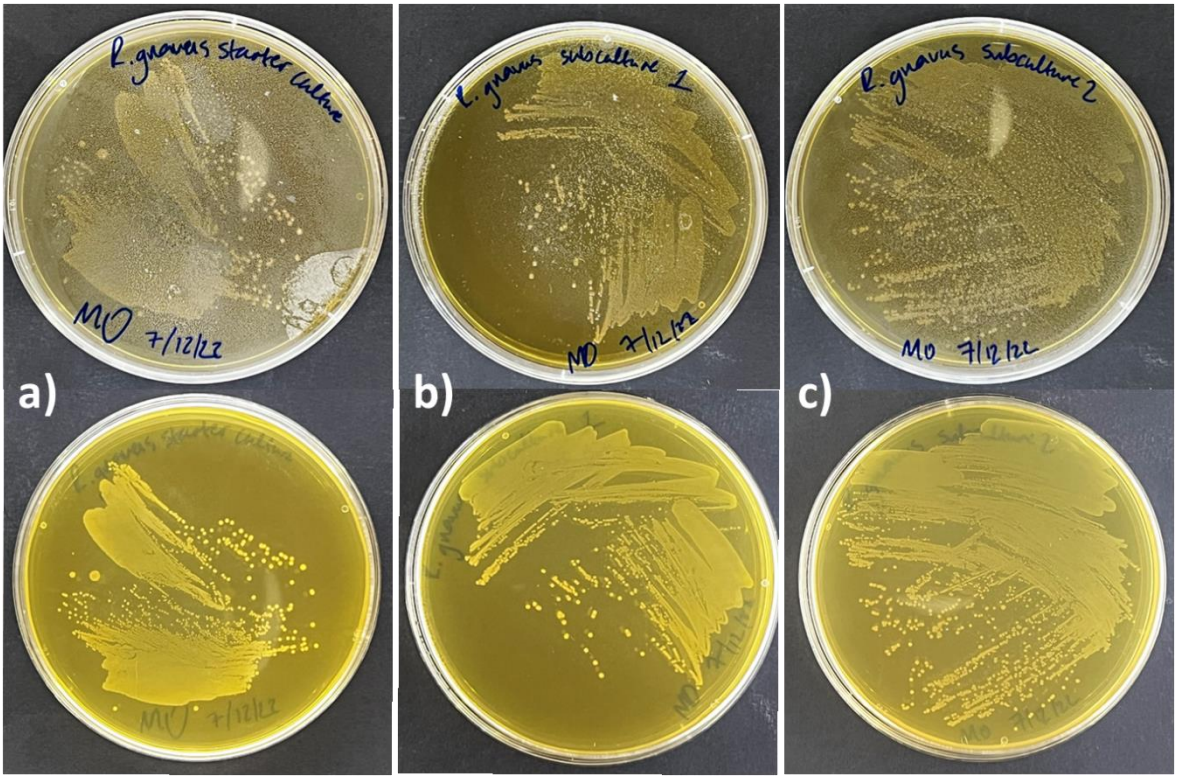


Figure 19 – *R. gnavus* ATCC 29149 4% starter culture **(a)** and 2% subculture **(b, c)** streaked on 1.5% BHI-YH agar, after 72 hours of incubation at 37 °C in strict anaerobic conditions, front and back of plates.

After confirming *R. gnavus* could form single colonies on plates (Figure 19) the double agar method used to isolate new *Enterobacter* phages was adapted for anaerobic use as previously described (section 3.8) but to investigate formation of bacterial lawn (as opposed to attempting to isolate phages at this stage).

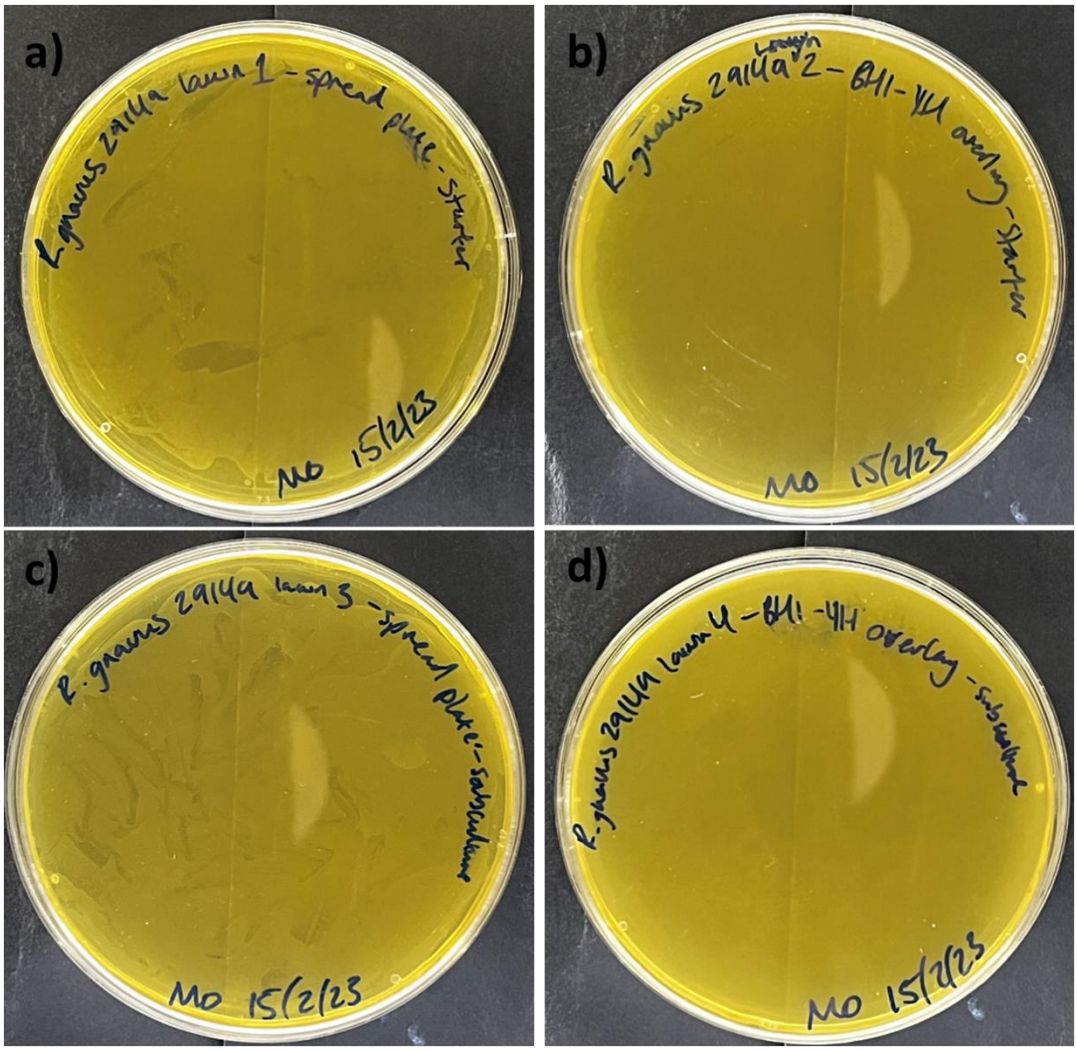


Figure 20 – Comparison of spread and overlay plates for *R. gnavus* for lawn formation. 200 µl of starter culture and subculture were poured onto agar plates and spread uniformly **(a, c)** or combined with 4 ml of cooled BHI-YH agar and poured onto plates **(b, d).**

Though both methods formed a semblance of a lawn, the spread plates (Figure 20a, c) did not form uniformly and were without an overlay agar and the overlay plates (Figure 20b, d) were not dense enough to be considered adequate lawns. the method was then adapted to include elements of the protocol of Buttimer and colleagues who reported the successful isolation of multiple temperate phages targeting *R. gnavus,* (Buttimer, et. al., 2023). Specifically, the use of anaerobe basal broth (ABB) media as opposed to BHI-YH. To determine the optimal media for use between the two and the optimal growth period, a comparison of both was done alongside Xena Dyball by growing a starter culture of *R. gnavus* ATCC 29149 from 2% BHI-YH glycerol stock overnight and inoculating into either ABB or BHI-YH to grow a subculture for 8 hours as opposed to 24 hours. Optical density (OD) at 600 nm, colony forming units and lawn formation was compared.

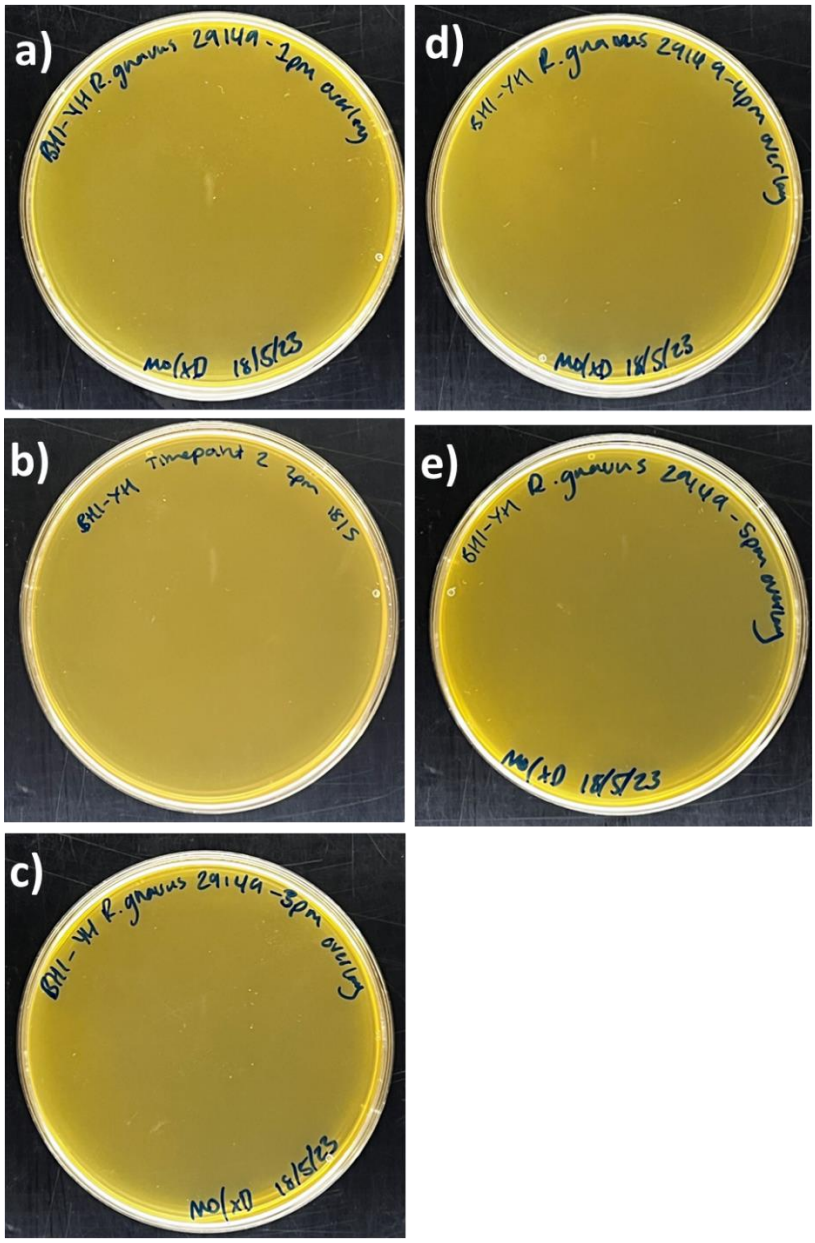


Figure 21 – Agar overlay lawns of *R. gnavus* ATCC 29149 with 0.5% BHI-YH. Lawns were poured out following subculture growth for a period as follows: **a)** at 1pm (t=4 hours), **b)** at 2pm (t=5 hours), **c)** at 3pm (t=6 hours), **d)** at 4pm (t=7 hours), **e)** at 5pm (t=8 hours). This was done once alongside Xena Dyball.

Assessing lawn formation of *R. gnavus* ATCC 29149 in BHI-YH (Figure 21) there was no distinct sign of the growth of a lawn. No clear differences were observed between the different time points to indicate that a lawn had formed.

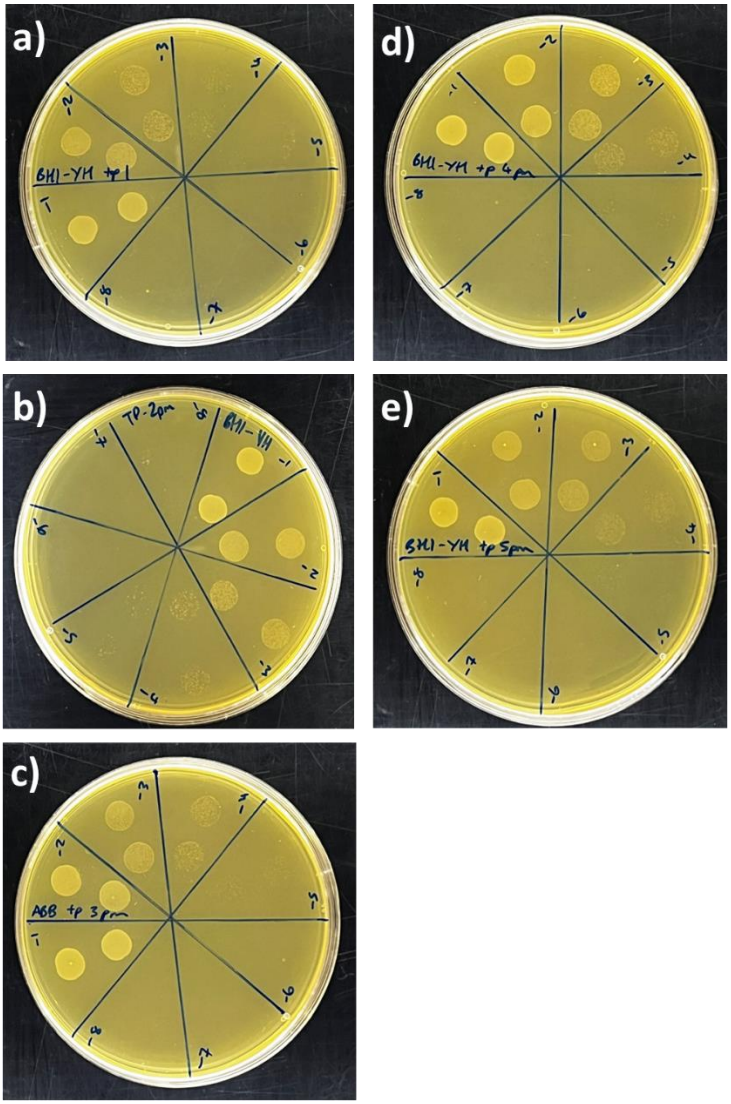


Figure 22 – Dilution series of *R. gnavus* ATCC 29149 (diluted in BHI-YH) on 1.5% BHI-YH agar for colony-forming units (CFU) per ml counts. *R. gnavus* subculture was grown for a period and then spotted on to the agar plate as follows: **a)** t=4 hours, **b)** t=5 hours, **c)** t=6 hours, **d)** t=7 hours, **e)** t=8 hours. This experiment was done once alongside Xena Dyball.

Formation of *R. gnavus* ATCC 29149 colonies was observed for all timepoints at varying dilution factors (Figure 22). After 4 hours of growth (Figure 22a) colonies began to form at a 10^{-3} dilution factor, and more were visible at 10^{-4} though not as greatly. After 5 hours (Figure 22b-d) colonies consistently formed at a 10^{-4} dilution factor. The colonies at the lowest dilution factor for each timepoint were counted for each spot and averaged to calculate the average number of colony-forming units (CFU) per ml (Figure 25).

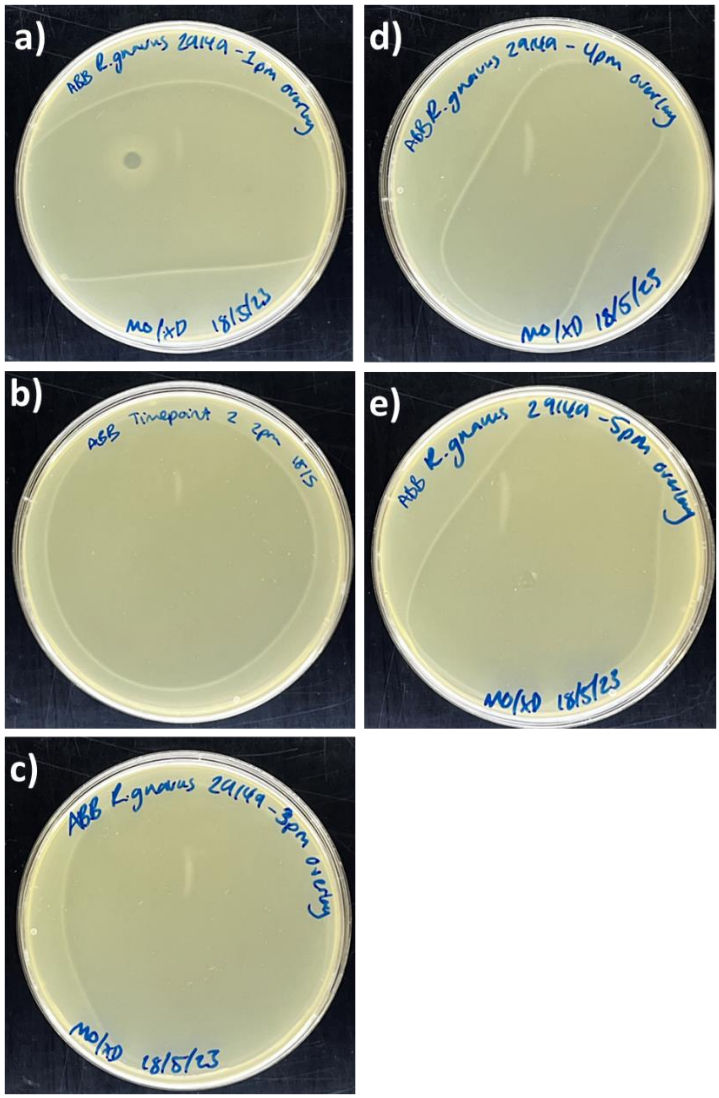


Figure 23 – Agar overlay lawns of *R. gnavus* ATCC 29149 with 0.5% ABB. Lawns were poured out following subculture growth for a period as follows: a) t=4 hours, b) t=5 hours, c) t=6 hours, d) t=7 hours, e) t=8 hours. This experiment was done once alongside Xena Dyball. An air bubble was visible in the overlay at t = 4 hours and should be disregarded.

Lawns of *R. gnavus* ATCC 29149 in ABB appeared to form at all timepoints (Figure 23). The appearance of a haloed section of lower opacity than the surrounding overlay was consistent across all timepoints and distinguished between the formation of an opaque lawn.

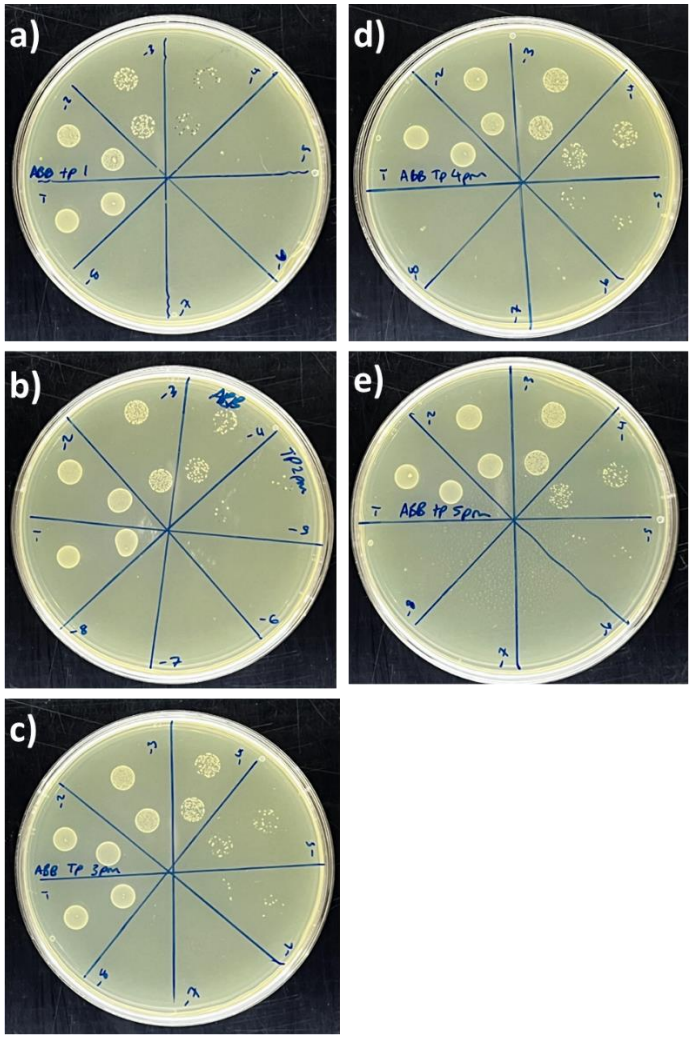


Figure 24 – Dilution series of R. gnavus ATCC 29149 (diluted in ABB) on 1.5% BHI-YH agar for colony-forming units (CFU) per ml counts. R. gnavus subculture was grown for a period and then spotted on to the agar plate as follows: a) t=4 hours, b) t=5 hours, c) t=6 hours, d) t=7 hours, e) t=8 hours. This experiment was done once alongside Xena Dyball.

Colony formation of *R. gnavus* ATCC 29149 was seen at all timepoints (Figure 24). At 1pM (Figure 24a) individual colonies began forming at 10^{-3} dilution factor and were able to be counted at 10^{-4}. After 5 hours of subculture incubation (Figure 24b-d) colonies began consistently forming at 10^{-4} dilution factor and stopped at 10^{-5}. After 6 hours (Figure 24c) there were multiple colonies observed as low as 10^{-6}. Colonies at the lowest dilution factor of each timepoint were counted and averaged between the two spots to calculate the average number of colony-forming units (CFU) per ml (Figure 25).

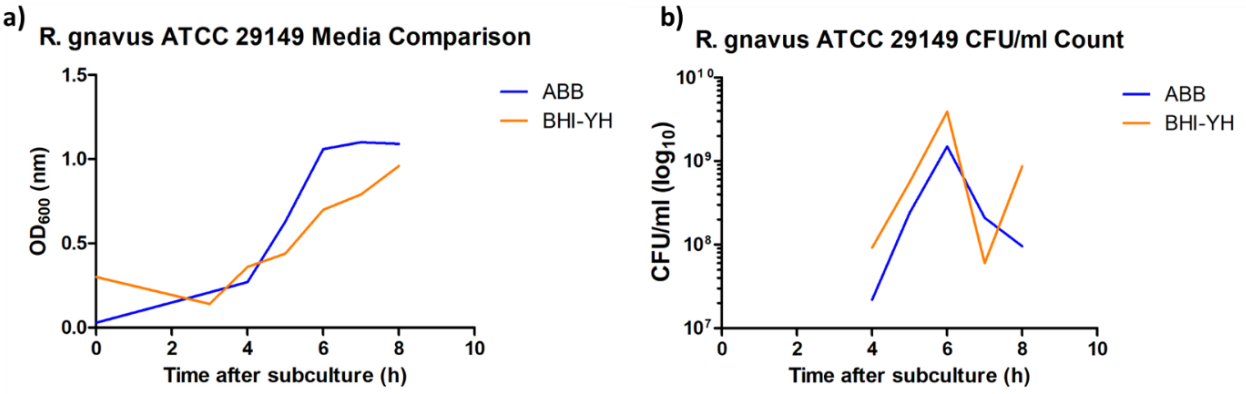


Figure 25 – Comparison of ABB and BHI-YH media for growing *R. gnavus* ATCC 29149 at time following subculture. **A)** OD_{600} comparison of both media. **b)** CFU count comparison of both media. Work was done alongside Xena Dyball and completed once.

OD_{600} was observed for *R. gnavus* ATCC 29149 in both media following subculture incubation for 8 hours (Figure 25a). At t=0, BHI-YH recorded a higher OD at approximately 0.4 compared to ABB at just over 0. For 3 hours following incubation, BHI-YH experienced a decrease in OD to approximately 0.2 while ABB had a sharp increase to around 0.3. From t=3 until t=8, BHI-YH had a steady increase in OD and ended at just under 1. At the same time, ABB experienced a steep increase particularly from t=4 to t=6 in which OD increased from approximately 0.4 to over 1 and then began to plateau before finishing at just over 1. When comparing CFU counts (Figure 25b) both media saw individual colonies began to form at only t=4. At this time, BHI-YH had approximately 10^8 CFU per ml (log_{10} scale) compared to over 10^7 CFU per ml for ABB. CFU for both media experienced a steep increase until t=6, with BHI-YH over $10^{9.}$ CFU per ml compared to approximately 10^9 CFU for ABB. In the next hour after this both media sharply decreased in CFU; at t=7, ABB had decreased to approximately 10^8 CFU per ml while BHI-YH observed around just under 10^8 CFU per ml. Between t=7 and t=8, ABB decreased again slightly to 10^8 CFU per ml while BHI-YH increased to approximately 10^9 CFU per ml.

It was determined that ABB would be the ideal media for *R. gnavus* growth, considering the higher OD between 4-8 hours (Figure 25a), comparable CFU count (Figure 25b) and lawn formation (Figure 23 and 24). In addition, the failure to yield a lawn using BHI-YH and the successful use by Buttimer and colleagues furthered the decision. It was also determined that the ideal growth period for the subculture would be between 6-8 hours of incubation, as opposed to the overnight growth period previously used with BHI-YH.

After achieving an optimal media and growth period, determining the ideal volume of *R. gnavus* to be used for lawn formation was required. Mitomycin C (MMC) was used to observe clearing on lawns of different volumes of *R. gnavus* as a precursor to attempting phage isolation, to observe the lawn condition using a known bactericidal. Alongside Xena Dyball, agar overlays of 200, 300, 400 and 500 µl of *R. gnavus* ATCC 29149 were plated on pre-reduced 1.5% ABB agar plates. After solidifying, 10 µl spots of 1, 4 and 32 µg/ml of MMC and ABB (as a negative control) were spotted onto each overlay. Plates were then incubated at 37 °C in strict anaerobic conditions overnight.

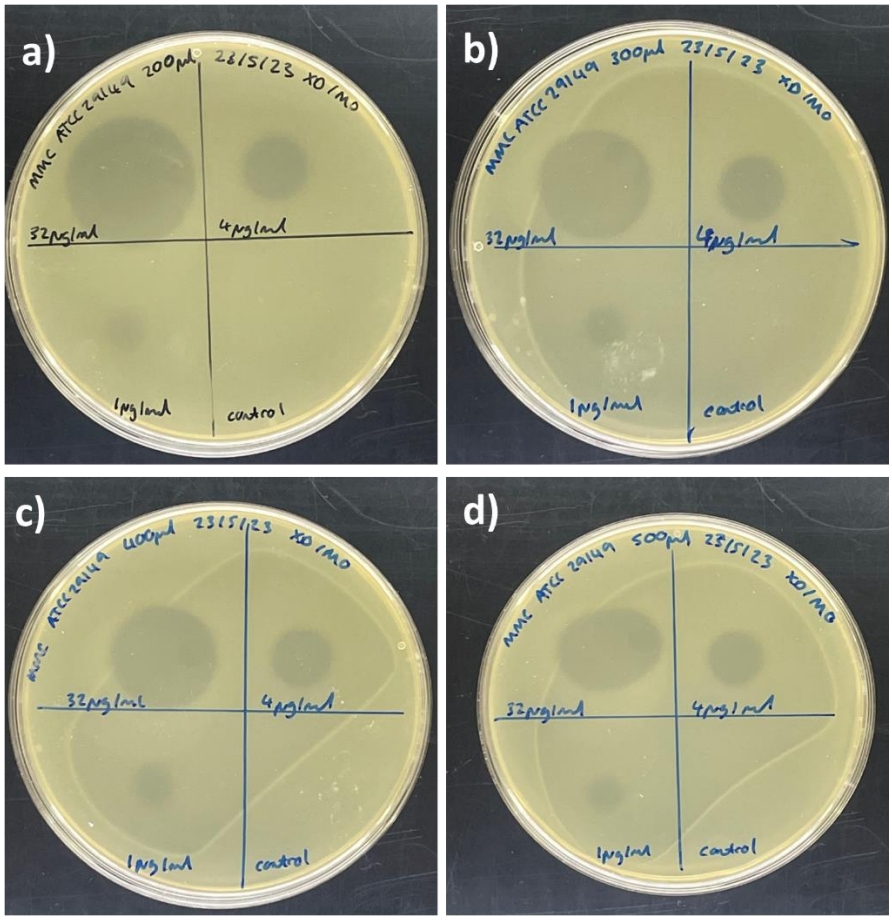


Figure 26 – Spots of mitomycin C on lawns of *R. gnavus* ATCC 29149 in 0.5% ABB. Mitomycin C concentrations 32, 4 and 1 µg/ml were spotted on to the plates alongside ABB as a control. Volumes of *R. gnavus* ATCC 29149 differed as follows: **a)** 200 µl, **b)** 300 µl, **c)** 400 µl and **d)** 500 µl.

Clearing was observed for all mitomycin C concentrations on all different volumes of *R. gnavus* ATCC 29149 (Figure 26). Across all plates, 32 µg/ml produced the largest zone of clearing followed by 4 and 1 µg/ml respectively. Clearing was most transparent on 200 µl of *R. gnavus* (Figure 26a) and became opaquer as the volume increased, with 500 µl showing the opaquest zones of clearing (Figure 26d). After observing the clearing zones on the different lawns, it was decided that 500 µl was the volume to be used to maximise the number of bacteria, but alternative volumes could be considered in different circumstances.

4.6.2 Development of *R. gnavus* phage isolation protocol

After establishing an optimal bacterial lawn, isolation of phages could proceed. *R. gnavus* strains ATCC 29149, E1, Finegold, CC55_001C and ATCC 35913 were used as target hosts. 46 faecal samples taken from a study cohort of elderly patients with varying risks of mental decline and pre-processed by Dr. Oliver Charity were pooled together and used for testing. In addition, wastewater and sewage samples collected by collaborators at Bangor University originating from Llangefni, Treboror, Wrexham, Cardiff, Flint, Newport and Ganol were also used. Due to low volumes of wastewater/sewage samples, multiple samples from different origins but the same timepoint were pooled together. Samples were centrifuged and filtered, as previously described for isolating *Enterobacter* phages, before being enriched as follows: 400 µl of faecal samples were added to 100 µl of 2% *R. gnavus* subculture of each strain. This was repeated for wastewater and sewage samples, instead using 1 ml of sample with 500

µl of *R. gnavus.* Contents were then transferred to 10 ml of anaerobic ABB and incubated overnight at 37 °C in strict anaerobic conditions. Enriched samples were centrifuged and filtered, producing a sample enriched by each strain of *R. gnavus* used. These enrichments were then spotted as previously described for the isolation of *Enterobacter* phages, on a lawn for each strain of *R. gnavus*. Enrichments were then spotted on to the plate before being incubated at 37 °C overnight in strict anaerobic conditions.

Table 21 – Samples used for testing for isolating phages against *R. gnavus* ATCC 29149, E1, Finegold, CC55_001C and ATCC 35913. *n* denotes the number of samples pooled together. *N* = no, *Y* = yes. Where clearing/single plaque formation was observed, the strain on which it was observed is noted. Samples were pooled together to maximise efficiency.

Sample used	**Sample description/origin**	**Clearing observed**	**Single plaques formed?**
MOTION Study Cohort	Faecal samples from elderly patients at varying risks of mental decline (n=46)	N	N
Wastewater	Wastewater samples originating from Ganol, Newport (n=4)	Y – *R. gnavus* Finegold	N
"Kata" Sewage	Sewage samples originating from Llangefni, Treboror, Wrexham, Cardiff, and Flint (n=13)	Y – *R. gnavus* ATCC 29149, E1, Finegold, CC55_001C	Y – *R. gnavus* ATCC 29149, E1, Finegold, CC55_001C

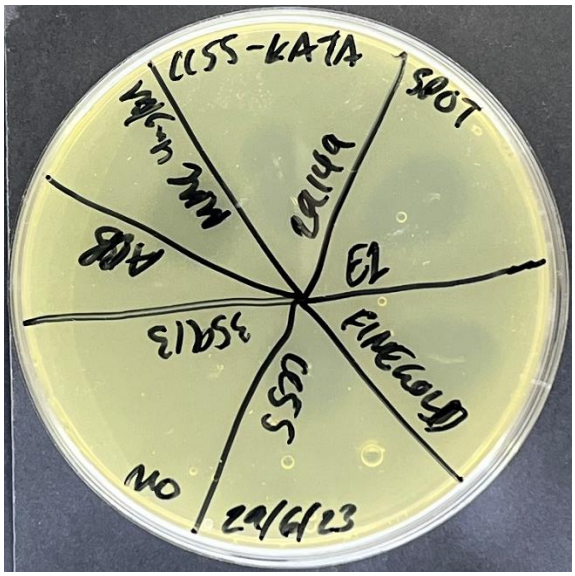


Figure 27 – Spot test of "Kata" sewage samples originating from Llangefni, Treboror, Wrexham, Cardiff, and Flint on a lawn of *R. gnavus* CC55_001C. 32 µg/ml of mitomycin C was spotted as a positive control alongside ABB for a negative control.

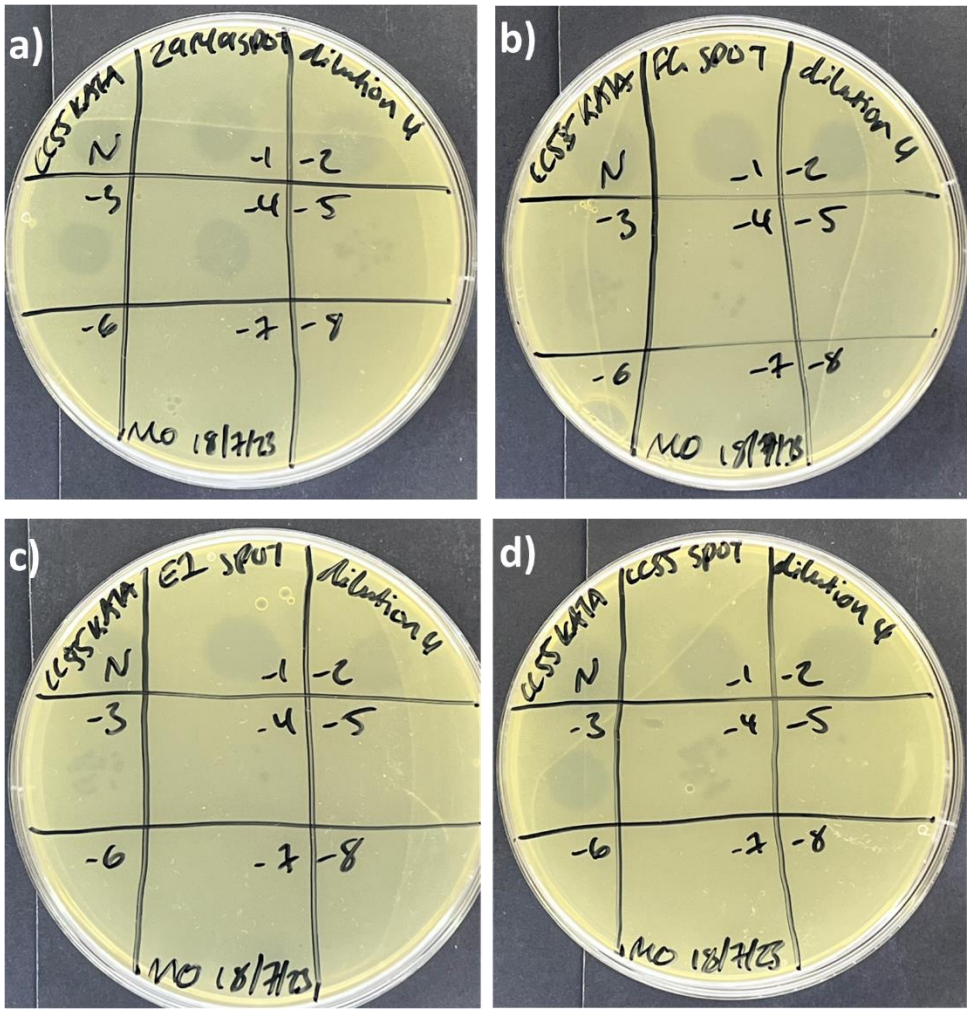


Figure 28 – Single plaque formations of "Kata" sewage samples for **a)** *R. gnavus* ATCC 29149, **b)** *R. gnavus* Finegold, **c)** *R. gnavus* E1 and **d)** *R. gnavus* CC55_001C.

Of the 63 samples tested, both the 4 wastewater samples and the 13 sewage samples produced clearance (Table 21). Wastewater samples produced clearance only on *R. gnavus* Finegold but no single plaques were formed. The Kata sewage samples enriched with *R. gnavus* ATCC 29149, E1, Finegold and CC55_001C produced clearance on a lawn of *R. gnavus* CC55_001C and produced single plaques (Figure 27). Plaque morphology across all plates was relatively similar - slightly opaque small plaques were formed. Following several rounds of single plaque purification as previously described for isolation of *Enterobacter* phages, plaques were amplified into stocks for storage and downstream analysis.

Table 22 - Phages isolated against *R. gnavus* CC55_001C, describing enrichment origin, plaque morphology, and phage titre. All phages were isolated from Kata sewage samples.

Phage name	**Strain enrichment**	**Plaque morphology**	**Phage titre (PFU/ml)**
MOR1	*R. gnavus* ATCC 29149	Slightly opaque, small plaques	1.3×10^{8}
MOR2	*R. gnavus* E1	Slightly opaque, small plaques	2.5×10^{6}
MOR3	*R. gnavus* Finegold	Slightly opaque, small plaques	7.0×10^{6}
MOR4	*R. gnavus* CC55_001C	Slightly opaque, small plaques	$^{1.7} \times 10^{7}$

5. Future work

With more time, more work could be completed to further the research project. The observation of the isolated *Enterobacter* phages using transmission electron microscopy (TEM) would be ideal to observe their morphologies and sizes. As a facultative anaerobe, *E. cloacae* is also capable of surviving in little to no oxygen conditions, so testing host-range and phage infection under anaerobic conditions could be insightful to see if there are any differences compared to aerobic conditions. In addition, testing of additional hosts could also be completed, particularly on other additional *E. cloacae* strains and *Enterobacter* bacteria due to their genetic similarities and prevalence in the gut microbiota, e.g., *Escherichia coli, Klebsiella pneumoniae,* etc., using both laboratory and clinical strains to determine if the phages were able to successfully infect these and how infection could differ compared to the host. The infection cycle of the phage could also be analysed using a one-step growth curve to learn about the growth of the phages and how this may differ not only between each other but in hosts and would be valuable data to transferring to a clinical setting. Looking at the assembly of the phages, MO1 and MO2 can only be considered draft genomes as there remains other contigs that could make up parts of the entire phage genome. Being able to determine genome completeness by investigating other assembled contigs and determine if they are part of the genome or not would allow for full genome assembly and completion. This could reveal more information about the phage genomes, including other predicted coding sequences and genes. Furthermore, being able to complete a comparative genomics analysis of each phage and their associated relatives would reveal more information about the genomes of each phage and what is shared with their relatives.

Due to time restrictions, the work done in isolating phages for *R. gnavus* concluded after phages had been isolated and amplified into phage stock for storage. If work was able to be continued on the project, there are multiple ways it could be furthered. Using TEM for observation of phage morphologies would be ideal. One-step growth curves of all phages would provide useful information about infection cycle dynamics and how this could differ between phages. Host-range analysis of the isolated phages could be carried out to determine if they are able to infect any additional hosts beyond the initial host strains. *R. gnavus* ATCC 35913 did not produce any clearing or plaques when tested alongside the other strains but could be tested again using enrichments of each phage to see if there would be any change to this. Clinical isolates of the strains of which phages have already been isolated could also be an ideal host and would assist in validating the use of these phages in potential phage therapy. In addition, the *Ruminococcus gnavus* has recently been reclassified as belonging to the *Mediterranibacter* genus (Togo, et. al., 2018). This includes *Ruminococcus faecis, Ruminococcus lactaris, Ruminococcus torques, Clostridium glycyrrhizinilyticus* – as such, these could also be viable hosts for the isolated *R. gnavus* phages. Alongside the host-range analysis, an efficiency of plaquing assay could also be done to compare how well the phages plaque on other hosts compared to the original hosts they were isolated from/

DNA extraction and quantification of the isolated phages could also be completed to find out the DNA concentration of each phage. This would then be used for whole-genome sequencing of the phages, with both Illumina and Nanopore sequencing to be done. Assembly of the genomes could then be completed using the HYPPA workflow as previously described, as this method is the most comprehensive for assembling phage genomes drawing on both types of sequencing data and multiple polishing steps. Following successful genome assembly, genome annotation could be undertaken using Pharokka as previously described. Genes of interest could then be investigated, particularly those relating to temperate phages, e.g., an integrase. The phages isolated by Buttimer, et. al., were all found to be temperate, so the discovery of a virulent phage targeting *R. gnavus*would make it the first to be isolated and published. Comparative genomics would then be undertaken; first of the genomes of the isolated phages between each other to identify any similarities or differences between them and to what extent. Secondly, a comparison to Buttimer, et. al.'s isolated phages would highlight any significant similarities between the two groups.

6. Discussion

Bacteriophages represent a vital component of the gut microbiome and other ecological niches, shaping microbial communities through predation of bacteria. Their role as mediators of microbial composition has been extensively studied, especially involving the gut microbiota and pathogenic and opportunistic bacteria. Research has displayed the potential of bacteriophage therapy in treating bacterial infections, biocontrol of food and combatting antibiotic resistance while also shedding a greater light on these viruses. However, much remains to be fully understood about bacteriophages including phage dynamics, bacteriophages of lesser studied bacteria and microbes, e.g., anaerobes, and the bacteriophage genome (including, perhaps, a "core" genome shared amongst all phages). The research presented aims to assist with expanding knowledge of bacteriophages for gut-associated bacteria *Enterobacter cloacae* NC10005 (*E. cloacae*) and *Ruminococcus gnavus* (*R. gnavus*). Both bacteria make up the "typical" gut flora of humans in different contexts. *E. cloacae* is an extensively drug resistant and opportunistic bacteria, able to resist an array of antibiotics and drugs when treated. *R. gnavus* is a commensal microbe that does not usually cause infection or harm but has been linked to IBD and inflammation within the gut, and knowledge of bacteriophages for anaerobic bacteria is scarce. Isolation of phages targeting these bacteria can not only help with increasing knowledge of bacteriophages but assembling a collection of phages to potentially be used in phage therapy. The aim of the project was as follows: to discover novel bacteriophages targeting *E. cloacae* and *R. gnavus* and investigate the use of these phages in treating bacterial infections and modulating the gut microbiota. In pursuit of this, three novel bacteriophages targeting *E. cloacae* were successfully isolated, amplified and had their genomes sequenced and annotated: MO1, MO2 and MO3. MO1 has been tentatively classified as an unclassified *Karamvirus* species and phage MO2 and MO3 have been tentatively classified as unclassified *Henuseptimarivus* species based on genomic and proteomic comparisons with relatives in public databases. Alongside this, four bacteriophages targeting *R. gnavus* CC55_001C were isolated and amplified: MOR1, MOR2, MOR3 and MOR4.

In attempting to isolate bacteriophages for *E. cloacae*, 12 samples were used: 8 wastewater samples and 4 river samples. *E. cloacae* is a prevalent member of the gut microbiota and tends to be present in human faecal matter and accompanying areas where this or contact with these faeces may be, e.g., wastewater, sewage, and bodies of water where people may frequent or run-off from sewage/wastewater may be present. As such, it is thought that bacteriophages are also present and samples from these areas are the primary interest when trying to isolate bacteriophages for bacteria

of the gut. Some samples have different timepoints of collection at the same origin as the composition may differ depending on when the sample is collected, so samples from the same place at different times could yield different results. Three bacteriophages targeting *E. cloacae* were found from samples originating from Chester Wastewater Treatment Works, Wrexham Five Fords WwTw and Llangefni Sewage Works. They were named MO1, MO2 and MO3, respectively once it was confirmed they were bacteriophages, each with slightly differing plaque morphology. MO1 formed small, transparent plaques; MO2 large, transparent plaques with halos (turbid outer rings), and MO3 with large plaques that then began to present a halo, unlike in the initial clearing observed (Figure 6). Plaques possessing a halo can be for a variety of reasons. Halos are typically associated with decreasing lytic activity due to aging of the bacterial lawn as it dries, but can also be attributed to lysis inhibition, a phenomenon occurring because of intracellular viral communication (Jurczak-Kurek, et. al., 2016; Abedon, 2009). When infecting susceptible host cells, there exists a delicate balance between the number of free phages and the number of unlysed host cells. Initial infection of host cells begins with a single free phage that binds to the host and ends with the cell being lysed and phage progeny being released. After some time, the ratio between free phage and unlysed host cells swings in favour of the phage, typically because of "superinfection", in which multiple phages or phage particles infect a single bacterium. Phages can then trigger lysis inhibition, delaying lysis for several hours (Dressman and Drake, 1999). It's believed that this strategy exists because the population of the host has reduced to a point that phage progeny may not have any hosts left to infect, acting as a method of preservation for both the phage and the bacteria (Erez, et. al., 2017). In addition, it's also thought that the appearance of a halo could result from the diffusion of depolymerases - phage enzymes capable of degrading bacterial polysaccharides - through the bacterial lawn and can serve as a marker for phages that encode the genes for them (Glonti, et. al., 2010; Cornelissen, et. al., 2011). Phage-derived polymerases are of particular interest due to their potential application in treating biofilms and associated diseases, where breakdown of bacterial polysaccharides is useful (Topka-Bielecka, et. al., 2021). Other studies of isolated phages against *E. cloacae* strains and the *E. cloacae* complex tend to present similar plaque morphologies to MO1, small, clear plaques with no halo (Morozova, et. al., 2021; Cieślik, et. al., 2022; Nasr-Eldin, et. al., 2023). Nasr-Elden and colleagues also managed to isolate an *E. cloacae* complex phage with a large plaque size and halos more consistent with MO2 and MO3. These are consistent with the plaque morphologies observed for the isolated *E. cloacae* phages.

For the host-range of the isolated *Enterobacter* phages, *E. ludwigii* strains HVP30 and HVP31 were chosen as *E. ludwigii* is included within the *E. cloacae* complex and would demonstrate the ability of the phages to infect an additional host within the complex. Initial results revealed that MO1, MO2 and MO3 could all successfully infect (form single plaques) against HVP30 (Figure 7). Plaque morphologies were consistent with those against *E. cloacae* NC10005. When infecting *E. ludwigii* HVP31, MO1 was unsuccessful in forming single plaques while MO2 and MO3 did. Clearing of the bacterial lawn indicates bacterial lysis has occurred at higher concentrations of phage, but the inability to form plaques could suggest a resistance to MO1 that is only overcome by higher titres. Unlike with *E. cloacae* NC10005, no enrichments of *E. ludwigii* strains with phages were done prior to testing for host-range. Thus, a repeat with an enrichment step could allow enough incubation time for MO1 to successfully bind and lyse *E. ludwigii* HVP31, possibly overcoming the resistance when repeated. Plaques formed by MO2 and MO3 were more consistent with those formed against *E. cloacae* NC10005 – large, transparent and accompanying a halo. The host-range was later repeated by Dr. Hannah Pye, who tested against *E. cloacae* NC10005 and *E. ludwigii* HVP30 and HVP31 again but also *E. cloacae* strains PO27E, PO69D, PO95LP, clinical E. cloacae isolate GOSH2 and *E. coli* strain ES2833 (Figure 9). When infecting *E. cloacae* NC10005, MO1 now formed single plaques with a morphology like MO2 and MO3 (as opposed to initially being distinct; Figure 7) - large with a halo. In addition, MO1

now formed single plaques against both *E. ludwigii* strains where it previously failed to do so (Figure 8), producing plaques smaller than those formed against *E. cloacae* NC10005 but larger than the plaques observed for MO1 against HVP30. MO2 and MO3 formed plaques of the same morphology against HVP30 and HVP31 when tested again. It should be noted that Dr. Pye also did not perform an enrichment step for any of the bacterial strains tested. The change in plaque morphology for MO1 against *E. cloacae* NC10005 and the new formation of plaques against *E. ludwigii* HVP31 suggest a change to the phage plaquing ability between the initial isolation/host-range test and the one carried out by Dr. Pye - a likely cause could be a sudden mutation in MO1 that caused this during its storage period. The efficiency of plating (EOP) was calculated for all phages against *E. ludwigii* HVP30 and HVP31, using *E. cloacae* NC10005 as a baseline (Figure 10). In comparison to *E. cloacae* NC10005, all phages had lower EOP for *E. ludwigii* HVP30 while for *E. ludwigii* HVP31, all EOP values were greater than that of *E. cloacae* NC10005. Greater EOP of all phages against *E. ludwigii* HVP31 suggest it is more susceptible to infection than both *E. cloacae* NC10005 and *E. ludwigii* HVP30.

MO1, MO2 and MO3 were initially sequenced using Illumina and assembled using Shovill (Seeman, 2017) to generate assemblies of 633,770, 50,130 and 170,917 bp into 87, 1 and 7 contigs respectively. Visualisation of assembly graphs using Bandage (Figure 11) shows all the DNA fragments from the assembly, including the assembled contigs, and was necessary to determine which contigs were representative of the genome and which could just be poor assemblies, such as homopolymeric contigs that Illumina is prone to, or contamination, represented by smaller branching lines as seen in the assembly MO1 (Figure 11a2). Ideal phage assembly would display a single circular contig, as was the case for MO2 (Figure 11b). Though displayed as circular, dsDNA tailed phage genomes are not truly circular when packaged in the capsid. Phage genomes can be circularly permuted to give the appearance of a circular genome but this is not truly representative, and circularization also tends to be an artefact of assembly (most likely due to biases for bacterial genomes). Within the context of the Bandage image the formation of a circular contig is ideal for identification of the phage genome. MO2 and MO3 had better assemblies compared to MO1, with just one and two circularly assembled contigs respectively, and lack of branches that would indicate poor assembly. MO1 has several of the latter and additional assembled circular contigs (complete genome assembly would involve the resolution of these other contigs, however due to time constraints the first and largest contig was taken as representing most of the phage genome) and as such, MO1 was a draft genome requiring further curation before it could be considered fully assembled and complete. The genome of MO2 was taken as the single assembled circular contig. With MO3, two circular contigs had assembled, suggesting the presence of two phages in the sample. The first contig (Figure 11c1) was taken as representing the phage genome of MO3, and the second contig (Figure 11c2) was further investigated and will be discussed later. While genome assembled was being completed, a hybrid and poly-polish phage assembly workflow for the assembly of complete phage genomes was published by Elek, et. al (Elek, et. al., 2023). This method was used to reattempt genome assembly of MO1, MO2 and MO3 and required long read sequencing with Oxford Nanopore Technology.

Following this methodology, phages were able to be assembled following an iteration of assembly using Flye and Canu from long-read data. Where MO1, MO2 and MO3 had initially assembled into 87, 1 and 7 contigs respectively with Illumina, long-read assembly had assembled phages into 1, 9 and 1 contigs respectively. These contigs of MO1, MO2 and MO3 were, in total, 50,013, 57,045 and 58,696 bp respectively (Tables 5 - 7). Iterations of both long-read and short-read polishing followed (Table 8 – 10) to refine this further but the number of contigs and total length remained the same for all phages and each were given a consensus quality score barring MO1 that could not undergo the second iteration of short-read polishing (Table 10). MO2 and MO3 were scored as having consensus >99% quality, indicating a high degree of confidence in the assemblies. A comparison between both

assemblies was done to determine which was ideal for taking the phage genomes forward (Table 11). Discrepancies between both assemblies was observed; Shovill had assembled MO1 and MO3 into a greater number of contigs while MO2 had assembled into more contigs with the HYPPA workflow. The genome length of MO1 was over three times as large from the Shovill assembly compared to HYPPA, whereas MO2 and MO3 were within 7000 and 12000 bp respectively of each other with HYPPA generating smaller genomes. MO1 had the biggest difference between both assemblies in both genome length and contigs assembled and when doing a BLASTN comparison of both assemblies (Table 12), this difference was further observed. MO1 reported no significant similarities in nucleotide sequence similarity between either assembly, indicating both were distinct from each other. MO2 and MO3's respective contigs were found to be >88% query cover and 100% identity, aside from two contigs in MO2 assembled from HYPPA. Due to time constraints, the largest contig of MO2 assembled via HYPPA was taken as the phage genome and it being a draft genome. Relatives of assembled HYPPA phage genomes revealed the clustering of all three phages with several *Enterobacter, Escherichia* and *Klebsiella* phage relatives at ~>80% intergenomic similarity (Table 13, Figure 12). Several of these relatives had comparable genome sizes (Table 13) to MO1, MO2 and MO3 (which was to be expected as members of the *Enterobacteriaceae* family). MO1 and MO3 displayed >99% intergenomic similarity between each other, strongly suggesting they may be identical. When deciding which assembly to take forward for further analysis for each phage, the differences in workflow between each method were strongly considered. Most phage genome assemblies rely on either short-read or long-read data, each with their own advantages and disadvantages and rarely are both used in tandem with one another. Hybrid assemblies are generally considered to be most advantageous for especially complex phage communities, due to being able to study the genomic diversity of an abundance of viruses in ecological niches while accounting for errors that each individual assembly may bring forth (Shen and Millard, 2021; Zablocki, et. al., 2021). When considering individual phage isolates, the considerable costs hybrid assemblies can incur make it not recommended for most phage genome sequencing projects. In the context of HYPPA, the use of both short-read and long-read data together, being able to cover pitfalls that may occur with just one method, coupled with the multiple iterations of polishing, I believe to be the ideal standard for bacteriophage genome assembly that offers a more robust method. When deciding to take the HYPPA assembly forward for MO2, this was done with the intention of resolving the genome that had assembled into multiple contigs as opposed to the one with Shovill. Though Shovil had one complete genome assembled, the 100% QC and identity of the eight of nine HYPPA contigs to the one contig (Table 12) suggested that all could be parts of the same genome that could be resolved. This was taken forward with the intention of being able to fully assemble the genome of MO2 from the different contigs assembled from HYPPA but time constraints meant that this could not be done and only the largest contig could be taken for further analysis. Regarding MO3, the second contig assembled from Shovill (Figure 11c2) was already determined to belong to a prophage (discussed later) that had induced from the host during the isolation process before proceeding with the HYPPA assembly, so its lack of appearance in the HYPPA assembly was preferred for attempting to generate a single, lytic, phage genome for MO3. Inclusion of the second contig in attempting to resolve a single genome would've resulted in a mismatched contig that would have been a combination of *E. cloacae* NC10005's genome and MO3's, rather than representing the latter. Though assembly with HYPPA was not able to generate a result to be taken forward for MO1 the hybrid poly-polish approach should be considered ideal for curating bacteriophage genomes, as previously stated. Currently, this methodology was only validated with three of the ten przondoviruses reported, and further usage of this workflow with additional, diverse isolated phages will highlight the reliability of this method.

Following the comparison between both assemblies (Table 14) and the hybrid and poly-polish approach of HYPPA, it was determined that HYPPA would serve as the complete method of genome assembly, and the assembled genomes from this taken forward, except for MO1. MO1 had the greatest discrepancy between both assemblies and no significant similarities were found between each other. While the HYPPA assembled genome would be ideal, the near identical intergenomic similarity to MO3 (Figure 12) suggests a possible contamination of MO1 or incorrect assembly. Instead, MO1 assembled with Shovill was taken forward as representing its phage genome (as a draft) and MO2 and MO3 assembled with HYPPA were taken forward as representing their phage genomes.

Phage genome annotation was initially performed using Prokka (Seeman, 2014). This method predicted most of each phage genome as being made up of hypothetical proteins, failing to provide much meaningful information about their genomics (Table 15). Annotation with Prokka resulted in >93% of MO1, MO2 and MO3's genomes being predicted as hypothetical or unknown. As a prokaryotic genome annotation tool, Prokka is made up of dependencies and other tools that favour bacteria and archaea and are not optimised for bacteriophages. The structural annotation performed by Prodigal (Hyatt, 2010), was successful in identifying CDS, but even this could be incorrect as bacteria tend to have larger open reading frames (ORFs) compared to viruses, so smaller or whole CDS could be missed. The functional annotation relies on the databases as follows: (1) ISFinder (Siguier, et. al., 2006); (2) the NCBI Bacterial Antimicrobial Resistance Reference Gene Database (NCBI Resource Coordinators, 2016); and (3) UniProtKB (SwissProt) (The UniProt Consortium, 2021). All three databases heavily bias bacteria and are heavily constructed around bacterial data (as well as archaeal), resulting in the tool being unable to resolve many most of the functions. To resolve this, Pharokka (Bouras, et. al., 2023) was used as an alternative genome annotation tool. For structural annotation, PHANOTATE (McNair, et. al., 2019) is used and is currently the only gene prediction tool specialised for bacteriophages. When completing functional annotation, Pharokka relies on the PHROG database, a collection of >930,000 proteins built on >17,000 reference viruses of prokaryotes (Terzian, et. al., 2021). In addition, the Comprehensive Antibiotic Resistance Database (CARD; Alcock, et. al., 2020) and the Virulence Factor Database (VFDB; Chen, et. al., 2005) are also used for functional annotation. As a trio of databases compared to those used by Prokka, these are tailored more specifically to bacteriophages, and can better predict and annotate CDS (Table 15). Following annotation with Pharokka, all phages had a massive decrease in the percentage of their genomes predicted to be hypothetical, going from >93% with Prokka to <74% with Pharokka. While MO1 and MO2 had around half of their genomes still predicted to be hypothetical, MO3 had 72.3% as a majority. Explicit functions and the number of CDS for each function were predicted (Table 16, Figure 13). Despite the differences in genome size, all phages were predicted to have the same number of connector genes. Of all phages, MO1 was the only one to contain CDS with two functions that neither MO2 nor MO3 possessed. MO1 was predicted to contain genes that were considered morons, auxiliary metabolic genes and/or host takeover. Moron genes are highly diverse genes that are beneficial for the host bacterial cell or the phage but not required for the phage's life cycle (and named as such because they add "more on" the phage genome) (Tsao, et. al., 2018.). These genes can be multifaceted; expressing proteins that protect their host via superinfection exclusion or adapting the host to a unique environmental niche. They can also favour the phage, decreasing motility of the host or expressing anti-CRISPR proteins that inactivate the CRISPR-Cas bacterial immune system (Pawluk, et. al., 2014). Their role within the phage and the host varies, and more remains to be discovered about them. One predicted moron gene in MO1 is an NrdD-like anaerobic ribonucleotide reductase large subunit – ribonucleotide reductases are enzymes responsible for converting ribonucleotides to 2'-deoxyribonucleotides, the precursor for DNA synthesis and repair in virtually all living organisms. Its presence in MO1 could be attributed to

assisting its host with DNA synthesis and repair, as these processes are not only essential for ensuring host survival, but phages make use of host DNA when assembling progeny and issues in host DNA would directly impact their own ability to replicate. A beta-glucosyltransferase is also present, and found in Bacteriophage T4, as an enzyme that transfers glucose from uridine diphosphoglucose to a modified DNA using a beta-glycosidic bond in order protecting viral DNA from any restriction enzymes in the host that may cut it up. These moron genes, amongst others, also demonstrate how much remains to be fully understood about bacteriophage genomes. fIn addition to moron genes, MO1 also was predicted to contain tRNAs, 19 exact. The role of tRNAs within the phage genome has been ambiguous – because phages make use of host translational machinery during replication, including tRNAs, they do not necessarily need any of their own. Yet, tRNAs were first discovered to be carried by bacteriophage T4 and it was initially hypothesized that their function was to bias translation towards their own genes, as opposed to their hosts, and supplement codons that are less frequently used (Cowe and Sharp, 1991). It has also been recently suggested that phage tRNAs are able to evade tRNA-targeting hose defences via evolution to be insensitive to host anticodon nucleases (Berg, et. al., 2023).

No phages were predicted to contain any genes for integration & excision, virulence factors or antimicrobial resistance. In the context of research these are not especially important, however when considering the use of these phages for modulating the gut microbiota, the latter two are. Phages themselves encode various genes to assist with infection and ensure host survival, including those for toxicity and virulence (Abedon and LeJeune, 2007). General or specialised transduction can occur whereby genetic material from one bacterium is transferred to another by viruses (Feiner, et. al., 2015; Goh, 2016; Penadés, et. al., 2015). General transduction involves the mispackaging of bacterial DNA into capsid heads during the lytic cycle, creating phages with random pieces of the host bacterial genome in addition to the phage genome – a new bacterium then receives foreign bacterial DNA that may then integrate into the host genome via homologous recombination. Specialised transduction occurs during improper prophage excision; when exiting the host genome, the prophage can take with it a restricted set of genes exclusive to the host that are then packaged into phage progeny. These genes may then get inserted into the new host's chromosomes and grant them new characteristics. In this sense, phages can contribute to the spread of antimicrobial resistance by conferring bacteria with genes through horizontal gene transfer (Balcazar, 2014). Shiga-toxin producing *E. coli* (STEC) are one instance thought to occur due to horizontal gene transfer – the genes for the Shiga toxin have been demonstrated to be encoded by phages and are able to be passed on to *E. coli* strains *in vivo* via lysogenic conversion (Marinus and Poteete, 2013; Majowicz, et. al., 2014). If these phages were to be used clinically, this would be a factor to bear in mind to not worsen the issue of antimicrobial resistance or equip bacteria with virulence factors that could harm an individual.

While attempting to resolve the genome of MO3 assembled from Illumina using Shovill (before deciding to use the HYPPA method), the Bandage image revealed the assembly of two circular, separate contigs (Figure 11c) that could suggest the presence of two phages within the sample. The second assembled contig (Figure 11c2) was referred to as MO3.2 and annotated with Pharokka to discover if it was a phage or not (Figure 14). Genome annotation revealed many of the typical predicted phage genes, e.g., tail fiber proteins, endolysin, holin, exonuclease, etc. However, the presence of an integrase gene strongly suggested the possibility of MO3.2 being a prophage that had induced during the isolation process, as an integrase is usually present in temperate phages for integration into the host genome. PHASTER was used to determine if MO3.2 was a prophage or not and received a maximum score of 150 (Figure 15), indicating a very strong likelihood of MO3.2 being a prophage. To investigate the origin of the prophage, whether it had induced from the host *E. cloacae* NC10005 or could be a possible contaminant, the host and MO3.2 were run through BLASTN using *E.*

cloacae as the subject and MO3.2 as the query sequence (Table 17). Of the 45.4 kb length of MO3.2, 41.5 kb (91.4%) were found to be 100% identical to *E. cloacae* NC10005, confirming that the region had originated from the host as a prophage. This result also revealed that *E. cloacae* NC10005 was host to a prophage that we were unaware of and, fortunately, had only been induced during the isolation process of MO3. Like the previously mentioned note of virulence factors and antimicrobial resistance genes, this is not a major cause for concern within the context of research. However, as these phages are intended to be used for modulation of the gut microbiota and phage therapy, the presence of a prophage would need to be heavily considered. Strictly lytic phages are ideal for phage therapy because they can kill their hosts immediately, while temperate phages carry the risk of integrating into the host and laying dormant until certain conditions force the induction out of the host genome. Temperate phages also have the possibility of incorrectly excising from the host genome and potentially taking with them genes for antimicrobial resistance, which could be passed on to the next bacterium and subsequent future generations via horizontal gene transfer. While not ideal for conventional phage therapy, temperate phages have still been considered for use. One avenue is through their ability to carry moron genes that could reduce host fitness and being able to integrate into hosts without killing them. Temperate phages of *Pseudomonas aeruginosa* (*P. aeruginosa*) were shown to be able to reduce of host motility of lysogens that is key for bacterial virulence (Chung, et. al., 2012). This had knock-on effects of reducing bacterial loads and mortality in *Drosophila* species. Temperate phages have also been considered for use in the context of phage engineering to make them more suitable for phage therapy. Zhang, et. al., was able to engineer a temperate *Enterococcus faecalis* phage to make it exclusively lytic by deleting the genomic module responsible for lysogeny, creating a mutant incapable of lysogeny with a significantly extended host range (Zhang, et. al., 2013). The extended host range was an unintended, yet positive, phenotypic change but would need to be clearly identified before use in therapy. Beyond engineering the phage itself, temperate phages have also been explored to be used as delivery systems of synthetic gene networks, exploiting the capacity to integrate into the host genome. Edgar, et. al., was able to do so and modified a temperate lambda phage to reverse its host antibiotic resistance, making it susceptible to the drugs it had evolved resistance to (Edgar, et. al., 2011). Temperate phages do have their use within phage therapy but would need to be selected critically and evaluated before being used, so it is possible that MO3.2 could also be considered for use, in the event it could be separated from MO3 or induced and isolated from *E. cloacae* NC10005. This still raised the need for a distinction between a “production strain” to be used within the lab and a “clinical strain” to be used for treatment. *E. cloacae* NC10005 is currently adequate for testing bacteriophages and attempting isolation from, due to its rapid growth period and easy growth conditions, however, there eventually would be need for a clinical strain of *E. cloacae* (or another host that is susceptible to infection) that does not contain a prophage to not put patients or treated people at risk. Though this was not possible to be explored within the research project, this could be built upon with future work and those isolating phages for phage therapy purposes should bear this in mind.

Comparative genomic analyses were performed for MO1, MO2 and MO3. Relatives of MO1 were discovered (Table 18) and those shown had a high degree of identity, >95%, across >94% of its genome. These relatives, alongside those previously identified for MO2 and MO3 (Table 13) were all compared to identify intergenomic similarities (Figure 16). A distinct clustering between MO1 and its relatives and MO2 & MO3 and their relatives was observed. MO1 shared <0.1% intergenomic similarity between MO2 and MO3 and their associated relatives, suggesting it may fall under a different taxonomic classification compared to MO2 and MO3. This was exemplified with Figure 17 where the same clustering of phages and their respective relatives can be observed relative to other dsDNA viruses. The host group phyla of all phages and relatives were identified to be Pseudomonadota, of

which the *Enterobacteriaceae* family falls under where the host relatives and isolated phages derive from. MO1 and its relatives were predicted to belong to the *Straboviridae* family of viruses, whereas MO2, MO3 and their relatives were instead predicted to be part of the *Drexlerviridae* family instead. The taxonomy of the relatives was further researched using the NCBI Taxonomy Browser (Schoch, et. al., 2020) (Table 19, 20). All relatives of MO1 were classified as belonging to the subfamily *Tevenvirinae*, *Karamvirus* genus and are unclassified *Karamvirus* species except for *Enterobacter* phage pg7. The high intergenomic similarity between MO1 and its relatives along with the clustering and assigned taxonomy led us to, tentatively, classify MO1 in the same vein as its relatives. That is, belonging to *Straboviridae* family*, Tevenvirinae* subfamily*, Karamvirus* genus and is an unclassified *Karamvirus* species of dsDNA virus. The taxonomy of relatives for MO2 and MO3 (Table 20) varied more compared to MO1's relatives. *Salmonella* phage KKP 3830 has currently not been found to be classified with any taxonomy from family to species but is within the class *Caudoviricetes* and, as such, can be said to be an unknown *Caudoviricetes* species for the time being. All other relatives were found to be classified as belonging to the *Drexlerviridae* family and *Tempervirinae* subfamily. Most relatives were found to belong to the *Henuseptimavirus* genus and are unclassified *Henuseptimavirus* species. *Escherichia* phage Henu7 was the only phage that is currently species classified. Phages identified as being unclassified *Henuseptimavirus* species also tended to cluster most closely to MO2 and MO3 (Table 20) with >60% intergenomic similarity (Figure 16). Those with a lower intergenomic similarity were found to belong to the genera *Hanrivervirus* or *Tlsvirus,* the latter of which contained two *Tlsvirus* YSP2 species. The results of these (Figure 16, Table 20) led us to tentatively classify MO2 and MO3 as belonging to the *Drexlerviridae* family, *Tempevirinae* subfamily, *Henuseptimavirus* genus and are unclassified *Henuseptimavirus* species. Thus, we tentatively propose that we have successfully isolated a novel *Karamvirus* dsDNA virus (MO1) and two novel *Henuseptimavirus* species.

Looking at the genome similarities between MO2 and MO3, both phages share a high proportion of their genomes with each other (Figure 18). In the case of MO2, virtually all its genome is shared with MO3, while the latter has a significant portion of its genome that does not align with any of MO2's. As MO2 is a draft genome, and not fully complete, it is possible that with completion there may be more of its genome that aligns with MO3. MO3 was found to possess genes particularly associated with the head that MO2 did not possess, named those involved in the capsid maturation and morphogenesis. This could give an indication that within the genome of MO2, it is the head portion that is missing from its full assembly, given how similar these two phages. The presence of deoxynucleoside monophosphate kinases and polynucleotide kinases was intriguing, as the former is typically involved in the production of ATP. Polynucleotide kinases tend to play key roles in DNA and RNA repair, as is evidenced by a similar protein found in bacteriophage T4 (Wang, et. al., 2002), though the exact method is not clearly defined. A repeat of this analysis undertaken with the complete genome of MO2 could give further clues into identifying if this gene is also present in MO2 and what its function could be.

Within bacteriophage literature, phages of anaerobic bacteria are scarce and few have been isolated and characterized relative to phages of aerobic bacteria, especially when considering strict anaerobes. Phages against *Clostridium* spp. are currently the most abundant amongst anaerobic bacteria, particularly those against *Clostridioides difficile* (*C. difficile*) and *Clostridium perfringens* (*C. perfringens*) as these are the two most clinically relevant species (Czepiel, et. al., 2019; Kiu and Hall, 2018). *C. difficile* is especially significant – despite there being over 40 isolated lysogenic phages, there have currently been no strictly lytic phages reported and isolated for *C. difficile* (Venhorst, et. al., 2022). Of the currently reported isolated phages, most have been isolated from animals as opposed to humans, making it increasingly difficult to use as a point of reference when looking at phages for anaerobes of the gut microbiota. Thus, the gap in knowledge necessitated the development of a protocol for the isolation of phages targeting *R. gnavus* was required. Prior to Buttimer and colleagues

successfully isolating bacteriophages against *R. gnavus* earlier this year, there were no reports of phages against *R. gnavus* or attempts to do so (Buttimer, et. al., 2023). This meant having to rely upon the brief amount of knowledge that was present while drawing upon what was learnt during the isolation of *Enterobacter* phages and apply these principles to *R. gnavus*. Before beginning to attempt phage isolation, establishing a sufficient lawn that *R. gnavus* would be able to grow and propagate on was necessary as the double agar method was going to be used for phage isolation. Being introduced to *R. gnavus* by Dr. Emmanuelle Crost, learning how to grow *R. gnavus* and in which conditions was necessary. The original growing conditions of *R. gnavus* in BHI-YH for over 48 hours, as according to protocols of the Juge lab, had yet to be tested for growth on agar plates and was done so first (Figure 19). Formation of single colonies on BHI-YH 1.5% agar indicated growth was possible, and that the double agar method used to isolate *Enterobacter* phages was viable for use and could be adapted to aerobic conditions. Thus, this was done in the attempt to form a bacterial lawn fit for use (Figure 20). Bacterial growth was observed but not to the extent that would be ideal to be considered for lawn growth, mainly lacking the density of the lawn that was present in lawns of *E. cloacae*. Buttimer and colleagues successfully reported the isolation and characterisation of multiple temperate phages targeting *R. gnavus*, providing a framework which could be followed and adapted within this project (Buttimer, et. al., 2023). As opposed to BHI-YH, ABB was used as the media of choice for growing *R. gnavus* and isolating phages against it. A comparison of both media was done to identify which media would be ideal alongside Xena Dyball (Figure 21 - 25). BHI-YH could not form visibly distinct lawns compared to ABB; the latter had the appearance of a slightly more transparent region observed within its agar plates that allowed for a distinction between dense bacterial growth. This region, observed by a thin outlined strip (Figure 23) could have been due to the method of agar pouring or because of the media. Within the growth period, OD_{600} for ABB continued to rise and overtook that of BHI-YH after over four hours of incubation, eventually plateauing after six hours. CFU counts for both media displayed a decrease after six hours of incubation, unexpectedly of both though BHI-YH saw an increase after ~6.5 hours. From the results of the comparison, we determined that ABB would be the ideal media for following *R. gnavus* growth, owing to the success by Buttimer, et. al., the visibility of the lawn, and higher OD_{600}. In addition, the growth period could now be shortened to an overnight starter culture and a following six to eight hours of subculture incubation, as opposed to the 48 hours originally used by the Juge group. The ideal volume of *R. gnavus* to be used for lawn formation was next to be determined using mitomycin C, a known bactericidal of *R. gnavus*, with Xena Dyball (Figure 26). Results determined 500 µl of bacterial inoculant was the ideal volume to be used to maximise the number of bacteria, due to the having the densest lawn formation and smallest zones of clearing from mitomycin C. With the optimal lawn conditions established, attempted isolation of phages could now occur.

As a prevalent member of the gut microbiota, it was deduced that phages targeting *R. gnavus* would most likely be found most abundant in faecal samples. *R. gnavus* ATCC 29149, E1, Finegold, CC55_001C and ATCC 35913 were all used as target hosts. This prompted the use of faecal samples from the MOTION study cohort – a group of elderly patients at varying risks of mental decline that were pre-processed by Dr. Oliver Charity prior to their use. In total, 46 different samples from patients were pooled into one sample to be used for phage enrichment of each strain to optimise the process. Alongside these, four wastewater and 13 sewage samples were also used, making a total of 63 samples tested against lawns of each strain (Table 21). The observation of clearing from wastewater samples against *R. gnavus* Finegold could not be confirmed by single plaque purification. Sewage samples collected by collaborators and enriched by hosts *R. gnavus* ATCC 29149, E1, Finegold and CC55_001C demonstrated both clearing and single plaque formation on *R. gnavus* CC55_001C (Figure 27, 28). The observed plaque morphology across all four phages was relatively similar: small, slightly opaque plaques. Similar plaque morphologies could indicate all phages may be similar, but this requires further investigation. Titres of each phage (Table 22) were calculated, and phage MOR1 had the greatest at 1.3×10^8 PFU/ml. DNA extraction and genome sequencing was unable to be performed due

to time constraints but the single plaque formation strongly suggests the successful isolation of, potentially, novel *R. gnavus* phages. Whether these phages are virulent or temperate remains to be seen and would be identified through further characterization assays and genome assembly and annotation. Thus, we propose the successful isolation of, seemingly, four phages targeting *R. gnavus* CC55_001C: MOR1, MOR2, MOR3, and MOR4.

7. Conclusion

Bacteriophages represent a vital aspect of the human gut microbiome as modulators of bacterial diversity and composition, which ultimately has downstream effects on the body. Their potential in phage therapy to combat bacterial infections and antimicrobial resistance has been realised but much remains to be known about the exact effects and outcomes of using phages in such a setting, as well as the regulatory framework, to make phage therapy commonplace. Regardless, bacteriophages represent a promising aspect of human health that requires further investigation.

This thesis describes the successful isolation of three bacteriophages able to successfully infect *Enterobacter cloacae* species NC10005 – *Enterobacter* phages MO1, MO2 and MO3. The work done describes further characterisation of these phages in several aspects; (1) expanded host range; (2) preliminary genome assembly; (3) genome annotation; (4) proposed taxonomic classification and (5) comparative genomics between associated relatives. The results of this thesis propose the isolation of novel unclassified *Karamvirus* species MO1, MO2 and MO3. Additionally, this thesis also describes the successful isolation of four, apparent, novel bacteriophages targeting *Ruminococcus gnavus* species CC55_001C: MOR1, MOR2, MOR3, MOR4.

8. References

Aagard, K., Petrosino, J., Keitel, W., Watson, M., Katancik, J., Garcia, N., Patel, S., Cutting, M., Madden, T., Hamilton, H., Harris, E., Gevers, D., Simone, G., McInnes, P., & Versalovic, J. (2012). The Human Microbiome Project strategy for comprehensive sampling of the human microbiome and why it matters, *The FASEB Journal*, 27(3), pp.1012-1022.

Abedon, S.T. and LeJeune, J.T. (2007). Why Bacteriophage Encode Exotoxins and other Virulence Factors, *Evolutionary Bioinformatics*, 1, pp.97-110.

Abedon, S.T. (2009). Disambiguating bacteriophage pseudolysogeny: an historical analysis of lysogeny, pseudolysogeny and the phage carrier state, *Contemporary Trends in Bacteriophage Research*, pp.285-307.

Ackermann, H.W. (2007). 5500 Phages examined in the electron microscope, *Archives of Virology*, 152(2), pp.227-243.

Ackermann, H.W. (2009). Phage classification and characterization, *Methods in Molecular Biology*, 501, pp.127-140.

Adams, M. J., et. al. (2017). 50 years of the International Committee on Taxonomy of Viruses: progress and prospects, *Archives of Virology*, 162, pp. 1441–1446.

Adriaenssens, E.M. and Brister, J.R. (2017). How to Name and Classify Your Phage: An Informal Guide, *Viruses*, 9(4), 70.

Afgan, E., et al. (2018). "The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update." Nucleic Acids Res 46(W1): W537-W544.

Alcock, B.P., Raphenya, A.R., Lau, T.T.Y., Tsang, K.K., Bouchard, M., Edalatmand, A., Hunyh, Nguyen, A.V., Cheng, A.A., Liu, S., Min, S.Y., Miroshnichenko, A., Tran, H., Werfalli, R.E., Nasir, J.A., Oloni, M., Speicher, D.J., Florescu, A., Singh, B., Faltyn, Hernandez-Koutoucheva, A., Sharma, A.N., Bordeleau, E., Pawlowski, A.C., Zubyk, H.L., Dooley, D., Griffiths, E., Maguire, F., Winsor., G.L., Beiko, R.G., Brinkman, F.S.L., Hsiao, W.W.L., Domselaar, G.V., & McArthur, A.G. (2020). CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database, *Nucleic Acids Research,* 48(D1), pp.D517-D525.

Altschul, S.F., et. al. (1990). Basic local alignment search tool, *Journal of Molecular Biology*, 215, pp.403-410. Available online at: https://blast.ncbi.nlm.nih.gov/Blast.cgi

Annavajhala, M.K., Gomez-Simmonds, A., and Uhlemann, A. (2019). Multidrug-Resistant Enterobacter cloacae Complex Emerging as a Global, Diversifying Threat, *Frontiers in Microbiology,* 10.

Antimicrobial Resistance Collaborators. (2022). Global burden of bacterial antimicrobial resistance in 2019, *The Lancet,* 399(10325), pp.629-655.

Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., and Wishart, D.S. (2016). PHASTER: a better, faster version of the PHAST phage search tool, *Nucleic Acids Research,* 44, pp.W16-W21. Available at: http://phaster.ca/

Balcazar, J.L. (2014). Bacteriophages as Vehicles for Antibiotic Resistance Genes in the Environment, *PLOS Pathogens*, 10(7), E1004219.

Bankevich, A., et. al. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing, *Journal of Computational Biology*, 19(5), pp.455-477.

Berg, D.F., Steen, B.A., Costa, A.R., and Brouns, S.J. (2023). Phage tRNAs evade tRNA-targeting host defenses through anticodon loop mutations, *eLife*, 12(e85183).

Bernstein, C. N., Banerjee, A., Targownik, L. E., Singh, H., Ghia, J. E., Burchill, C., Chateau, D., & Roos, L. L. (2016). Cesarean Section Delivery Is Not a Risk Factor for Development of Inflammatory Bowel Disease: A Population-based Analysis. Clin Gastroenterol Hepatol, 14(1), 50-57. https://doi.org/10.1016/j.cgh.2015.08.005

Biswas, B., Adhya, S., Washart, P., Paul, B., Trostel, A.N., Powell., B., Carlton, R., and Merril, C.R. (2002). Bacteriophage Therapy Rescues Mice Bacteremic from a Clinical Isolate of Vancomycin-Resistant Enterococcus faecium, *Infection and Immunity,* 70(1), pp.204-210.

Black, M., Bhattacharya, S., Philip, S., Norman, J. E., & McLernon, D. J. (2016). Planned Repeat Cesarean Section at Term and Adverse Childhood Health Outcomes: A Record-Linkage Study. *PLoS Med*, 13(3), e1001973.

Bouras, G., Nepal, R., Houtak, G., Psaltis, A.J., Wormald, P., and Vreugde, S. (2023). Pharokka: a fast scalable bacteriophage annotation tool, *Bioinformatics,* 39(1), btac776. Available at: https://github.com/gbouras13/pharokka

Browne, H.P., et. al. (2016). Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation, *Nature,* 533, pp.543-546.

Buttimer, C., Khokhlova, E.V., Stein, L., Hueston, C.M., Govi, B., Draper, L.A., Ross, R.P., Shkoporov, A.N., & Hill, C. (2023). Temperate bacteriophages infecting the mucin-degrading bacterium Ruminococcus gnavus from the human gut, *Gut Microbes,* 15(1), 2194794.

Campbell, A. (2003). The future of bacteriophage biology, *Nature Reviews Genetics*, 4(6), pp.471-477.

Chanishvili, N. and Sharp, R. (2009). A literature review of the practical application of bacteriophage research. Eliava Institute of Bacteriophage, *Microbiology and Virology*, *Tbilisi, Georgia*.

Chapman-Kiddell, C.A., Davies, P.S.W., Gillen, L., and Radford-Smith, G.L. (2010). Role of diet in the development of inflammatory bowel disease, *Inflammatory Bowel Diseases*, 16(1), pp.137-151.

Chen, L., Yang, J., Yu, J., Yao, Z., Sun, L., Shen, Y., & Jin, Q. (2005). VFDB: a reference database for bacterial virulence factors, *Nucleic Acids Research,* 33(suppl_1), pp.D325-D328.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor, *Bioinformatics,* 34(17), pp.i884-i890. Available at: https://github.com/OpenGene/fastp

Chung, I., Sim, N., and Cho, Y. (2012). Antibacterial Efficacy of Temperate Phage-Mediated Inhibition of Bacterial Group Motilities, *Antimicrobial Agents and Chemotherapy,* 56(11), pp.5612-5617.

Cieślik, M., Harhala, M., Orwat, F., Dabrowska, K., Górski, A., and Jończyk-Matysiak, E. (2022). Two Newly Isolated Enterobacter-Specific Bacteriophages: Biological Properties and Stability Studies, *Viruses,* 14(7), 1518.

Clokie, M.R.J., et. al. (2011). Phages in nature, *Bacteriophage*, 1(1), pp.31-45.

Coelho, G. D. P., Ayres, L. F. A., Barreto, D. S., Henriques, B. D., Prado, M., & Passos, C. M. D. (2021). Acquisition of microbiota according to the type of birth: an integrative review. Rev Lat Am Enfermagem, 29, e3446. https://doi.org/10.1590/1518.8345.4466.3446

Comeau, A.M., et. al. (2008). Exploring the prokaryotic virosphere, *Research in Microbiology*, 159(5), pp.306-313.

Cornelissen, A., et. al. (2011). The T7-Related Pseudomonas putida Phage φ15 Displays Virion-Associated Biofilm Degradation Properties, *PLOS ONE*, 6(4), e18597.

Cowe, E., and Sharp, P.M. (1991). Molecular evolution of bacteriophages: Discrete patterns of codon usage in T4 genes are related to the time of gene expression, *Journal of Molecular Evolution,* 33, pp.13-22.

Cowie, D.M. and Hicks, W.C. (1932). Observations on the bacteriophage III, *Journal of Laboratory and Clinical Medicine*, 17, 685.

Crost, E.H., Le Gall, G., Laverde-Gomez, J.A., Mukhopadhya, I., Flint, H.J., and Juge, N. (2018). Mechanistic Insights Into the Cross-Feeding of Ruminococcus gnavus and Ruminococcus bromii on Host and Dietary Carbohydrates, *Frontiers in Microbiology,* 9.

Czepiel, J, Dróżdż, M., Pituch, H.., Kuijper, E.J., Perucki, W., Mielimonka, A., Goldman, S., Wultańska, D., Garlicki, A., & Biesiada, G. (2019). Clostridium difficile infection: review, *European Journal of Clinical Microbiology & Infectious Diseases,* 38, pp.1211-1221.

Das, B., & Nair, G. B. (2019). Homeostasis and dysbiosis of the gut microbiome in health and disease. *Journal of Biosciences*, 44(5).

De Coster, W., D'Hert, S., Schultz, D.T., Cruts, M., and Broeckhoven, C.V. (2018). NanoPack: visualizing and processing long-read sequencing data, *Bioinformatics,* 34(15), pp.2666-2669. Available at: https://github.com/wdecoster/nanostat

DeGruttola, A.K., Low, D., Mizoguchi, A., and Mizoguchi, E. (2016). Current understanding of dysbiosis in disease in human and animal models, *Inflammatory Bowel Diseases*.

De Oliveira, D.M.P., Forde, B.M., Kidd, T.J., Harris, P.N.A., Schembri, M.A., Beatson, S.A., Paterson, D.L., & Walker, M.J. (2020). Antimicrobial Resistance in ESKAPE Pathogens, *Clinical Microbiology Reviews,* 33(3), pp.e00181-189.

Devoto, A.E., Santini, J.M., Olm, M.R., Anantharaman, K., Munk, P., Tung, J., Archie, E.A., Turnbaugh, P.J., Seed, K.D., Blekhman, R., Aarestrup, F.M., Thomas, B.C., & Banfield, J.F. (2019). Megaphages infect Prevotella and variants are widespread in gut microbiomes, *Nature Microbiology,* 4, pp.693-700.

d'Hérelle, F. (1917). Sur un microbe invisible antagoniste des bacilles dysentériques [On an invisible microbe antagonistic to dysenteric bacilli], Comptes Rendus Academie des Sciences [Reports Academy of Sciences], 165(11), pp.373-375.

d'Hérelle, F. (1926). The bacteriophage and its behavior, *Nature,* 183-185.

Dion, M.B., et. al. (2020). Phage diversity, genomics and phylogeny, *Nature Reviews Microbiology*, 18, pp,125-138.

Dominguez-Bello, M.G., Costello, E.K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N., & Knight, R. (2010). Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns, *Proceedings of the National Academy of Sciences,* 107(26), pp.11971-11975.

Dressman, H.K. and Drake, J.W. (1999). Lysis and Lysis Inhibition in Bacteriophage T4: rV Mutations Reside in the Holin t Gene, *Journal of Bacteriology,* 181(14), pp.4391-4396.

Edgar, R., Friedman, N., Molshanski-Mor, S., and Qimron, U. (2011). Reversing bacterial resistance to antibiotics by phage-mediated delivery of dominant sensitive genes, *Applied Environmental Microbiology,* 78(3), pp.744-751.

Ehrlich, S.D. and The MetaHIT Consortium. (2011). MetaHIT: The European Union Project on Metagenomics of the Human Intestinal Tract. In: Nelson, K. (eds) *Metagenomics of the Human Body.* Springer, New York, NY.

Elek, C., Brown, T.L., Viet, T.L., Evans, R., Baker, D.J., Telatin, A., Tiwari, S.K., Al-Khanaq, H., Thilliez, G., Kingsley, R.A., Hall, L.J., Webber, M.A., & Adriaenssens, E.M. (2023). A hybrid and poly-polish workflow for the complete and accurate assembly of phage genomes: a case study of ten przondoviruses, *Microbial Genomics,* 9(7).

Erez, Z., Steinberger-Levy, I., Shamir, M., Doron, S., Stokar-Avihail, A., Peleg, Y., Melamed, S., Leavitt, A., Savidor, A., Albeck, S., Amitai, G., & Sorek, R. (2017). Communication between viruses guides lysis-lysogeny decisions, *Nature,* 541(7638), pp.488-493.

Feiner, R., et. al. (2015). A new perspective on lysogeny: prophages as active regulatory switches of bacteria, *Nature Reviews Microbiology,* 13, pp.641-650.

Goh, S. (2016). Phage Transduction, *Methods in Molecular Biology*, 1476, pp.177-185.

Guerin, E. and Hill, C. (2020). Shining Light on Human Gut Bacteriophages, *Frontiers in Cellular and Infection Microbiology*, 10.

Ferretti, P., Pasolli, E., Tett, A., Asnicar, F., Gorfer, V., Fedi, S., Armanini, F., Truong, D. T., Manara, S., Zolfo, M., Beghini, F., Bertorelli, R., De Sanctis, V., Bariletti, I., Canto, R., Clementi, R., Cologna, M., Crifo, T., Cusumano, G., Gottardi, S., Innamorati, C., Masè, C., Postai, D., Savoi, D., Duranti, S., Lugli, G.A., Mancabelli, L., Turroni, F., Ferrario, C., Milani, C., Mangifesta, M., Anzalone, R., Viappiani, A., Yassour, M., Vlamakis, H., Xavier, R., Collado, C.M., Koren, O., Tateo, S., Soffiati, M., Pedrotti, A., Ventura, M., Huttenhower, C., Bork, P., & Segata, N. (2018). Mother-to-Infant Microbial Transmission from Different Body Sites Shapes the Developing Infant Gut Microbiome, Cell *Host Microbe*, 24(1), pp.133-145, e135.

Filippo, C.D., Cavalieri, D., Paola, M.D., Ramazzotti, M., Poullet, J.B., Massart, S., Collini, S., Pieraccini, G., & Lionetti, P. (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa, *Proceedings of the National Academy of Sciences,* 107(33), pp.14691-14696.

Fleming, A. (1929). On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to their Use in the Isolation of B. influenzæ, *The British Journal of Experimental Pathology*, 10(3), pp.226-236.

Fleming, A. (1945, June 26). Penicillin's finder assays its future, *The New York Times*. Accessed at: https://www.nytimes.com/1945/06/26/archives/penicillins-finder-assays-its-future-sir-alexander-fleming-says.html

Gilbert, J.A. (2015). Our unique microbial identity, *Genome Biology,* 16(1), 97.

Gilchrist, C.L.M. and Chooi, Y. (2021). clinker & clustermap.js: automatic generation of gene cluster comparison figures, *Bioinformatics*, 37(16), pp.2473-2475. Available at: https://cagecat.bioinformatics.nl/

Glonti, T., et. al. (2010). Bacteriophage-derived enzyme that depolymerizes the alginic acid capsule associated with cystic fibrosis isolates of Pseudomonas aeruginosa, *Journal of Applied Microbiology*, 108(2), pp.695-702.

Górski, A., Wazna, E., Dabrowska, B., Dabrowska, K., Switała-Jeleń, K., and Miedzybrodzki, R. (2006). Bacteriophage translocation, *FEMS Immunology and Medical Microbiology,* 46(3), pp.313-319.

Guerin, E. and Hill, C. (2020). Shining Light on Human Gut Bacteriophages, *Frontiers in Cellular and Infection Microbiology*, 10.

Hadley, P. (1928). The Twort-D'Herelle Phenomenon: A Critical Review and Presentation of a New Conception (Homogamic Theory) Of Bacteriophage Action, *Journal of Infectious Diseases*, 42(4), pp.263-434.

Hall, A.B., Yassour, M., Sauk, J., Garner, A., Jiang, X., Arthur, T., Lagoudas, G.K., Vatanen, T., Fornelos, N., Wilson, R., Bertha, M., Cohen, M., Garber, J., Khalili, H., Gevers, D., Ananthakrishnan, A.N., Kugathasan, S., Lander, E.S., Blainey, P., Vlamakis, H., Xavier, R.J., & Huttenhower, C. (2017). A novel Ruminococcus gnavus clade enriched in inflammatory bowel disease patients, *Genome Medicine,* 9(103).

Henke, M.T., Kenny, D.J., Cassilly, C.D., Vlamakis, H., Xavier, R.J., and Clardy, J. (2019). Ruminococcus gnavus, a member of the human gut microbiome associated with Crohn's disease, produces an inflammatory polysaccharide, *Proceedings of the National Academy of Sciences,* 116(26), pp.12672-12677.

Herzog, C. (1976). Chemotherapy of typhoid fever: a review of literature, *Infection,* 4(3), pp.166-173.

Hodyra-Stefaniak, K., Miernikiewicz, P., Drapała, J., Drab, M., Jończyk-Matysiak, E., Lecion, D., Kaźmierczak., Z., Beta, W., Majewska, J., Harhala, M., Bubak, B., Kłopot, A., Górski, A., & Dabrowska, K. (2015). Mammalian Host-Versus-Phage immune response determines phage fate in vivo, *Scientific Reports,* 5, 14802.

Ho, K. (2001). Bacteriophage Therapy for Bacterial Infections: rekindling a memory from the pre-antibiotics era, *Perspectives in Biology and Medicine*, 44(1), pp.1-16.

Hou, J.K., Abraham, B., and El-Serag, H. (2011). Dietary Intake and Risk of Developing Inflammatory Bowel Disease: A Systematic Review of the Literature, *American Journal of Gastroenterology,* 106(4), pp.563-573.

Huurre, A., Kalliomaki, M., Rautava, S., Rinne, M., Salminen, S., & Isolauri, E. (2008). Mode of delivery - effects on gut microbiota and humoral immunity. Neonatology, 93(4), 236-240. https://doi.org/10.1159/000111102

Hyatt, D., Chen, G., LoCascio, P.F., Land, M.L., Larimer, F.W., & Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification, *BMC Bioinformatics,* 11.

Jandhyala, S. M., Talukdar, R., Subramanyam, C., Vuyyuru, H., Sasikala, M., & Nageshwar Reddy, D. (2015). Role of the normal gut microbiota. World J Gastroenterol, 21(29), 8787-8803. https://doi.org/10.3748/wjg.v21.i29.8787

Jevons, M.P. (1961). "Celbenin"-resistant Staphylocci, *British Medical Journal*, 1, pp.124-125.

Jimenez, E., Marin, M. L., Martin, R., Odriozola, J. M., Olivares, M., Xaus, J., Fernandez, L., & Rodriguez, J. M. (2008). Is meconium from healthy newborns actually sterile? Res Microbiol, 159(3), 187-193. https://doi.org/10.1016/j.resmic.2007.12.007

Johnson, J.S., Spakowicz, D.J., Hong, B., Petersen, L.M., Demkowicz, P., Chen, L., Leopold, S.R., Hanson, B.M., Agresta, H.O., Gerstein, M., Sodergren, E., & Weinstock, G.M. (2019). Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis, *Nature Communications,* 10.

Jost, T., Lacroix, C., Braegger, C.P., Rochat, F., and Chassard, C. (2013). Vertical mother-neonate transfer of maternal gut bacteria via breastfeeding, *Environmental Microbiology,* 16(9), pp.2891-2904.

Jurczak-Kurek, A., et. al. (2016). Biodiversity of bacteriophages: morphological and biological properties of a large group of phages isolated from urban sewage, *Scientific Reports*, 6, 34338.

Kaur, D.C. and Chate, S.S. (2015). Study of Antibiotic Resistance Pattern in Methicillin Resistant Staphylococcus Aureus with Special Reference to Newer Antibiotic, *Journal of Global Infectious Diseases*, 7(2), pp.78-84.

Kaushik, D., et. al. (2014). Ampicillin: Rise Fall and Resurgence, *Journal of Clinical & Diagnostic Research*, 8(5), pp.ME01-ME03.

Kellenger, E. (2001). Exploring the unknown, *EMBO Reports,* 2(1), pp.5-7.

Kennedy, K.M., et. al. (2023). Questioning the fetal microbiome illustrates pitfalls of low-biomass microbial studies, *Nature*, 613, pp.639-649.

Kho, Z.Y., and Lal, S.K. (2018). The Human Gut Microbiome – A Potential Controller of Wellness and Disease, *Frontiers in Microbiology*, 9.

Kim, Y.J., Kang, H.Y., Han, Y., Lee, M.S., and Lee, H.J. (2017). A bloodstream infection by Ruminococcus gnavus in a patient with a gall bladder perforation, *Anaerobe,* 47, pp.129-131.

Kiu, R., and Hall, L.J. (2018). An update on the human and animal enteric pathogen Clostridium perfringens, *Emerging Microbes & Infections*, 7(141).

Knox, R. (1960). A New Penicillin (BRL 1241) Active Against Penicillin-resistant Staphylocci, *British Medical Journal*, 2(5200), pp.690-693.

Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P.A. (2019). Assembly of long, error-prone reads using repeat graphs, *Nature Biotechnology,* 37, pp.540-546. Available at: https://github.com/fenderglass/Flye

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation, *Genome Research,* 27, pp.722-736. Available at: https://github.com/marbl/canu

Krestownikowa, W. and Gubin, W. (1925). Die Verteilung and die Ausscheidungvon Bak-teriophagen im Meerschweinchen-organismus bei subkutaner Appli-cationsart [The distribution and excretion of bacteriophages in the guinea pig organism when administered subcutaneously.], *J. Microbiol., Patolog. i. Infekzionnich bolesney* [*Journal of Microbiology, Pathology, and Infectious Diseases*], 1(3).

Kristensen, K., & Henriksen, L. (2016). Cesarean section and disease associated with immune function. J Allergy Clin Immunol, 137(2), 587-590. https://doi.org/10.1016/j.jaci.2015.07.040

Kutateladze, M. and Adamia, R. (2010). Bacteriophages as potential new therapeutics to replace or supplement antibiotics, *CellPress: Trends in Biotechnology*, 28(12), pp.591-595.

Liang, G., Zhao, C., Zhang, H., Mattei, L., Sherrill-Mix, S., Bittinger, K., Kessler, L. R., Wu, G. D., Baldassano, R. N., DeRusso, P., Ford, E., Elovitz, M. A., Kelly, M. S., Patel, M. Z., Mazhani, T., Gerber, J. S., Kelly, A., Zemel, B. S., & Bushman, F. D. (2020). The stepwise assembly of the neonatal virome is modulated by breastfeeding. *Nature*, 581(7809), pp.470-474.

Lim, E.S., Zhou, Y., Zhao, G., Bauer, I.K., Droit, L., Ndao, I.M., Warner, B.B., Tarr, P.I., Wang, D., & Holtz, L.R. (2015). Early life dynamics of the human gut virome and bacterial microbiome in infants, *Nature Medicine,* 21(10), pp.1228-1234.

Lin, D.M., Koskella, B., and Lin, H.C. (2017). Phage therapy: An alternative to antibiotics in the age of multi-drug resistance, *World Journal of Gastrointestinal Pharmacology and Therapeutics*, 8(3), pp.162-173.

Lloyd-Price, J., Abu-Ali, G., and Huttenhower, C. (2016). The healthy human microbiome, *Genome Medicine,* 8(51).

Łoś, M. and Węgrzyn, G. (2012). Pseudolysogeny, *Advances in Virus Research*, 82, pp.339-349.

Lowy, F.D. (2003). Antimicrobial resistance: the example of Staphylococcus aureus, *The Journal of Clinical Investigation*, 111(9), pp.1265-1273.

Ludwig, W., Schleifer, K-H., and Whitman, W.B. (2009) Revised road map to the phylum Firmicutes. In: De Vos, P., Garrity, G.M., Jones, D., Krieg, N.R., Ludwig, W., Rainey, F.A., Schleifer, K-H., and Whitman, W.B. (eds), *Bergey's Manual® of Systematic Bacteriology: Volume Three The Firmicutes*. New York: Springer, 1–13.

Luria, S.E. and Delbrück, M. (1943). Mutations of bacteria from virus sensitivity to virus resistance, *Genetics*, 28(6), pp.491-511

Majowicz, S.E., et. al. (2014). Global incidence of human Shiga toxin-producing Escherichia coli infections and deaths: a systematic review and knowledge of synthesis. *Foodborne Pathogens and Disease*, 11, pp.447-455.

Malmuthuge, N., & Griebel, P. J. (2018). Fetal environment and fetal intestine are sterile during the third trimester of pregnancy. Vet Immunol Immunopathol, 204, 59-64.

Marinus, M.G. and Poteete, A.R. (2013). High efficiency generalised transduction in Escherichia coli O157:H7, *F1000 Research*, 2(7).

McCallin, S., Sarker, A.S., Barretto, C., Sultana, S., Berger, B., Huq, S., Krause, L., Bibiloni, R., Schmitt, B., Reuteler, G., and Brüssow, H. (2013). Safety analysis of a Russian phage cocktail: from metagenomic analysis to oral application in healthy human subjects, *Virology,* 443(2), pp.187-196.

McNair, K., Zhou, C., Dinsdale, E.A., Souza, B., and Edwards, R.A. (2019). PHANOTATE: a novel approach to gene identification in phage genomes, *Bioinformatics,* 35(22), pp.4537-4542. Available at: https://github.com/deprekate/PHANOTATE

Merabishvili, M., Pirnay, J., Verbeken, G., Chanishvili, N., Tediashvili, M., Lashki, N., Glonti, T., Krylov, V., Mast, J., Parsys, L., Lavigne, R., Volckaert, G., Mattheus, W., Verween, G., Corte, P., Rose, T., Jennes, S., Zizi, M., Vos, D., & Vaneechoutte, M. (2009). Quality-Controlled Small-Scale Production of a Well-Defined Bacteriophage Cocktail for Use in Human Clinical Trials. *PLoS ONE,* 4(3), e4944.

Mihara, T., Nishimura, Y., Shimizu, Y., Nishiyama, H., Yoshikawa, G., Uehara, H., Hingamp, P., Goto, S., & Ogata, H. (2016). Linking Virus Genomes with Host Taxonomy, *Viruses,* 8(3), 66.

Moraru, C., Varsani, A., and Kropinski, A.M. (2020). VIRIDIC - a novel tool to calculate the intergenomic similarities of prokaryote-infecting viruses, *Viruses,* 12(1268). Available at: https://rhea.icbm.uni-oldenburg.de/viridic/

Morozova, V., Jdeed, G., Kozlova, Y., Babkin, I., Tikunov, A., and Tikunova, N. (2021). A New Enterobacter cloacae Bacteriophage EC151 Encodes the Deazaguanine DNA Modification Pathway and Represents a New Genus within the Siphoviridae Family, *Viruses,* 13(7), 1372.

Mushegian, A.R. (2020). Are There 10^{31} Virus Particles on Earth, or More, or Fewer?, *Journal of Bacteriology*, 202(9).

Nasr-Eldin, M.A., Gamal, E., Hazza, M., and Abo-Elmaaty, S.A. (2023). Isolation, characterization, and application of lytic bacteriophages for controlling Enterobacter cloacae complex (ECC) in pasteurized milk and yogurt, *Folia Microbiologica*.

Naureen, Z., Malacarne, D., Anpilogov, K., Dautaj, A., Camilleri, G., Cecchin, S., Bressan, S., Casadei, A., Albion, E., Sorrentino, E., Beccari, T., Dundar, M., & Bertelli, M. (2020). Comparison between American and European legislation in the therapeutical and alimentary bacteriophage usage, *Acta Biomedica*, 91(Suppl 13), e2020023.

NCBI Resource Coordinators. (2016). NCBI Bacterial Antimicrobial Resistance Reference Database, *NCBI*. Available at: https://www.ncbi.nlm.nih.gov/bioproject/313047

Nishimura, Y., Yoshida, T., Kuronishi, M., Uehara, H., Ogata, H., and Goto, S. (2017). VIPTree: the viral proteomic tree server, *Bioinformatics,* 33(15), pp.2379-2380. Available at: https://www.genome.jp/viptree/

Oren, A. and Garrity, G.M. (2021). Valid publication of the name of forty-two phyla of prokaryotes, *International Journal of Systematic and Evolutionary Microbiology,* 71(10).

Panlilio, A.L., et. al. (1992). Methicillin-resistant Staphylococcus aureus in U.S. hospitals, 1975-1991, *Infection Control & Hospital Epidemiology*, 13(10), pp.582-586.

Park, K., Cha, K.E., and Myung, H. (2014). Observation of inflammatory responses in mice orally fed with bacteriophage T7, *Journal of Applied Microbiology,* 117(3), pp.627-633.

Pawluk, A., Bondy-Denomy, J., Cheung, V.H.W., Maxwell, K.L., and Davidson, A.R. (2014). A New Group of Phage Anti-CRISPR Genes Inhibits the Type I-E CRISPR-Cas System of Pseudomonas aeruginosa, *mBio,* 5(2), pp.e00896-14.

Penadés, J.R., et. al. (2015). Bacteriophage-mediated spread of bacterial virulence genes, *Current Opinion in Microbiology,* 23, pp.171-178.

Rock, C., and Donnenberg, M.S. (2014). Human Pathogenic Enterobacteriaceae, *Reference Module in Biomedical Sciences*.

Rothschild, D., Weissbrod, O., Barkan, E., Kuilshikov, A., Korem, T., Zeevi, D., Costea, P.I., Godneva, A., Kalka, I.N., Bar, N., Shilo, S., Lador, D., Vila, A.V., Zmora, N., Pevsner-Fischer, M., Israeli, D., Kosower, N., Malka, G., Wolf, B.C., Avnit-Sagu, T., Lotan-Pompan, M., Weinberger, A., Halpern, Z., Carmi, S., Fu, J., Wijmenga, C., Zhernakova, A., Elinav, E., & Segal, E. (2018). Environment dominates over host genetics in shaping human gut microbiota, *Nature,* 555, pp.210-215.

Saga, T. and Yamaguchi, K. (2009). History of Antimicrobial Agents and Resistant Bacteria, *Japan Medical Association Journal*, 52(2), pp.103-108.

Sagheddu, V., Patrone, V., Miragoli, F., Puglisi, E., and Morelli, L. (2016). Infant Early Gut Colonization by Lachnospiraceae: High Frequency of Ruminococcus gnavus, *Frontiers in Pediatrics,* 2(4), 57.

Santajit, S., and Indrawattana, N. (2016). Mechanisms of Antimicrobial Resistance in ESKAPE Pathogens, *Biomed Research International,* 2475067.

Sarker, S.A., Sultana, S., Reuteler, G., Moine, D., Descombes, P., Charton, F., Bourdin, G., McCallin, S., Ngom-Bru, C., Neville, T., Akter, M., Huq, S., Qadri, F., Talukdar, K., Kassam, M., Delley, M., Loiseau, C., Deng, Y., Aidy, S.E., Berger, B., & Brüssow, H. (2016). Oral Phage Therapy of Acute Bacterial Diarrhea With Two Coliphage Preparations: A Randomized Trial in Children From Bangladesh, *eBioMedicine,* 4, pp.124-137.

Saussereau, E. and Debarbieux, L. (2012). Chapter 4 – Bacteriophages in the Experimental Treatment of Pseudomonas aeruginosa Infections in Mice, *Advances in Virus Research*, 83, pp.123-141.

Schoch, C.L., Cifuo, S., Domrachev, M., Hotton, C.L., Kannan, S., Khovanskaya, R., Leipe, D., Mcveigh, R., O'Neill, K., Robbertse, B., Sharma, S., Soussov, V., Sullivan, J.P., Sun, L., Turner, S., & Karsch-Mizrachi, I. (2020). NCBI Taxonomy: a comprehensive update on curation, resources and tools, *Database (Oxford)*, baaa062.

Seeman, T. (2014). Prokka: rapid prokaryotic genome annotation, *Bioinformatics*, 30(14), pp.2068-2069.

Seeman, T. (2017). Shovill: Faster SPAdes assembly of Illumina reads. Available at: https://github.com/tseemann/shovill

Servick, K. (2016). Beleaguered phage therapy trial presses on, *Science,* 352(6293), 1506.

Shen, A., and Milard, A. (2021). Phage Genome Annotation: Where to Begin and End, *Phage (New Rochelle),* 2(4), pp.183-193.

Siguier, P., et. al. (2006). ISfinder: the reference centre for bacterial insertion sequences, *Nucleic Acids Research*, 34, pp.D32-D36. Available at: https://www-is.biotoul.fr/index.php

Smith, H.W. and Huggins, M.B. (1982). Successful Treatment of Experimental Escherichia coli Infections in Mice Using Phage: its General Superiority over Antibiotics, *Microbiology Society*, 128(2).

Smith, H.W. and Huggins, M.B. (1983). Effectiveness of Phages in Treating Experimental Escherichia coli Diarrhoea in Calves, Piglets and Lambs, *Microbiology Society,* 129(8).

Smith, H.W., et. al. (1987). Factors Influencing the Survival and Multiplication of Bacteriophages in Calves and in Their Environment, *Microbiology Society*, 133(5).

Solís, G., de los Reyes-Gavilan, C.G., Fernández, N., Margolles, A., and Gueimonde, M. (2010). Establishment and development of lactic acid bacteria and bifidobacteria microbiota in breast-milk and the infant gut, *Anaerobe,* 16(3), pp.307-310.

Soothill, J.S. (1992). Treatment of experimental infections of mice with bacteriophages, *Journal of Medical Microbiology,* 37(4), pp.258-261.

Stewart, C.J., Ajami, N.J., O'Brien, J.L., Hutchinson, D.S., Smith, D.P., Wong, M.C., Ross, M.C., Lloyd, R.E., Doddapaneni, H., Metcalf, G.A., Muzny, D., Gibbs, R.A., Vatanen, T., Huttenhower, C., Xavier, R.J., Rewers, M., Hagopian, W., Toppari, J., Anette-G. Ziegler, She, J., Akolkar, B., Lernmark, A., Hyoty, H., Vehik, K., Krischer, J.P., & Petrosino, J.F. (2018). Temporal development of the gut microbiome in early childhood from the TEDDY study, *Nature,* 562, pp.583-588.

Stojanov, S., Berlec, A., and Štrukelj, B. (2020). The Influence of Probiotics on the Firmicutes/Bacteroidetes Ratio in the Treatment of Obesity and Inflammatory Bowel disease, *Microorganisms,* 8(11), 1715.

Tannock, G.W., Lawley, B., Munro, K., Pathmanathan, S.G., Zhou, S.J., Makrides, M., Gibson, R.A., Sullivan, T., Prosser, C.G., Lowry, D., & Hodgkinson, A.J. (2013). Comparison of the compositions of the stool microbiotas of infants fed goat milk formula, cow milk-based formula, or breast milk, *Applied and Environmental Microbiology,* 79(9), pp.3040-3048.

Terzian, P., Ndela, E.O., Galiez, C., Lossouarn, J., Bucio, R.E.P., Mom, R., Toussaint, A., Petit, M., & Enault, F. (2021). PHROG: families of prokaryotic virus proteins clustered using remote homology, *NAR Genomics and Bioinformatics,* 3(3), lqab067.

The Human Microbiome Project Consortium. (2012). Structure, function and diversity of the healthy human microbiome, *Nature*, 486, pp.207-214.

The UniProt Consortium. (2021). UniProt:the universal protein knowledgebase in 2021, *Nucleic Acids Research*, 49(D1), pp.D480-D489.

Togo, A.H., Diop, A., Bittar, F., Maranichi, M., Valero, R., Armstrong, N., Dubourg, G., Labas, N., Richez, M., Delerce, J., Levasseur, A., Fournier, P., Raoult, D., & Million, M. (2018). Description of

Mediterraneibacter massiliensis, gen. nov., sp. nov., a new genus isolated from the gut microbiota of an obese patient and reclassification of Ruminococcus faecis, Ruminococcus lactaris, Ruminococcus torques, Ruminococcus gnavus and Clostridium glycyrrhizinilyticum as Mediterraneibacter faecis comb. nov., Mediterraneibacter lactaris comb. nov., Mediterraneibacter torques comb. nov., Mediterraneibacter gnavus comb. nov. and Mediterraneibacter glycyrrhizinilyticus comb. nov, *Antonie Van Leeuwenhoek,* 111(11), pp.2107-2128.

Topka-Bielecka, G., et. al. (2021). Bacteriophage-Derived Depolymerases against Bacterial Biofilm, *Antibiotics*, 10(2), 175.

Tsao, Y., Taylor, V.L., Kala, S., Bondy-Denomy, J., Khan, A.N., Bona, D., Cattoir, V., Lory, S., Davidson, A.R., and Maxwell, K.L. (2018). Phage Morons Play an Important Role in Pseudomonas aeruginosa Phenotypes, *Journal of Bacteriology,* 200(22), pp.e00189-18.

Turnbaugh, P.J., Ley, R.E., Hamady, M., Fraser-Liggett, C.M., Knight, R., & Gordon, J.I. (2007). The Human Microbiome Project, *Nature*, 449, pp.804-810.

Turner, D., Adriaenssens, E.M., Tolstoy, I., and Kropinski, A.M. (2021). Phage Annotation Guide: Guidelines for Assembly and High-Quality Annotation, *Phage (New Rochelle),* 2(4), pp.170-182.

Turner, D., Shkoporov, A.N., Lood, C., Millard, A.D., Dutilh, B.E., Alfenas-Zerbini, P., van Zyl, L.J., Aziz, R.K., Oksanen, H.M., Poranen, M.M., Kropinski, A.M., Barylski, J., Brister, J.R., Chanisvili, N., Edwards, R.A., Enault, F., Gillis, A., Knezevic, P., Krupovic, M., Kurtböke, I., Kushkina, A., Lavigne, R., Lehman, S., Lobocka, M., Moraru, C., Swift, A.M., Morozova, V., Nakavuma, J., Muñoz, A.R., Rūmnieks, J., Sarkar, B.L., Sullivan, M.B., Uchiyama, J., Wittman, J., Yigang, T., & Adriaenssens, E.M. (2023). Abolishment of morphology-based taxa and change to binomial species names: 2022 taxonomy update of the ICTV bacterial viruses subcommittee, *Archives of Virology,* 168(2), 74.

Twort, F. (1915). An Investigation On The Nature of Ultra-Microscopic Viruses, *The Lancet*, 186(4814), pp.1241-1243.

Venegas, D.P., De la Fuenta, M.K., Landskron, G., González, M.J., Quera, R., Dijkstra, G., Harmsen, H.J.M., Faber, K.N., & Hermoso, M.A. (2019). Short Chain Fatty Acids (SCFAs)-Mediated Gut Epithelial and Immune Regulation and Its Relevance for Inflammatory Bowel Diseases, *Frontiers in Immunology*, 10.

Verbeken, G., Pirnay, J., Vos, D., Jennes, S., Zizi, M., Lavigne, R., Casteels, M., & Huys, I. (2012). Optimizing the European regulatory framework for sustainable bacteriophage therapy in human medicine, *Archivum Immunologiae et Therapiae Experimentalis* [*Archives of Immunology and Experimental Therapy*], 60(3), pp.161-172.

Vinolo, M.A.R., Rodrigues, H.G., Nachbar, R.T., and Curi, R. (2011). Regulation of Inflammation by Short Chain Fatty Acids, *Nutrients,* 3(10), pp.858-876.

Wang, J., Hu, B., Xu, M., Yan, Q., Liu, S., Zhu, X., Sun, Z., Tao, D., Ding, L., Reed, E., Gong, J., Li, Q.Q., and Hu, J. (2006). Therapeutic effectiveness of bacteriophages in the rescue of mice with extended spectrum beta-lactamase-producing Escherichia coli bacteremia, *International Journal of Molecular Medicine,* 17(2), pp.347-355.

Wang, L.K., Lima, C.D., and Shuman, S. (2002). Structure and mechanism of T4 polynucleotide kinase: an RNA repair enzyme, *The EMBO Journal*, 21(14), pp.3873-3880.

Watanabe, R., Matsumoto, T., Sano, G., Ishii, Y., Tateda, K., Sumiyama, Y., Uchiyama, J., Sakurai, S., Matsuzaki, S., Imai, S., & Yamaguchi, K. (2007). Efficacy of Bacteriophage Therapy against Gut-Derived Sepsis Caused by Pseudomonas aeruginosa in Mice, *Antimicrobial Agents and Chemotherapy,* 51(2), pp.446-452.

Weber-Dabrowska, B., Mulczyk, M., and Górski, A. (2000). Bacteriophage therapy of bacterial infections: an update of our institute's experience, *Archivum Immunologiae et Therapiae Experimentalis* [*Archives of Immunology and Experimental Therapy*], 48(6), pp.547-51.

Wick, R.R., and Holt, K.E. (2022). Polypolish: short-read polishing of long-read bacterial genome assemblies, *PLoS Computational Biology*, 18(e1009802).

Wick, R.R., Schultz, M.B., Zobel, J., and Holt, K.E. (2015). Bandage: interactive visualization of de novo genome assemblies, *Bioinformatics,* 31(20), pp.3350-3352.

World Health Organization (WHO). Antimicrobial Resistance. Available online: https://www.who.int/health-topics/antimicrobial-resistance

Wright, C., and Wykes, M. (2020). Medaka; available at: https://github.com/nanoporetech/medaka

Yang, L., Sakandar, H.A., Sun, Z., and Zhang, H. (2021). Recent advances of intestinal microbiota transmission from mother to infant, *Journal of Functional Foods,* 87.

Yuan, Y., and Gao, M. (2017). Jumbo Bacteriophages: An Overview, *Frontiers in Microbiology,* 8.

Zablocki, O., Michelsen, M., Burris, M., Soloneko, N., Warwick-Dugdale, J., Ghosh, R., Pett-Ridge, J., Sullivan, M.B., and Temperton, B. (2021). VirION2: a short- and long-read sequencing and informatics workflow to study the genomic diversity of viruses in nature, *PeerJ,* 9(e11088).

Zhang, H., Fouts, D.E., DePew, J., and Stevens, R.H. (2013). Genetic modifications to temperate Enterococcus faecalis phage φEf11 that abolish the establishment of lysogeny and sensitivity to repressor, and increase host range and productivity of lytic infection, *Microbiology*, 159(Pt_6).

Zhang, X., Yu, D., Wu, D., Gao, X., Shao, F., Zhao, M., Wang, J., Ma, J., Wang, W., Qin., Chen, Y., Xia, P., & Wang, S. (2023). Tissue-resident Lachnospiraceae family bacteria protect against colorectal carcinogenesis by promoting tumor immune surveillance, *Cell Host & Microbe,* 31(3), pp.418-432.e8.

Zimin, A.V., and Salzberg, S.L. (2020) The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies, *PLoS Computational Biology*, 16(e1007981)