
# Towards Phage Therapy: Characterization of *Clostridioides difficile* Bacteriophages from a South African Strain Collection

By

Cara Golding

A Thesis Presented for the Degree of
MASTER OF SCIENCE

In the
Division of Medical Microbiology, Department of Pathology,
UNIVERSITY OF CAPE TOWN

*September 2023*

Primary Supervisor: Dr Lynthia Paul

Co-supervisor: Dr Brian Kullin

Declaration:

I, Cara Golding, hereby:

a. declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.
b. grant the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate

Date: 2 September 2023

# Table of Contents

# List of tables

# Table of Figures

# Abbreviations

| | |
|---|---|
| Abi | Abortive infection |
| AMR | Antimicrobial resistance |
| BHI | Brain-heart infusion |
| CDI | *Clostridioides difficile* infection |
| CRISPR | Clustered regularly interspaced short palindromic repeats |
| EtBr | Ethidium bromide |
| DNA | Deoxyribonucleic acid |
| dNTP | Deoxyribonucleoside triphosphate |
| FDA | Food and Drug Administration |
| FMT | Faecal microbiota transplant |
| GIT | Gastrointestinal tract |
| MMC | Mitomycin C |
| $OD_{600}$ | Optical density at 600nm wavelength |
| ORF | Open reading frame |
| PAM | Protospacer adjacent motif |
| PBS | Phosphate buffered saline |
| PCR | Polymerase chain reaction |
| PEG | Polyethylene glycol |
| PMC | Pseudomembranous colitis |
| PTLP | Phage tail-like particle |
| RM | Restriction modification |
| RNA | Ribonucleic acid |
| rRNA | Ribosomal ribonucleic acid |
| RT | Ribotype |
| SM | Storage media |
| TEM | Transmission electron microscopy |
| UV | Ultraviolet |
| VLP | Virus-like particle |
| WGS | Whole genome sequence |

# ABSTRACT

*Clostridioides difficile* infection is a worldwide public health concern that affects persons with gastrointestinal dysbiosis, notably hospitalised patients on antibiotic therapy and those with other types of gastrointestinal conditions. This opportunistic infection is caused by a Gram-positive, spore-forming bacillus, with conditions including diarrhoea, pseudomembranous colitis, and toxic megacolon. Antibiotic resistance has been reported for the standard of care antibiotics, while newer drugs are too costly for resource-limited clinical settings. Faecal microbiota transplantation (FMT) is highly effective but may pose a risk to immunocompromised individuals. Phage therapy is being increasingly investigated as a therapy option, notably for persons who cannot receive FMT. However, the host and regional specificity of *C. difficile* phages must first be characterised to ensure the correct phage cocktail is used for therapy. Although *C. difficile* phages have been described in other parts of the world, analysis of local South African strains has not been conducted prior to this study. This project aimed to isolate and characterise phages from a stored collection of South African *C. difficile* isolates, focusing on ribotype 017, which has been identified in Western Cape hospitalised patients with tuberculosis. Using a bioinformatic approach, we extracted phage genomes from previously generated *C. difficile* genome assemblies, assessed genomic relatedness, organised phage genomes into functional modules, and identified host defence systems. Phage induction was done using mitomycin C in liquid cultures and plaque overlay assays. Attempts were made to purify phages and generate PCR amplicons for sequence confirmation. The results of this study demonstrate the presence of phages in local *C. difficile* isolates and provide evidence for their classification as *Caudoviricetes*. Further studies are needed to determine the specific taxonomy, since recent updates have rendered previous morphology-based classification of phages inadequate.

## PROBLEM STATEMENT

Antibiotic resistance has been declared a global threat by the World Health Organization and the Centres for Disease Control & Prevention (CDC) (1, 2). *Clostridioides difficile* is considered globally as a leading cause of antibiotic-associated infectious diarrhoea (3). Affected persons have a 25% chance of developing recurrence following the initial episode of *Clostridioides difficile* infection (CDI) (4) and after 2 recurrent episodes, recurrence can be as high as 65% (5). With the increase in antibiotic resistance, newer therapies are urgently needed. One alternative, namely faecal microbiota transplantation (FMT), shows high efficacy, but it is not without risk, as serious adverse effects include infection and death (6, 7). The use of bacteriophages (phages) might be a therapeutic avenue for predisposed persons with drug resistant CDI. Phage therapy, already showing success in Eastern European countries as an accepted antibacterial therapy, is gaining interest in Western countries (8). The use of narrow host range phages, such as *C. difficile* phages prevents killing of non-targeted (beneficial) host microbes and is, therefore, useful to prevent further opportunistic infections. However, to be optimally effective, this requires knowledge of what specific strains or ribotypes are circulating in different geographical areas. Limited knowledge on South African *C. difficile* ribotypes and associated phages exist. Therefore, this study aims to fill that knowledge gap, aiming to pioneer an avenue for phage therapy in South Africa.

## LITERATURE REVIEW

CDI is a global health threat associated with the use of antibiotics. The disease is characterised by diarrhoea and inflammation of the colon. The Centres for Disease Control & Prevention (CDC) is cognisant of the threat posed by CDI on patient health, as well as the associated cost of care associated with CDI, particularly in the case of recurrent disease (9). In South Africa, CDI has been reported in 22% of patients with diarrhoea in some settings. Importantly, given the high prevalence of the disease in sub-Saharan Africa, tuberculosis has been identified as an additional risk factor, likely due to the intensive antimicrobial therapy used in treatment (10, 11). This review will (i) briefly explore the organism responsible for CDI, its pathogenesis, and currently approved therapies (Part A). The focus will then shift (ii) to phage therapy and its emerging role to treat bacterial infections such as CDI (Part B).

## **PART A:** *C. DIFFICILE*, ITS PATHOGENESIS, THERAPY AND DRUG RESISTANCE.

Hall and O'Toole first described a hitherto unknown microorganism that was isolated from the gut of a healthy neonate in 1935 (12). A Gram-positive, strictly anaerobic rod with subterminal spores (13), the microorganism was named *Bacillus difficilis* to reflect the difficulty of isolating it (12). In 1978, it was renamed to *Clostridium difficile*, drawing its etymology from the Greek word 'kloster' denoting 'spindle' (14, 15). In 2016, following the restriction of the *Clostridium* genus to contain only *C. butyricum* and related species, *C. difficile* along with *Clostridium mangenotii* were placed in the newly formed genus *Clostridioides* based on alignment of the 16S rRNA gene (16). However, *Clostridium difficile* is still frequently used, and the nomenclature is still accepted (17).

### *C. difficile* infection

#### Disease presentation

CDI encompasses a wide range of symptoms, ranging from mild to moderate self-limiting diarrhoea to more severe pseudomembranous colitis (PMC). Potentially fatal complications such as fulminant colitis (toxic megacolon) are also possible. CDI patients most commonly present with colitis in the absence of a pseudomembrane (18). It is most severe in the rectum, sigmoid and descending colon and, to a lesser extent, the right colon (19).

*C. difficile* toxins TcdA and TcdB are both implicated in the development of disease pathology by stimulating production of proinflammatory cytokines and destruction of the cellular cytoskeleton (20, 21). Raised yellow plaques between 2 and 10mm in diameter are characteristic features of PMC and can be seen by sigmoidoscopy of the colorectal mucosa (22). *C. difficile* colitis becomes fulminant with the onset of septic shock and related toxic effects (23).

Toxic megacolon is a rare complication of fulminant *C. difficile* colitis (24) which occurs in up to 3% of patients (25), with an associated mortality rate as high as 50% (23). The characteristic features of toxic megacolon are colonic distension and systemic toxicity, which includes raised temperature, hypotension, imbalances of electrolytes and decreased levels of consciousness (26, 27). Patient outcomes may be improved with early diagnosis and aggressive therapy (28).

## Hospital-acquired vs community-acquired disease

Historically CDI was thought to be a hospital-acquired disease, specifically associated with antibiotic use. All antibiotics commonly used are implicated in the risk to develop CDI, although it more often develops after use of clindamycin, fluoroquinolones and broad-spectrum cephalosporins (29). However, recent studies have found that up to 50% of CDI cases are community-associated, often without prior antibiotic use or other predisposing risk factors (30-33). Two recent studies reported asymptomatic carriers in 10 - 15% of the tested population (34, 35). While 40% of children under 1 year are asymptomatic carriers, this decreases to under 5% between 2 and 3 years of age (36). Community-acquired CDI is usually defined as the development of symptoms <48h of hospital admission, without any prior history of hospitalization or antibiotic use (37), whereas hospital-acquired CDI is defined as the onset of symptoms 48h after hospital admission (38-40). However, determining timing of infection is sometimes difficult and it is estimated that many community-acquired CDI cases may go undiagnosed, as it is generally tested at request of the physician in patients with known risk factors (37).

*Clostridioides difficile* transmission occurs via the faeco-oral route, from person to person or instrument to person, with the infectious agent being the dormant spore (18). Spores are intrinsically resistant to commonly used antimicrobial agents (41) and disinfectants (42). They are heat-stable, can survive aerobic environments, and the acidity of the stomach. Resistance to antimicrobial agents is ascribed to metabolic dormancy, as many antimicrobial agents target vegetative and metabolically active cells (43). Spores are problematic in hospitals and have been found to survive for several months on hospital surfaces (39, 44). Hospitalised patients on antimicrobial therapy are therefore at risk, as broad-spectrum antimicrobials kill the normal (protective) microbiota of the gastrointestinal tract, resulting in overgrowth of C. difficile (45, 46).

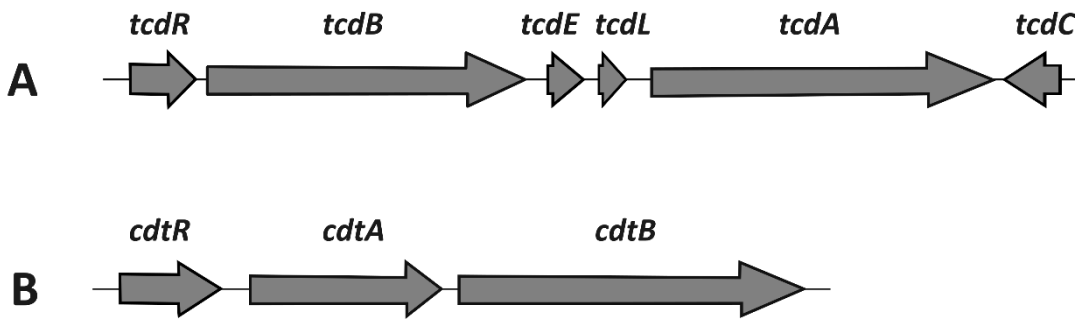



Figure 1. Schematic organization of *C. difficile* pathogenicity islands. A) PaLoc, *tcdB* and *tcdA* encoding large clostridial toxins, *tcdR* encodes a positive regulator, and *tcdC* encodes a negative regulator, *tcdE* encodes a holin-like protein and *tcdL* encodes a remnant endolysin B) CdtLoc, *cdtA* and *cdtB* encodes the binary toxins, *cdtR* encodes a transcription regulator.

## Virulence

Firstly, disease is initiated by spore germination, a critical step in the aetiology of CDI. Spore germination is regulated by primary bile acids, particularly cholic acid and its derivatives, however that alone is not enough, thus amino acid co-germinants are required (47). In the spore coat/outer membrane lies CspC, belonging to the Csp family of subtilisin-like serine proteases, which acts as a unique germinant receptor in C. difficile (48). Following germination, relevant toxins and proteins (Table 1) are produced; these interact with host mucosa to induce diarrhoea (18, 49). The virulence of the organism is mainly driven by the production of toxin A, toxin B and *C. difficile* transferase (also known as CDT or binary toxin) (Table 1, Figure 1).

The genes for toxin A and B are located on the 19.6kb pathogenicity locus (PaLoc) of *C. difficile*, which contains toxin and other virulence genes. The *tcdB* and *tcdA* genes encode for two exotoxins, Toxin A (TcdA) and Toxin B (TcdB), both belonging to the large *Clostridial* toxin family (49, 50).

Situated between *tcdB* and *tcdA* is *tcdE* and *tcdL*. The TcdE/TcdL system originates from the phage holin/endolysin system and has been adapted to assist in toxin transport. The *tcdE* gene encodes a holin-like protein, and *tcdL* encodes an inactive endolysin-like protein. Together, TcdE and TcdL are thought to mediate toxin transport out of the *C. difficile* cell (51).

Transcription of *tcdA* and *tcdB* require the upstream positive regulator, *tcdR*, which encodes an alternative sigma factor (52). The *tcdC* gene, found downstream of *tcdA*, encodes a negative

regulator of *tcdA* and *tcdB* expression. *In vitro*, this regulator is mainly expressed in the exponential phase. Expression decreases towards the stationary phase when toxin production begins (53). Conversely, tcdR expression is upregulated during the late log phase and stationary phase to commence toxin production (53).

Toxins A and B exert cytotoxic effects by interfering with the actin cytoskeleton, consequently causing cell rounding and membrane blebbing (54). This is achieved via a uridine diphosphate (UDP) glucose hydrolase dependent mechanism which transfers a sugar moiety from the toxin to the guanosine triphosphatases (GTPases) 1 Rho, Rac and Cdc42, resulting in their inactivation. GTPases from the Rho family are responsible for many cell processes, but they largely regulate the actin cytoskeleton and their inactivation by TcdA and TcdB results in cell rounding and consequently cell death (54).

Table 1. Virulence factors and known regulators facilitating the development of CDI.

| ***Locus*** | ***Virulence factor*** | ***Characteristic*** | ***Reference*** |
|---|---|---|---|
| ***PaLoc*** | TcdA and TcdB | Protein exotoxins that disrupt the actin cytoskeleton and consequently induce cell rounding and membrane blebbing, resulting in cell death. | (54) |
| | TcdC | Negative regulator of TcdA and TcdB | (52) |
| | TcdE/TcdL | Holin/endolysin system thought to be responsible for the release of toxins out of the bacterial cell | (51) |
| | TcdR | Positive regulator of TcdA and TcdB which encodes an alternative sigma factor. | (52) |
| ***CdtLoc*** | CDT | *Clostridioides difficile* transferase (CDT) or binary toxin | (55) |
| | CdtR | Upstream regulator of *cdt*AB | (56) |

TcdB is more potent than TcdA as the catalytic activity of the UDP-glucose-dependent intracellular reaction in TcdB was found to be approximately five-fold greater than TcdA (57). One study found that holin proteins do not form channels that are consistent in size (58). Thus TcdB (266kDa) is likely to pass through more channels, including the narrower channels, than

[1] GTPase, hydrolyses nucleotide guanosine triphosphate (GTP) into guanosine diphosphate (GDP)

TcdA (308kDa). Together this would accelerate the cytotoxic effect caused by TcdB, hence making it the most potent of the two.

Some *C. difficile* strains (e.g., strains belonging to RT027 and RT078) produce a binary toxin, also known as *C. difficile* transferase (CDT) (56) . CDT consists of two components: the binding component (CDTb) and the enzymatic component (CDTa), which are encoded by *cdtA* and *cdtB* both found in the binary toxin locus (CdtLoc) (Figure 1B) (59, 60). Like Toxin A and B, the *cdtAB* operon expression is under the control of an upstream transcriptional regulator, *cdtR* (59).

**Table 2.** Typing methods used for *C. difficile.*

| Typing method | Description |
|---|---|
| MLST | This technique involves partial sequencing of a set of housekeeping genes. Alleles are assigned a number from a global database, and this will yield a multi-locus allelic profile which is used to assign isolates to the corresponding sequence type (61). |
| PCR ribotyping | The ribosomal 16S-23S rRNA intergenic spacer sequences of organisms are amplified. Differences in the number and size of amplified products are used to classify a species into subtypes, named ribotypes (62). |
| PFGE | Whole genome DNA is digested using a rare-cutting restriction enzyme, which produces large DNA fragments. The unique band patterns, called pulsotypes, are detected using pulsed-field gel electrophoresis (63). |
| REA | This involves digestion of total genomic DNA by the *Hind*III restriction enzyme and using agarose gel electrophoresis to separate DNA fragments into unique band patterns (64). |
| Toxinotyping | Characterisation is done using variations in the PaLoc region using PCR-restriction fragment length polymorphism-based methods (65) |

MLST, multilocus sequence typing; REA, restriction endonuclease analysis; PFGE, pulsed-field gel electrophoresis.

## Typing and Epidemiology of *C. difficile*

*C. difficile* isolates responsible for outbreaks are typed as described in **Table 2**. PCR ribotyping is the most common typing scheme used to describe strains isolated from patients, especially during outbreaks (66, 67). The more prevalent *C. difficile* ribotypes are described in Table 3.

**Table 3.** Prevalent *C. difficile* ribotypes responsible for outbreaks

| Ribotype | Regions detected | Features | Host |
|---|---|---|---|
| **027** | North America, Europe (68-71) | TcdA, TcdB and CDT-positive; *tcd*C has 1bp frameshift deletion and 18bp deletion (68). | Humans (68) |
| **078** | Europe, Asia (72, 73) | TcdA, TcdB and CDT-positive; *tcd*C has 39bp deletion (74) | Humans, animals (74, 75) |
| **017** | Asia, Europe, Africa (10, 76, 77) | TcdA-negative and TcdB-positive (76) | Human (78) |

Strains belonging to the epidemic PCR ribotype (RT) 027 have been associated with more severe disease in some settings. RT 027 strains produce TcdA, TcdB and binary toxin. Some RT 027 strains have been shown to produce 16 times more TcdA and 23 times more TcdB than other non-RT 027 strains *in* vitro (79). Several other RT 027 strains retain potent virulence even with low toxin production indicating the role of other virulence factors (80). The hyper-colonization phenotype observed in low-toxin producing strains suggests that lowering toxin production is a fitness strategy that promotes efficient gut colonization (80). An *in vivo* study by Vitucci *et al*. (81) tested the virulence of epidemic ribotypes (including RT 027 and RT 078) and non-epidemic ribotypes in two animal models, namely murine and hamster models, and they found that the epidemic ribotypes were more virulent than the non-epidemic ribotypes. RT 027 strains have been associated with outbreaks in North America and Europe (68-71). The United States has had a decrease in CDI incidence of 6% over the last decade (82), with RT 027 having 10% prevalence (83). Canada experienced a decrease in RT 027 prevalence from 25 to 9.4%. The current predominant ribotype in North America is RT 106 (83, 84).

In Europe, RT 014 and RT 078 are the most prevalent (85). While RT 014 is also the most prevalent ribotype in Australia (86-90). Strains belonging to RT 014 and 078 have largely been

associated with animal disease, predominantly porcine and bovine animals, as well as in retail meat products (75, 91, 92). Although reports of human disease caused by these strains have been rising, the specific mode of transmission, whether through zoonotic transmission or the food chain, remains unclear (74, 92).

The most frequently isolated RTs in the Middle East are RT 001, 002, 027, and 017 (93). While RT 017 is the dominant ribotype Asia and is endemic to East and Southeast Asia (94), and is the primary ribotype for non-outbreak cases (95). Isolates belonging to RT 017, are TcdA-negative and TcdB-positive, and are occasionally involved in epidemics (96). RT 017 is also the primary RT described in diarrhoeal patients from the Western Cape Province of South Africa (97), and is the most prevalent RT in patients admitted to specialist TB hospitals (98). An Egyptian study has reported that RT 097 and 078 were the most prevalent toxigenic strains, while RT 039 was the prevalent non-toxigenic strain (99). Ribotype data from other African countries are lacking.

### Zoonotic transmission

As stated before, *C. difficile* is typically found in the gastrointestinal tract of humans and other mammals (100-104); and hence is ubiquitous in environments inhabited by them (102). Initial studies of household pets observed that the majority of isolated strains were non-cytotoxigenic, with higher carriage of isolates in animals that had recently undergone antibiotic treatment (103). However, recent work reported known “human” ribotypes and drug resistant isolates in animals (105, 106), which suggests a potential zoonotic risk of animal strains to pet owners.

### Prevention of Hospital Acquired CDI

Intercepting spore transmission with sporicidal agents in the hospital setting is crucial for disease prevention. Spores are resistant to commonly used hard surface disinfectants, as a consequence sporicidal agents must be used to combat spore longevity (107). Spores are able to survive on hospital hard surfaces for up to five months (44), whilst vegetative cells, which are shed alongside spores in the stool, survive up to 15 minutes (108). Transmission occurs via the aforementioned hard surface contamination (109-111) and by the contaminated hands of healthcare personnel (42, 107). Chlorine-based disinfectants have been shown to have efficacious sporicidal activity, and are recommended by some governments (112), however this has been proved to be ineffective when tested under more realistic practices (42). Medical personnel can safely use alcohol-based hand sanitizers before and after handling patients

without the risk of transmitting spores, though handwashing with antimicrobial soaps is still recommended (113).

## CDI Therapy and Antibiotic Resistance

### Recommended treatment

Vancomycin is the preferred antimicrobial agent for treating CDI. Previously metronidazole was used for treatment of mild cases, while vancomycin was reserved for severe cases (114). It has since been shown that vancomycin is a superior treatment option; it reduces the risk of recurrence, substantially reduces 30-day all-cause mortality, and shows overall better rates of clinical success (115-117).

### Newer antibiotics

Fidaxomicin is a novel, narrow-spectrum, Food and Drug Administration (FDA)-approved antibiotic for CDI treatment. It has been shown to be superior to vancomycin with less recurrent episodes 25 days after treatment and causing less disruption to the microbiota (118). This is a breakthrough in antibiotic treatment strategies, owing to the fact that 25% patients are likely to develop a recurrence of CDI (4). This bactericidal agent works by inhibiting RNA polymerase (118). The major disadvantage preventing fidaxomicin use from becoming commonplace is its exorbitant price, which can be 3 times higher than other drugs used to treat CDI (119).

## Non-Conventional Therapy: Probiotics and Faecal Microbiota Transplantation (FMT)

The human gastrointestinal tract is populated by a complex ecosystem of immune cells, epithelial cells, and microorganisms. The resident microbiota plays a vast array of roles in human physiology, one of which is colonisation resistance, which is the phenomenon whereby the resident microbiota inhibits colonisation by invading pathogens (120). The onset of disease in CDI occurs with the disruption of the microbiota, usually as a result of antibiotic use (45).

### Probiotics

The World Health Organization has defined probiotics as, “live microorganisms that, when administered in adequate amounts, confer a health benefit on the host” (121). Regarding treatment of CDI, it is believed that administration of the appropriate probiotics should aid in restoration of the intestinal microbiota, thereby inhibiting *C. difficile* proliferation. However, the certainty of this claim, and thus, the optimal probiotic mixture, is yet to be determined.

Lawrence *et al.* (122), published the results of a placebo-controlled trial, and found that probiotics containing *Lacticaseibacillus rhamnosus* GG (LGG), a strain commonly prescribed for diarrhoeal treatment, has no beneficial effect on the host in preventing recurrence of CDI. Older studies done in 1987 and 1995 have shown that LGG can successfully treat recurrent CDI, however, both were open-label trials with small samples sizes (123, 124). A trial done by Hickson *et al*. (125) shows a probiotic with the most potential for treating CDI. They used probiotic mixtures containing *Lacticaseibacillus casei* Imunitass*, Streptococcus thermophilus* and *Lactobacillus delbrueckii* subsp. *bulgaricus.* No patients receiving probiotic mixtures and 17% of patients in the control group developed CDI (4, 125). Many studies done on probiotics as a CDI treatment method have small sample sizes and are done under conditions which might have biases and may therefore be considered unreliable. Further randomized, double-blind, placebo-controlled trials must be done with larger sample sizes to provide more accurate results.

### Faecal microbiota transplantation

FMT is a relatively new treatment method for CDI. As stated above, the onset of clinical disease is induced by disruption of the native microorganisms of the GIT. Thus, the principal curative factor of FMT lies in restoration of the indigenous microbiota by use of stool transplantation from a healthy donor (126). Fresh or frozen FMT can be administered via colonoscopy, gastric tube, and enema; with the latter being the most convenient and still yielding high cure rates (127, 128). Total clinical success rates very between 80% and 90% (128-130).

There is no consensus on whether FMT should be done on immunocompromised patients. A systematic review by Shogbesan *et al*. reported that, out of 303 immunocompromised patients, two deaths occurred, both of which were post-transplantation patients (131). Success rates of 88% for single FMTs and 93% for multiple FMTs in immunocompromised patients were observed, thus suggesting similar efficacy and safety to that of immunocompetent patients. DeFilipp *et al*. has reported 2 cases of bacteraemias following FMT in immunocompromised patients (132). Both patients acquired strains of extended spectrum beta-lactamase (ESBL)-producing *Escherichia coli* which was isolated from blood cultures. Both were treated with carbapenems, however one patient succumbed to the infection. It was found that both patients received stool capsules from the same donor, who had not been screened for ESBL-producing microorganisms since at the time it was not routinely tested for  (132). Hyun *et al.*

investigated the efficacy and safety of FMT in elderly patients with CDI, which yielded a success rate of 66.7%. However, there were two patient deaths, one of which may have been associated with either CDI or FMT, but the exact cause cannot be definitively determined (133). In a recent systematic review, Marcella *et al*. investigated the adverse effects of FMT from 2000 to 2020 and reported five deaths associated with the procedure (7). They also report that four of the five deaths occurred when FMT was delivered via the upper gastrointestinal tract, but a review by Wang *et al*. reports a higher incidence of severe adverse effects via the lower gastrointestinal tract (6). Further research must be done to identify the safest route of FMT administration and to determine whether FMT is suitable for high-risk patients.

## **PART B.** BACTERIOPHAGES, BACTERIAL PREDATORS POTENTIALLY USEFUL TO TREAT CDI

Phages are viruses that infect bacteria and archaea, and can be lytic or lysogenic, shaping the diversity of microbes in natural environments. Although often described as the natural predators of bacteria, a mutualism between phage and host is not uncommon, as phages may confer benefits to the host, such as virulence factors, or antimicrobial resistance genes, whilst the bacterium provides the machinery for idle replication when in the latent phase. That is until unfavourable conditions arise, and the virus assembles and replicates within the host, whereby mature virions lyse the bacterial cell, resulting in host cell death. Phages are typically highly specific to their hosts, targeting a particular strain or species, although some rare phages are known to infect more than one genus (134, 135).

### History of Phage Therapy

The first phage plaques were observed by Frederick Twort in 1915, but it was Felix d'Herelle who discovered that viruses – which he named bacteriophages – were responsible for the bacterial lysis, presenting as clear spots (lysed bacterial cells) on bacterial lawns. He saw their potential for therapy, and successfully used phages to treat four cases of bacterial dysentery (136). Using phages as a therapy resulted in numerous complications that were not fully understood at the time, including the host innate immune response eliminating phages (137). Scientific understanding of the biological and physiological processes of phages were largely hindered by technological limitations of that era. This resulted in the abandonment of phage trials following the discovery of the antibiotics in the 1940s (138, 139). However, phage therapy continued in some Eastern European countries. Of late, research into phage therapy has

skyrocketed as Western scientists search for alternatives to antibiotics in order to combat antimicrobial drug resistance and the possibility of a return to the "pre-antibiotic era" (**Table 4**).

Table 4. Successful phage-based treatments reported in the literature.

| Infectious disease | Organism | Reference |
|---|---|---|
| Pneumonia | *Pseudomonas aeruginosa* | (140) |
| Osteomyelitis | *Staphylococcus aureus* | (141) |
| Urinary tract infection | *Pseudomonas aeruginosa* | (142) |
| Prosthetic joint infection | *Staphylococcus aureus* | (143) |
| Prosthetic valve endocarditis | *Staphylococcus aureus* | (144) |
| Pneumonia | *Achromobacter xylosoxidans* | (145) |

## An overview of temperate phages

Phages typically have two life cycles, lytic or lysogenic, and while those exhibiting both life cycles are known as temperate phages (139). The lytic infection is a productive infection resulting in the death of the host cell, while the lysogenic infection the phage is integrated into the host genome as a prophage. The accepted definition of a temperate phage is given by its ability to lysogenize (146). However, this definition includes all phages displaying lysogenic cycles, regardless of whether they are released by host cell lysis (lytic infection) or those that are chronically released (chronic infection). As this study's focus is on *Caudoviricetes*, and given that the latter is typically seen in filamentous phages, henceforth the term "temperate" will refer to phages that are released from the host cell via lysis and have the ability to lysogenise (147).

The temperate phage's ability to switch between lysogenic and lytic states are controlled by a gene regulatory circuit, or a genetic switch (148). Unlike the lytic phase, which is a productive infection producing mature virions, the lysogenic state is passive and does not threaten viability of the host (149). In the lysogenic state, the phage, which is now called a prophage, has its genetic material integrated into the host chromosome, or less commonly it may exist as a plasmid within the host (150). When integrated, the prophage replicates passively alongside the host, and it remains stable in this state until triggered by external stressors (150, 151). As seen in *Escherichia coli* lambda phage, the lysogenic state is maintained by the expression of

the CI protein, a structural homologue of the bacterial LexA protein (152), which represses transcription of early lytic genes (153). This state will persist until CI is inactivated, a switch that occurs during the host SOS response induced by DNA damage or other stressors. This thereby triggers the irreversible switch into development of the lytic phage, this process is known as 'induction' (152). Laboratory experiments typically carry out prophage induction using UV irradiation or a chemical mutagen. The best studied temperate phage is the lambda phage of *E. coli*, and the genetic switch between lytic and lysogenic phases in lambda phages is well described. The RecA protein is activated and interacts with the phage CI repressor to undergo self-cleavage (154, 155). Low CI protein levels allows expression of Cro, a regulatory protein that represses the lysogenic promotor $P_{RM}$ from which CI is expressed, thereby inhibiting recovery of CI levels. This results in expression of protein Q, which stimulates transcription of lytic genes (156, 157). As a result, the viral genome is excised from the bacterial chromosome and new viral particles are produced. Endolysins accumulate in the cytosol and holins accumulate in the cell membrane. At a time pre-determined by the holin, the membrane will become permeable to the endolysins, and the cell membrane will rupture, releasing the mature virions (158).

The newly leased free phages will go on to infect new host cells. Once infection is established, the phage must decide whether to remain lytic, or to enter the lysogenic phase. This decision is determined by the levels of CII protein, whereby CII levels exceeding threshold will drive lysogeny, and insufficient levels will shift the phage into the lytic cycle (159).

### Phage Taxonomy

Recent changes of phage taxonomy have been made by the International Committee on Taxonomy of Viruses (ICTV). Previously, all tailed phages belonged to the order *Caudovirales*. The three families of tailed phages included *Siphoviridae*, *Myoviridae*, and *Podoviridae*, and their classification was based on their morphology. However, numerous studies had shown that classification based on morphology was misleading, as members of the same families were considerably diverse and with no shared evolutionary history (160-162). The ICTV thereby abolished the order *Caudovirales* and its families *Siphoviridae*, *Myoviridae*, and *Podoviridae* (163). A binomial naming format is now employed, using a genus and species name. All tailed phages now belong to the class *Caudoviricetes*, and while new orders and families have been established, many subfamilies and genera have not been assigned higher ranks. As of this

ratification, morphology-based identifiers are still acknowledged, but they have been deemed to lack taxonomic significance. While many *C. difficile* phages have been assigned at the levels of genus and species, family and order ranks have not yet been established. *C. difficile* phages relevant to this study are listed in Table 5, with their current names, former name, and former family. Although aware of the changes in taxonomy, historical phage names were used to remain consistent with the literature.

Table 5. Taxonomic changes of *C. difficile* phages relevant to this study.

| Current name | Historical name | Former family* |
|---|---|---|
| *Lubbockvirus CD119* | phiCDHM19 | *Myoviridae* |
| *Sherbrookevirus CDHM11* | phiCDHM11 | *Myoviridae* |
| *Sherbrookevirus CDHM14* | phiCDHM14 | *Myoviridae* |
| *Yongloolinvirus C2* | phiC2 | *Myoviridae* |
| *Yongloolinvirus MMP01* | phiMMP01 | *Myoviridae* |
| *Colneyvirus CD27* | phiCD27 | *Myoviridae* |
| *Colneyvirus MMP02* | phiMMP02 | *Myoviridae* |

*Morphology-based taxa

## Morphology of *C. difficile* Phages

In 1983 Sell *et al*. developed the first typing scheme for *C. difficile* phages when they isolated the first *C. difficile* phages (164). All described phages of *C. difficile* were tailed phages with icosahedral capsids and double-stranded deoxyribonucleic acid (DNA) genomes, and therefore belonged to the now obsolete order *Caudovirales* (currently classified as part of the class *Caudoviricetes*). Myoviruses and siphoviruses were found to be present in *C. difficile* (165-167). Myoviruses have a long, non-flexible contractile tail which can range from 70 to 250nm and capsids ranging from 50 to 73nm in diameter. Siphoviruses have non-contractile tails which are longer on average (168), ranging from 190 to 432nm and capsids between 50 and 64nm in diameter (169, 170).

## Phage specificity

*C. difficile* phages have a narrow host range, and several are suggested to be strain specific (139, 171). Host specificity can be attributed to various factors, such as the presence of specific receptors on the surface of the host cell, also CwpV superinfection exclusion system which blocks phage DNA injection, and antiphage systems such as Clustered Regularly Interspaced

Short Palindromic Repeats (CRISPRs) (172-175). Interestingly, specificity is also determined by spatial and temporal pressures, but this likely because the circulating host strains directly influence the phage population (176). Furthermore, it has been shown that phages of a particular geographical region showed little infectivity when used against hosts from a different region (177, 178). Host specificity is beneficial in CDI, as it limits the damage done to the indigenous microbiota of the gut. However, this specificity implies that knowledge of phages circulating in particular geographical locales is needed in order to select the appropriate phages for use in therapy. Known phages isolated from *C. difficile* are described in Table S3.

Table 6. Geographic distribution of characterised *C. difficile* phages

| Region | Myovirus | Siphovirus | |
|---|---|---|---|
| Thailand | ϕHR24, ϕHN10, ϕHN16-1, ϕHN16-2, ϕHN50 | None | (170) |
| Germany/Canada | None | phiCD211 (phiCDIF129T)* | (179, 180) |
| Singapore | ϕC2, ϕ630-1, ϕ630-2 | None | (181) |
| UK | ϕ CD27; phiCDHM2, phiCDHM3, phiCDHM5, phiCDHM6; ϕ CD1801, ϕ08011, ϕ418, ϕ2301 | None | (182-184) |
| Ireland | None | ϕCD6356, ϕCD6365 | (165) |
| USA | ϕ CD119, phiCDMH1 | None | (174, 185) |
| Australia | ϕCD2, ϕCD5, ϕCD8 | ϕCD6 | (186) |
| Canada | ϕCD481-1, ϕCD481-2; ϕCD505, ϕCD506, ϕCD508; ϕCD24-2, ϕCD526; ϕMMP02, ϕMMP04; ϕCD5, ϕCD52, ϕCD630 | ϕCD111, ϕCD146; ϕCD8-1, ϕCD8-2, ϕCD24, ϕCD38-1, | (166, 187-189) |
| Iraq | CDKM15, CDKM9 | | (190) |
| Costa Rica | None | phiCD5763, phiCD2955 | (191) |
| China | JD032 | None | (192) |
| Poland | phiCDKH01 | None | (193) |

*Identical phage independently isolated and characterised by different researchers

Proponents of phage therapy suggests treatment can be administered as a monophage therapy, or a combination of phages called a phage cocktail. Monophage therapy requires precise identification of the offending pathogen on the strain level and determining what the corresponding phage may be (194). In contrast, phage cocktails consist of various different phages which increases the possibility that the target bacteria will fall within the host range (195). Therefore, phage cocktails might compensate for the limitation conferred by narrow

host ranges. However, all *C. difficile* phages identified so far are lysogenic, which potentially poses a problem due to superinfection exclusion systems. Nevertheless, studies have shown that cocktails using lysogenic phages are effective in batch fermentation and animal models (183, 196).

### A Spanner in The Works: Bacterial Immunity Against Phages

Bacteria have developed a complex immune system to protect against phage infection. The most common anti-phage defence mechanisms include the restriction-modification (RM) system, abortive infection (Abi), and the CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR-associated) systems.

The RM system targets and eliminates invading DNA using two types of enzymes, a restriction endonucleases (REase) and a methyltransferases (MTase) (197). The MTase modifies host DNA sequences, which are identical sequences found on the foreign DNA. The modification of these sequences allows for discrimination between self and non-self DNA by the REases, which will then target and cleave the invading DNA (197). Following this, the RecBCD complex is recruited to digest the foreign DNA starting from the breaks done by the REases (198). RecBCD then interacts with Chi (Crossover hotspot instigator) to initiate the RecA DNA repair process (198).

The Abi system involves a programmed cell death, which is activated when other defence systems fail at eliminating the invading phage (199). Abi is initiated when the RexA sensor protein identifies protein-DNA complexes arising from phage infection. RexA will then stimulate RexB, which induces the formation of ion channels, leading to a rapid loss of membrane potential and depletion of ATP, ultimately resulting in the death of the infected cell (200, 201). This cellular suicide occurs before the invading phages can spread to other bacterial cells (201).

The CRISPR system, comprising the CRISPR and *cas* genes, which are upstream of the leader sequence, confers adaptive immunity against invasion of phages and foreign DNA (202). A CRISPR array contains short repeating palindromic sequences which are separated by variable sequences known as spacers; the latter consists of DNA fragments acquired from previous phage infections (203). Spacers enable bacterial recognition and subsequent destruction of

incoming genetic material from infecting bacteriophages. Exposure to new viruses adds new spacers to the CRISPR array (204). New spacers are added downstream of the leader sequence, thereby arranging the spacers in chronological order (204, 205).

The CRISPR system works in three steps: Adaptation, expression, and interference. Briefly, in the adaptation step, a short viral sequence, called the protospacer, is inserted into the CRISPR locus as a new spacer. In the expression phase, CRISPR genes are transcribed into long precursor CRISPR RNA (pre-crRNA) and following expression of *cas* genes, are processed into mature short crRNAs and subsequently form a crRNA-Cas complex. In the final phase, the crRNA-Cas protein complexes target and destroy the invading genetic material (204, 206).

Phages developed several counterattacks to CRISPR-Cas systems. The simplest counter system is evasion of CRISPR-Cas immunity by mutation of the regions recognized by the CRISPR-Cas system, which includes the complementary region between the crRNA and the protospacer (207, 208). A single nucleotide mutation is enough to evade recognition by CRISPR-Cas immunity (207, 209). Phages also encode their own CRISPR arrays and destroy the genomic regions for host innate immunity (210).

### Notable Observations Regarding Phage Infection

#### *In vivo* phage induction

A study by Meessen-Pinard *et al*. has shown that induction of *C. difficile* phages occurs *in vivo* (189). The researchers initially aimed to isolate free phages from stool samples, but instead discovered spontaneously formed plaques when culturing *C. difficile* isolates from stool. Certain phages were also observed to be more susceptible to induction than others. Interestingly, they found that these clinical *C. difficile* lysogens released a higher phage titre than the laboratory generated counterparts.

#### Bacterial host phenotypic changes after phage infection

Changes in toxin production have been observed in *C. difficile* isolates infected with phages. These changes were first described by Goh *et al.* (211), and although similar phenotypic changes have been reported (188), it is still to be fully understood.

This has important implications for the use of phages as a therapeutic option since phage integration during therapy may lead to unintended changes in strain virulence. Goh *et al.* studied the effects of infecting five toxigenic *C. difficile* strains with four different phages (211).

When TcdA levels were determined, only one lysogen had a two-fold increase during bacterial toxin production, with the remaining lysogens showing no change. One lysogen showed increased *tcdA* transcription, yet there was no detectable change in TcdA production. On the other hand, 66% of lysogens had a significant increase in TcdB production yet there was no change in *tcdB* transcription. Additionally, the authors found that lysogenization of non-toxigenic strains did not change non-toxin producers into toxigenic *C. difficile*. The authors speculated that because most changes were not at the transcriptional level, phage infection might instead function by altering the amount of toxin that was released from the bacterial cell. For example, it is known that *tcdE* encodes a phage holin-like protein (212), and this might assist with toxin release, and this could explain the increased toxin release following lysogenization (211). Holins are thought to form channels that are large enough to allow TcdB (266kDa) to pass through, but not the larger TcdA (308kDa) (211). While the size of the holin channels can vary (158), they could still contribute to the increase of TcdB toxin levels.

However, the above observations were contrasted by a study from Govind *et al*., who found that transcription was downregulated in all PaLoc genes within *C. difficile* isolates lysogenized with ϕCD119 (213). Amongst these changes, they found a 50% reduction in TcdA secretion when compared to parent strains. Lysogenization of ϕCD119 in several other strains produced similar results and thus these changes were not considered strain-specific. The authors found that the phage regulator protein, RepR, was expressed in *C. difficile*-ϕCD119 lysogens. RepR downregulates the promoters of *tcdR* and *tcdA*, and by inhibiting *tcdR*, transcription of *tcdA* and *tcdB* are consequently downregulated. Thus, it was concluded that RepR was responsible for the transcriptional changes in *C. difficile*-ϕCD119 lysogens.

Sekulovic *et al*. showed that the amount of TcdA and TcdB found extracellularly increased by 1.6-fold and 2.1-fold, respectively, upon introduction of ϕCD38-2, a siphovirus (188). This was found in an isolate which embodied all characteristics of a 'hypervirulent' NAP1/027 isolate and contained an additional myovirus. However, these results could only be replicated in one other isolate, which belonged to RT 027. Three other isolates belonging to PCR ribotypes 014, 035 and 027 failed to show a significant increase in extracellular toxin levels. They speculate that the reasons for the increase in toxin production are complex and might involve phage-phage interactions or other mechanisms.

The impact of phage infection on toxin production seems to be dependent on both the phage and host characteristics, and while variable effects have been observed, further research is

necessary to understand factors at play.

## Phage therapy for CDI

When compared with antibiotic therapy or FMT, there are a limited number of reports that describe treatment of CDI using phage therapy (183, 214). Various i*n vitro* assays have shown the use of phages to successfully clear C. *difficile*. One study employed batch fermentation models to evaluate the use of ϕCD27 as a treatment option (215). The researchers demonstrated the efficacy of phage administration for both prophylactic and remedial treatment. A follow-up study used an *in vitro* colon model system to study the potential of ΦCD27 administration as prophylactic treatment for CDI over a 35 day period (216). The researchers found a substantial reduction in *C. difficile* cells, toxin levels, and no adverse effects on the gut microbiota, however, they noted an increase in *Enterobacteriaceae*, *Lactobacillus* spp., *Bacteroides* spp., and aerobes. More recently, Nale *et al*. showed that a combination of phages, known as a phage cocktail, was able to alleviate disease symptoms and reduce bacterial burden of *C. difficile* in hamster models (183). In another study by Nale *et al*., using a Wax Moth model, phage application reduced the bacterial burden and improved longevity (217). It was also found to be successful in prophylaxis against CDI. Furthermore, the researchers reported the prevention of biofilm formation and successful penetration of biofilms in the Wax Moth models (217).

The first animal model using a phage (CD140) to treat CDI was done by Ramesh *et al*. using hamster models (218). Three groups (1a, 1b, and 1c) of six each were challenged with *C. difficile* then administered with CD140 suspension. Group 1a received a single dose, and 1b and 1c were administered phage suspension every 8 h for 48 h and 72 h, respectively. Only one phage-treated hamster died. Interestingly, autopsy revealed that the *C. difficile* isolate it carried was resistant to CD140. This study highlights the efficacy of phage therapy in treating CDI. Both single and multiple doses of treatment were shown to confer protection against CDI. However, there were limitations. Phage therapy was shown to not prevent future infections, as after phage treatment, as all hamsters died when they were administered with clindamycin and challenged again with *C. difficile* two weeks after the initial experiments. This is expected, as the host cells would be eliminated, and continued colonization of phages would not be possible due to lack of hosts. One major drawback of phage therapy is host strain susceptibility, as

resistant *C. difficile* will not be protected against CDI. However, this can be circumvented by using a cocktail of phages to increase the host range. This was done by Nale *et al*. using a batch fermentation model (196). This study investigated the effectiveness of phages for prophylactic and remedial regimens. The prophylaxis group was administered phages 2 h prior to *C. difficile* infection, while the remedial group received phages 5 h after *C. difficile* infection. Both groups successfully cleared *C. difficile* until the experiment conclusion (72 h). This study shows that phage cocktails are effective for prevention and treatment of CDI (196). To date, the literature shows that phage therapy is efficacious in *in vitro* assays and *in vivo* animal models, but human trials are yet to be evaluated.

Studies have shown FMT as a highly effective means to treat CDI. At present, it is unknown which element of the microbiota transplant is responsible for infection resolution, as it may be phages, virus-like particles (VLPs), bacterial debris, antimicrobial compounds, or metabolic products (219). An open-label pilot study by Ott *et al*. (219) investigated the efficacy of using sterile faecal filtrate in treating CDI in five patients. They found rapid resolution of symptoms in 100% of treated patients, even six months – two years post-procedure. Transferring faecal filtrates minimises the risk which comes with standard FMT, thus making it a safer alternative for immunocompromised patients.

As demonstrated by the abovementioned study, live bacteria may not be the sole active agent in FMT, and phage activity may play a role. Since the viromes of CDI patients and healthy individuals are markedly different, it can be said that they display dysbiosis of the intestinal virome (126). In a study by Zuo *et al*., an association was found between phage transfer during FMT and the success of FMT treatment for CDI. Furthermore, they reported that successful FMT treatment was more likely when donors had a greater *Caudovirales*[2] richness than the recipient (126). While the study findings clearly outline changes in the intestinal virome, it is important to note that the presence of phages does not necessarily indicate their role in the success of the treatment, as they could instead serve as a marker of a more resilient bacterial community. Since the researchers purified virus-like particles (VLPs) and not infectious phages, their results more accurately represent the bacteria present, rather than the phage activity. Therefore, the increased diversity of *Caudoviricetes* most likely represent the increased

[2] Currently reclassified as *Caudoviricetes*

diversity of bacteria in the donor stools.

In South Africa, phage therapy is under investigation to treat infections in the veterinary, avian and apian worlds (220-222). To date, there are no peer-reviewed publications regarding the treatment of human infections using phage therapy in South Africa. However, a study by Van Zyl and colleagues detailed phages on the skin of healthy human volunteers from Cape Town (223).

Drug resistant *C. difficile* have been reported in South Africa (97, 98). FMT, as indicated earlier, is risky in persons predisposed to opportunistic infections. South Africa has a large population of people living with HIV, diabetes, and tuberculosis (224, 225). These conditions often require antimicrobial therapy for infections for these immune compromised groups, predisposing them to the development of antibiotic-associated CDI. Therefore, an alternative to FMT is needed, for recurrent cases and in the absence of non-expensive (effective) drugs such as fidaxomicin, or if vancomycin resistance (albeit rare) occurs. Phage therapy could be such an alternative, but more knowledge on the regional prevalence of phages and associated ribotypes must be acquired.

## STUDY RATIONALE

*C. difficile* phages have been described in various countries across the globe, however South African phages have not been described thus far. *C. difficile* phages are known for their narrow host range (186-188, 190) and genetic data indicate that this specificity is further shaped by the bacterial strains present in the immediate geographical region (177, 178). The literature shows that RT 017 strains predominant in the Western Cape and are particularly prevalent in tuberculosis patients treated in TB health care facilities in the region (226). These patients are potential reservoirs of antimicrobial resistant (AMR) strains, increasing the possibility for nosocomial and community infections. Alternative strategies, other than expensive fidaxomicin and potentially harmful FMTs (for immunocompromised individuals), are needed. As such, phage cocktails, from local *C. difficile* isolates, may be useful to eliminate CDI in persons who have drug resistant CDI, but who cannot receive FMT.

The presence of prophages has been described in most sequenced *C. difficile* genomes (196). We, therefore, expect to find prophages in the collection provided. The knowledge gap

regarding phages from South African *C. difficile* isolates will be addressed in this study via molecular characterisation of prophages from a local collection of *C. difficile* isolates. These isolates and their genomes were obtained from a previous study on *C. difficile* in Cape Town diarrhoea patients. The ribotype and genome sequences are available for some of the isolates. This study will thus inform on *C. difficile* phage types present in South Africa, data which is currently lacking completely.

## AIMS AND OBJECTIVES

This study aimed to isolate and characterise bacteriophages from a collection of South African *C. difficile* isolates. To achieve this aim, the objectives were to:

i. Screen for lytic *C. difficile* phages circulating in sewage water sourced from the greater Cape Town area.
ii. Determine the prevalence of prophage sequences in South African *C. difficile* RT 017 strains through a bioinformatic analysis of a set of sequenced genomes from a local *C. difficile* collection, specifically searching for phage DNA signatures in the genomes as well as previously described putative anti-phage mechanisms.
iii. Induce phage lysis from *C. difficile* isolates identified to harbour prophages using media containing a DNA damaging agent.
iv. Isolate and identify phages from plaques using PCR and sequencing.

# CHAPTER 2

# RAW SEWAGE SCREEN FOR LYTIC PHAGES


## SUMMARY

*C. difficile* has been isolated in various natural and manmade environments either directly inhabited by mammals or where mammalian contamination could have occurred. Consequently, the wastewater system is an ideal environment for *C. difficile* to reside. Given the predatory nature of phages on their bacterial host, and their specificity to their hosts, it is likely that an environmental sample with the presence of the host population, would also contain their associated phages. Currently, the literature regarding *C. difficile* phages largely deals with those isolated through induction, and free phages from environmental sources are not commonly reported. Given the difficulty in isolating phages by induction, a reservoir of free phages would be advantageous. This chapter aimed to identify phages with lytic activity in pre-treated wastewater. Double agar overlays and phage spot test assays were performed to identify phage plaques. Conventional PCR was done to confirm presence of phages. While the double agar overlay assay suggested the presence of a plaque, it was not possible to amplify phage holin genes from the plaque material.


## Introduction

Phages are found in all environments across the globe and are considered the most abundant and ubiquitous biological entity, with an estimated population of $10^{30}$ particles in the biosphere (175, 227). They are ubiquitous, and globally the most abundant phages are tailed phages, which belong to the class *Caudovirocetes* (previous taxonomically classified as order *Caudovirales*) (228, 229). Phages can be found in any environmental niche populated by their host, and because of their host predation strategies, any changes in host abundance and diversity will likewise affect the respective phage population (149, 230). This is not limited to geographical environments, as phages form part of the microbiota found in various systems in the human body, including, but not limited to the skin, oral cavity, respiratory system, gastrointestinal tract, and more (231-234). Therefore, provided that the host is present, these environments may serve as reservoirs for phage isolation.

The current literature has only a few reports of successful *C. difficile* phage isolation from the environment, and these include soil, sediment, and environmental estuarine samples (190, 235). Isolation methods typically involve enriching the environmental sample with the host strain and incubating them for an extended period. Although this shows that environmental isolation is possible, it introduces a bias toward phages with activity against the particular strain used in the amplification.

To date, only a limited number of studies have attempted to isolate free *C. difficile* phages from sewage samples. Meessen-Pinard *et al*. tested 30 sewage samples and 59 faecal samples (189). They found six faecal samples contained phages infecting their selected *C. difficile* test strains. Comparative genomic analyses of *C. difficile* isolates and phages from the same faecal samples suggested that these phages were induced from the host bacteria *in vivo* and excreted in the human faeces. Although these temperate phages were free of the host, they were not strictly lytic. The authors did not find free phages in the sewage samples. Others have also attempted to isolate strictly lytic phages from sewage samples and faecal samples, albeit without success (165, 186). Since Meessen-Pinard *et al*. (189) have shown that *C. difficile* phages have been found in stool samples, it is possible that phages may exist free of their host in the sewage system.

Despite the potential role of sewage as a reservoir for *C. difficile* phages, the current literature on their presence in sewage samples is limited (165, 186, 189). To address this knowledge gap, this chapter aimed to isolate locally circulating *C. difficile* phages from sewage samples in the greater Cape Town area. Local South African *C. difficile* clinical isolates were used as indicator strains to detect free phages, which is crucial given that phages typically exhibit a narrow host range that may have evolved to target locally circulating strains. This study contributes to the current body of knowledge on the prevalence and diversity of phages in sewage, and its potential as a reservoir for *C. difficile* phages.

## METHODS AND MATERIALS

### Ethics statement

Ethical approval for this study was obtained from the Human Research Ethics Committee (HREC) at the Faculty of Health Sciences, University of Cape Town, under reference numbers 764/2020. Additionally, authorization was secured to utilize isolates collected from a project with ethics approval number 310/2008.

### Sewage sample collection and processing

Treated wastewater samples from a Cape Town wastewater treatment plant were used. An initial filtration step to remove solid faecal material was done using paper filters followed by a second filtration using 0.8µm membrane filters (MilliporeSigma, Darmstadt, Germany). Samples were then centrifuged (Eppendorf AG, Hamburg, Germany) at 10,000×g for 5 minutes and underwent a final filtration using 0.22µm membrane filters (MilliporeSigma, Darmstadt, Germany). Presumptive phage filtrates were stored at 4°C.

### Bacterial culture media and growth conditions

*C. difficile* isolates were routinely cultured at 37°C in anaerobic jars containing gas-generating sachets (AnaeroPack®, Mitsubishi Gas Chemical, Tokyo, Japan), and cultured in brain heart infusion (Bacto™, Thermo Fisher Scientific, Massachusetts, USA) (BHI) broth containing 0.05% (w/v) L-cysteine hydrochloride (Glentham Life Sciences, Corsham, England) and 0.5% (w/v) yeast extract (Merck, Wadeville, Gauteng, South Africa). BHI was supplemented with 10mM $MgCl_2$ (Merck, Wadeville, Gauteng, South Africa) and 10mM $CaCl_2$ (Merck, Wadeville, Gauteng, South Africa) (BHIS) to increase stability of phages. Agar (Scharlau, Barcelona, Spain) was added to solidify media as needed, with final concentrations of 1.5% and 0.67% (w/v) for normal and soft agar, respectively. Media were pre-warmed and pre-reduced in an anaerobic environment for a minimum of 12 hours before use.

### Detection of phages from sewage filtrate using plaque assays

To identify phage plaques and purify phages from sewage filtrate, spot test assays and double agar overlays were performed using *C. difficile* SA05842529 and SA221002 as indicator strains, both of which were isolated from CDI patients attending a large tertiary hospital in Cape Town, South Africa (97, 98).

### Growth optimization

Growth curves of twenty *C. difficile* isolates were established to determine growth kinetics. Isolates included two RT 017 reference strains, M68 and CD630. Briefly, 200µl BHI broth was inoculated using a single colony and incubated overnight (16 h) in the anaerobic chamber. The following day, 1:10 and 1:20 dilutions were made using 20µl and 10µl culture into 180µl (1:10 dilution) and 190µl (1:20 dilution) fresh BHI media, respectively. Growth was monitored by measuring the optical density at 600nm ($OD_{600}$) at three-minute intervals using a plate reader (Stratus, Cerillo, Charlottesville, USA). For each isolate, growth curves were conducted in duplicate. The mean optical density and standard deviation were then calculated from the duplicate values at each time point.

To determine the optimal growth of bacterial lawns in the spot test assays, three different ODs of the indicator strains were inoculated into the top agar. Briefly, overnight cultures were diluted 1:10 into BHI broth and incubated anaerobically at 37°C for four hours until the culture reached the mid-log growth phase. Thereafter, three aliquots (1ml final volume) were made into fresh BHI at 1:1, 1:2, and 1:3 dilutions. Each aliquot was added to 5ml top BHI agar (final concentration 0.67% (w/v)) and poured onto a bottom agar plate (final concentration (1.5% w/v)). Plates were allowed to solidify for 20 minutes (anaerobically) then inverted and incubated overnight in the anaerobic chamber (Bugbox, Baker Ruskinn), under an atmosphere of nitrogen (85%), carbon dioxide (10%), and hydrogen (5%), at 37°C and 60% relative humidity.

### Spot test assays

Spot test assays were performed using a 1:10 dilution of overnight liquid culture into BHIS broth that was incubated for 2-3 hours. Thereafter, 1ml of liquid culture was added into 5ml BHIS top agar (final concentration 0.67% w/v), and immediately overlaid on a BHI bottom agar plate (final agar concentration 1.5% w/v). Double agar overlays were allowed to set for 10 minutes at room temperature on the bench, then 5µl phage solution was pipetted in the centre

of the plate and allowed to dry for an additional 5 minutes before being incubated overnight at 37°C, anaerobically.

### Double agar overlay plaque assays

Overnight liquid cultures were diluted 1:10 into BHIS broth to a final volume of 10ml and allowed to grow for three hours to reach the early log phase. Thereafter, 50µl of phage filtrate was added, and incubated for an additional 30 – 60 minutes to allow phages to adsorb. Thereafter, 1ml of liquid culture was added into 5ml BHIS soft agar (final concentration 0.67% (w/v)), which had been kept molten at 55°C in a water bath, and immediately overlaid on a BHI bottom agar plate (final agar concentration 1.5% (w/v)). Double agar overlays were allowed to set for 10 - 15 minutes at room temperature on the bench, then incubated overnight at 37° overnight under anaerobic conditions.

### Phage lysate storage media

Following phage double agar overlay and spot test assays, observed plaques were picked with sterile pipette tip, and added to 1ml BHIS liquid media or SM buffer (100mM NaCl, 8mM $MgSO_4$, 50mM Tris-Cl (pH 7.6), 0.001% (w/v) gelatine). The phage plaque agar plugs were crushed and allowed to diffuse out of the media at room temperature overnight (236). To remove unlysed bacterial cells, the solution was centrifuged at 4000×g for 1 minute and filtered through 0.22 µm syringe filter membranes (MilliporeSigma, Darmstadt, Germany).

### Phage DNA purification

Prior to phage DNA extraction, filtered material was treated with DNAse I to remove cell-free host DNA. Digests consisted of 450µl phage filtrate, 50µl 10X DNase I buffer (New England BioLabs, Ipswich, Massachusetts, United States), 1µl DNase I (1U/ml final concentration), and 1µl RNase A (39µg/ml final concentration) (New England BioLabs) and were incubated at 37°C for 1.5h. DNAse I and RNAse A were inactivated with 20µl of 0.5M EDTA (20 mM final concentration) and heated for 10 minutes at 75°C. Thereafter, phage capsid heads were digested using 1.25µl of 20mg/ml proteinase K, incubated at 56°C for 1.5 hours, without shaking. Proteinase K was heat inactivated at 90°C for 10 minutes. The resulting crude phage DNA preparation was stored at 4°C.

### *C. difficile* Genomic DNA extraction

Genomic DNA extraction from cultured cells of *C. difficile* clinical isolate SA221002 was done using the Quick-DNA Fungal/Bacterial Miniprep Kit (Zymo Research, Irvine, California, USA) according to the manufacturer's protocol. Briefly, an overnight liquid culture of *C. difficile* SA221002 was centrifuged on a benchtop microcentrifuge at 10,000 × g for 5 minutes, the supernatant discarded, and the pellet resuspended in 200µl PBS and 750µl Bashing Bead Buffer. Cells were disrupted via mechanical lysis using bead beating (one cycle of 5 minutes on the Vortex Genie2 instrument). The lysed material was centrifuged at 10,000 × g for 1 minute, and 400µl supernatant transferred to a spin column fitted in a collection tube (provided in extraction kit). This was centrifuged at 10,000 × g for 1 minute and 1.2ml genomic lysis buffer was added. Half of this suspension was centrifuged at 10,000 × g for 1 minute at a time. The flow-through was discarded and 200µl pre-wash buffer was added with a clean collection tube and centrifuged at 10,000 × g for 1 minute. To this, 500µl wash buffer was added to the spin column and centrifuged at 10,000 × g for 1 minute. Finally, the column was transferred to a clean 1.5ml microcentrifuge tube and 50µl elution buffer was added and centrifuged at 10,000 × g for 30 seconds.

### Detection of phage holin genes using PCR

To establish a positive control for phage holin PCRs, *in silico* PCR analysis first performed using available genomes, previously published primers and FastPCR [(version 6.8.04) accessed 13 April 2022]. This was done to discern if published holin gene primers amplified the target regions in host strains SA221002 and SA05842529. The *in silico* PCR targeted both the myovirus and siphovirus holin genes of *C. difficile* using the primer pairs described in **Table 7** (169).

Table 7. Primer pairs used in conventional PCR assays.

| Genes | Forward and reverse primer (5' – 3') | Tm (°C) | Size (bp) | Reference |
|---|---|---|---|---|
| Myovirus holin | CDHmyoF: TATACCAGAGCAGTTRCTRA<br>CDHmyoR: CMTCCTTCAAYTGTTTGTAA | 48 | 227 | (169) |
| Siphovirus holin | CDHsiphoF: TTATGCGCTTTGCTRTTYAA<br>CDHsiphoR: MGTTTTCATTGCTCCCATTT | 46 | 150 | (169) |
| 16S rRNA | 27F: AGAGTTTGATCCTGGCTCAG<br>1525R: AAGGAGGTGWTCCARCC | 56 | 1500 | (237) |

Tm, annealing temperature.

Following this, conventional PCR reactions to amplify phage holins were made up using 1X buffer (Green Go Taq Flexi Buffer, Promega, Madison, WI, USA), 3mM $MgCl_2$ (Promega, Madison, WI, USA), 0.25mM dNTPs (New England Biolabs, Ipswich, Massachusetts), 1µM each primer (Table 7), 0.0026U/µl Taq polymerase (GoTaq G2 Flexi, Promega, Madison, WI, USA), and 2µl template DNA, up to a final volume of 25µl. PCRs were performed using the 2720 Thermal Cycler (Applied Biosystems, Waltham, Massachusetts, USA), starting with an initial denaturation of 94°C for 2 minutes. This was followed by 35 cycles of denaturation at 94°C for 45s, annealing for 45s at 48°C for myoviruses and 46°C for siphoviruses, then elongation for 1 minute at 72°C. A final elongation was done for 10 minutes at 72°C, followed by 1 minute of cooling at 4°C. PCR products were run on 2% agarose gels (SeaKem LE Agarose, Lonza, Basel, Switzerland) (stained with ethidium bromide) for 1h at 100V and visualized under UV transillumination (GelDoc XR+ Molecular Imager, Bio-Rad, Hercules, California, USA) using Image Lab (version 6.1.0). A 100 bp DNA ladder (Promega) was used to determine the band size.

To detect residual bacterial DNA in purified phage solution, 16S rRNA PCR was done using published primer pairs (Table 7) (237). PCR reactions of 25µl was prepared with 1X buffer (Green Go Taq Flexi Buffer, Promega, Madison, WI, USA), 2mM $MgCl_2$ (Promega, Madison, WI, USA), 0.2mM dNTPs (New England Biolabs, Ipswich, Massachusetts), 0.2µM each primer, 0.0026U/µl Taq polymerase (GoTaq G2 Flexi, Promega, Madison, WI, USA), and 3µl of template DNA. PCRs were run with an initial denaturation at 94°C for 5 minutes, followed by 35 cycles of 30s denaturation at 94°C, 30s annealing at 56°C, and 45s elongation at 72°C, and a final elongation step of 5 minutes at 72°C. PCR products were run on 2% agarose gel as described above, and a 1kb DNA ladder (New England Biolabs) was used.

## RESULTS

### Growth curve and growth optimisation

A growth curve was established to determine when *C. difficile* isolates reached the mid-log growth phase. Two dilutions of overnight cultures were used (1:10 and 1:20 dilutions). Both dilutions reached mid-log at similar time points of approximately 200 minutes (3.33 hours) (Figure 2). Henceforth, all subsequent experiments requiring cultures to be in the mid-log

phase ($OD_{600}$=0.3) were diluted 1:10 from overnight (16 h) cultures, and then incubated anaerobically for 3 – 4 hours.

### Double agar overlays and spot test assays

To optimise the growth of the indicator strains in double agar overlays, various dilutions of the indicator strains were prepared in the top agar layer. Cultures of indicator strains SA221002 and SA05842529 were prepared, with $OD_{600}$ values of 0.9 and 0.7, respectively. Three dilutions of each were made and used in the top agar of the double agar overlays. Dense growth with uniform distribution in the overlays was observed for all dilutions of both strains. However, 1:3 dilutions showed a subtle decrease in turbidity. Therefore, for all further experiments an initial $OD_{600}$ of 0.3 was used.

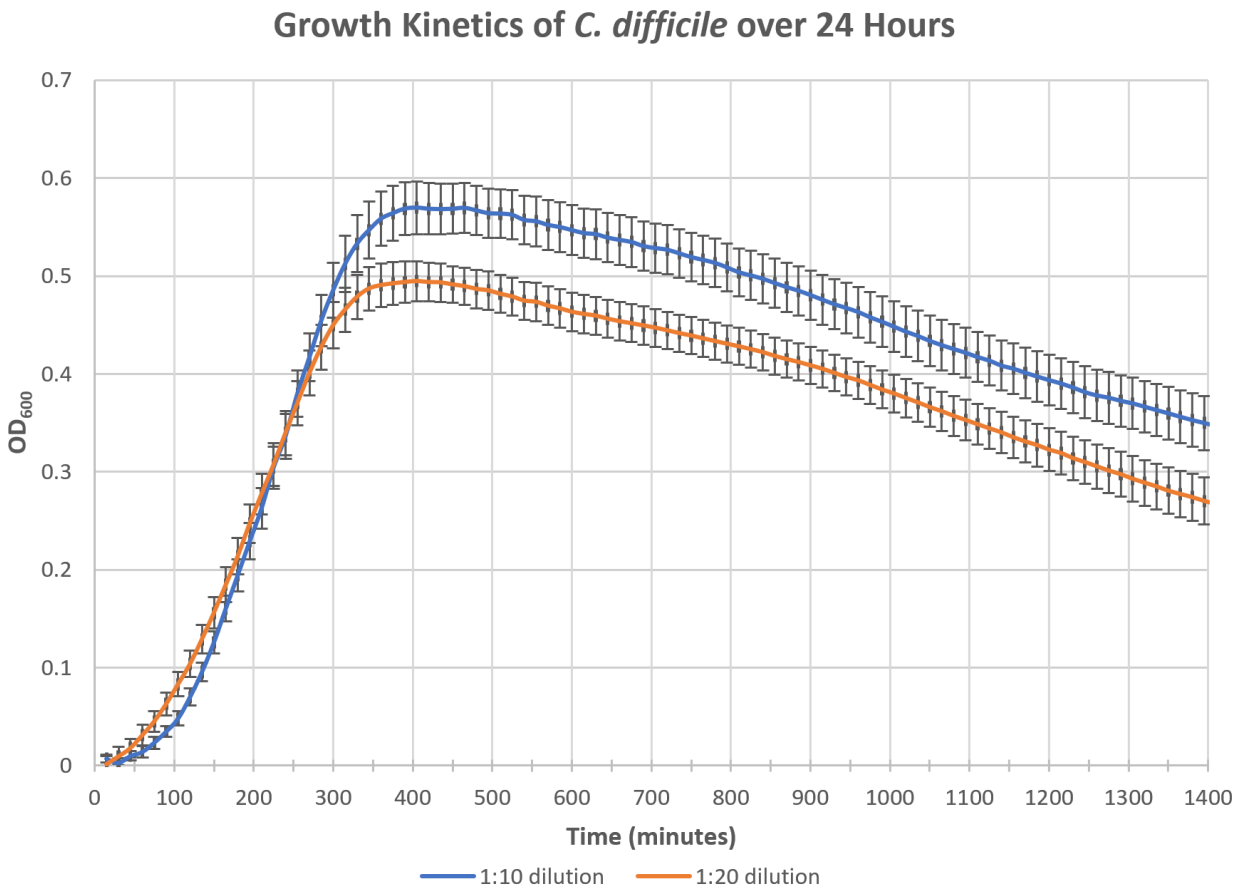



**Figure 2.** Effect of initial dilution on *C. difficile* growth kinetics. Twenty *C. difficile* RT 017 isolates were grown anaerobically for 24 hours. $OD_{600}$ values were recorded at three-minute intervals and average $OD_{600}$ values were plotted. Cultures were grown in duplicate. Blue line represents the 1:10 dilution and orange line represent the 1:20 dilution. Error bars represent standard error of the mean (SEM). Both 1:10 and 1:20 dilutions reached mid-log phase after approximately 200 minutes.

### Spot test assays and double agar overlay assays

To determine if the sewage filtrate contained phages that could infect the test strains, spot test assays were performed. Spot test assays were optimised by pipetting 3, 5, and 10 µl volumes of phage filtrate onto the solidified top agar layer. Despite these optimisations, spot test assays did not produce any plaques.

Double agar overlays were performed, as this allows for phage amplification and adsorption. A single, small (1 – 2 mm), clear plaque was observed on SA05842529 indicator strain (Figure 3). This result was not reproducible and no other double agar overlays yielded plaques.

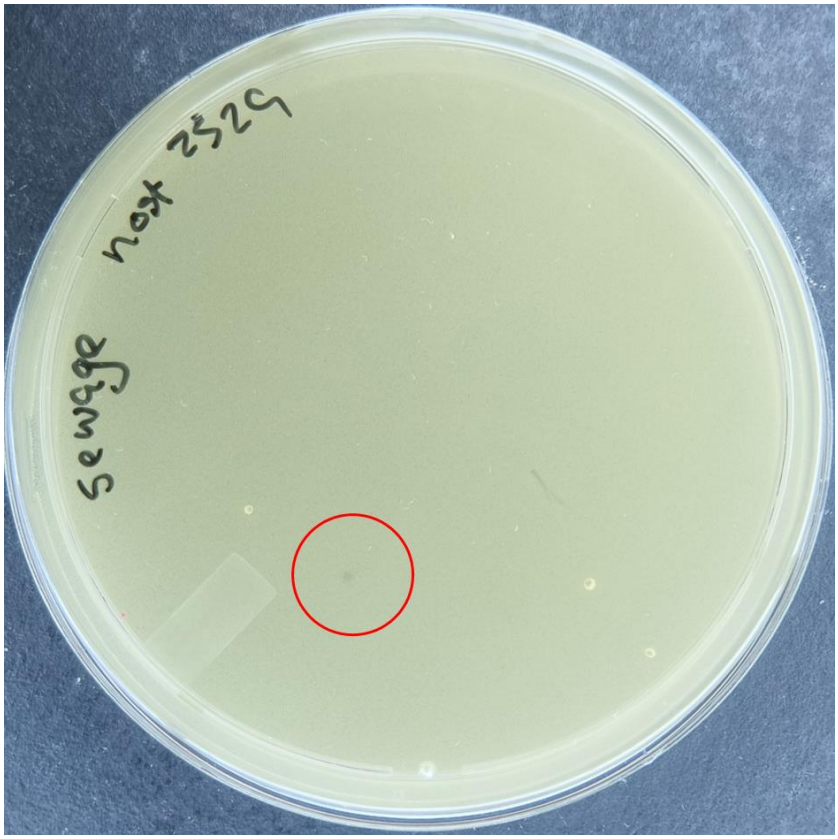



**Figure 3**. A potential plaque (red circle) formed on double agar overlay from sewage filtrate using the SA05842529 host strain.

### Phage holin gene PCR

A crude phage DNA preparation from the single identified plaque was carried out. To confirm the absence of bacterial DNA in this preparation, the solution was used as template in a 16S rRNA PCR amplification. As expected, no product was obtained for the crude phage DNA preparation (Figure 4).

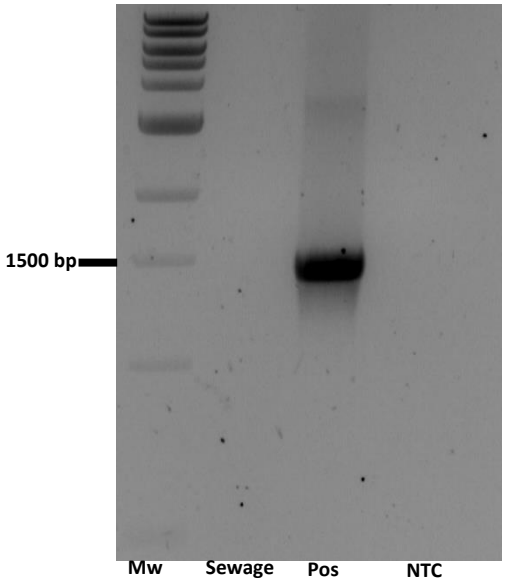



Figure 4. 16S PCR of a sewage-derived plaque. No bacterial amplification was observed. Mw, 1kb DNA ladder; Pos, positive control; NTC, non-template control.

To generate positive controls for the holin PCRs, the SA05842529 and SA221002 indicator strains underwent PCR using the myovirus and siphovirus primer pairs specific to the respective holin genes. Both strains yielded a product for the myovirus holin PCR (Figure 5), but unfortunately, it was not possible to amplify a siphovirus holin product from either strain. Subsequent experiments used SA221002 as the positive control for the myovirus holin PCR.

To determine whether the putative plaque formed on the SA05842529 lawn contained phage DNA, the crude phage DNA preparation described above was subjected to PCR amplification using both holin primer pairs (Figure 6). Neither primer pair yielded a product despite several PCR attempts.

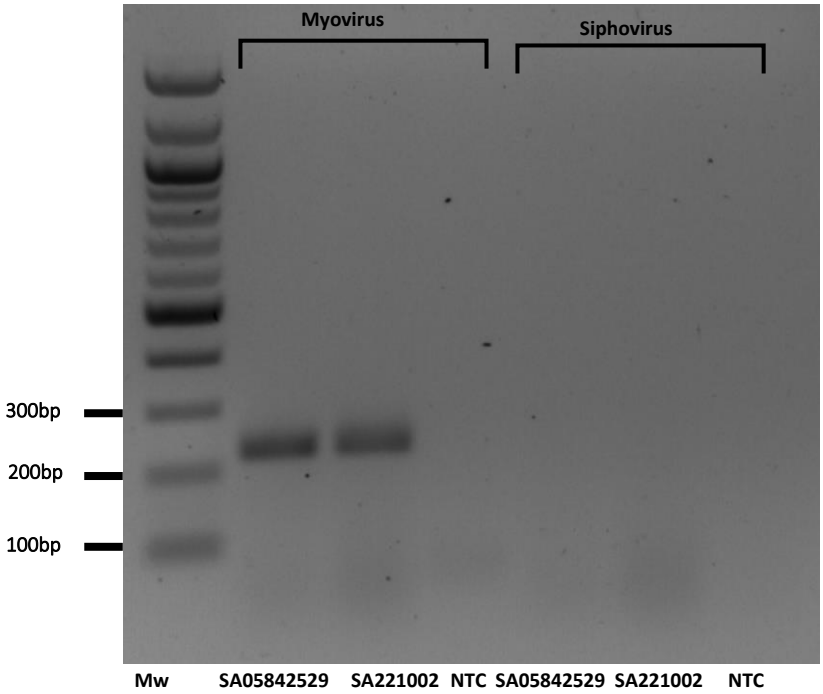



Figure 5. Phage-specific holin gene PCR of SA05842529 and SA221002. (Left) Myovirus holin PCR amplified a band of 227bp on both host strains; (Right) Siphovirus holin was not amplified from either strain. Mw, molecular weight marker (100bp); NTC, non-template control.

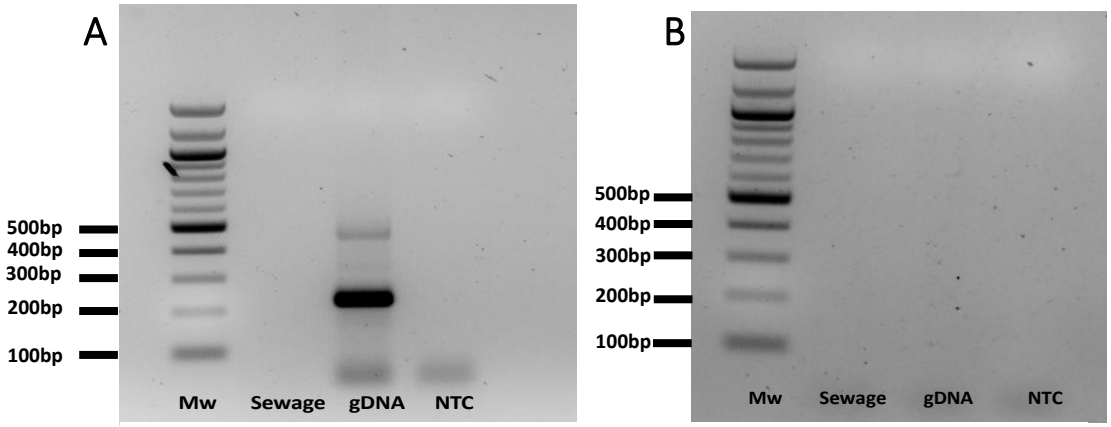



Figure 6. PCR of holin gene regions of myoviruses and siphoviruses for sewage filtrates. A) No amplification for myovirus in sewage, the genomic DNA of test strain SA221002 shows band of expected size B) No amplification for siphovirus PCR. Mw, 100bp molecular weight marker; gDNA, genomic DNA of SA221002; NTC non-template control.

## Discussion

The recovery of *C. difficile* phages from environmental sources, such as wastewater systems, could establish a reservoir of phages. It could potentially yield virulent phages, which would be highly beneficial for phage therapy as virulent phages directly lyse their host cells, thus avoiding the potential of integration into the host genome. Additionally, this would increase the likelihood of obtaining a diverse range of phages from the host pool. Given that phages must co-exist with their host, any phages isolated from environmental sources are likely to exhibit host specificity ranges targeting circulating isolates. Therefore, *C. difficile* hosts isolated from patients in Cape Town and representing the predominant strain types present in the region (226) were used to screen for phages within the sewage samples.

Double agar overlay and spot test assays of sewage phage filtrates were largely unsuccessful. Nevertheless, a potential plaque was observed in the double agar overlay assay, and experiments were done to assess whether this was a true plaque, or simply an artefact in the agar. PCR amplification of the plaque was attempted using previously published primers to conserved holin sequences. Amplification was unsuccessful, potentially due to low DNA concentration within the plaque or DNA degradation during the crude DNA purification process. Since repeat experiments failed to produce additional plaques, the most likely possibility is that the zone of clearing was an artefact. Nevertheless, it is possible that the plaque was caused by a novel phage that harboured a divergent holin gene lacking the conserved primer binding sites. However, given the lack of a positive control for at least one of the holin primer pairs (siphovirus), further experiments will be needed to confirm this. Further investigations could use rolling circle amplification on extracted phage DNA from the plaque, followed by next generation sequencing to detect novel phage sequences.

The low salt concentration used in plaque assays could potentially have impacted the ability of phages to form plaques. The addition of $MgCl_2$ and $CaCl_2$ enhance the stability and attachment of phages (238). This study supplemented BHI with 10mM $MgCl_2$ and 10mM $CaCl_2$, which is typically only used for liquid media. Plaque formation could be improved by increasing the $MgCl_2$ concentration in the soft agar to 1M (239).

An alternative explanation for the failure to produce a PCR product could be that the plaque resulted from the activity of a phage tail-like particle (PTLPs), a class of bacteriocin. These are lethal, non-infectious, host-produced particles which structurally resemble a phage tail (240). Upon attachment to receptors on the target cell, rapid membrane depolarization occurs, resulting in cell death (241). Given that bacteriocins induce their lethal effects without causing the host cell to lyse, they would not require holin or endolysins. A study by Gebhart *et al*. (242) has annotated the open reading frames of *C. difficile* bacteriocins and noted the absence of holin and endolysin genes. This could explain why the PCR did not produce a positive result for the phage holin gene. Transmission electron microscopy (TEM) would provide evidence whether the plaque contained PTLPs, however, this was beyond the scope of the chapter. Therefore, it is possible that the plaque was produced by the activity of a PTLP. Future research could explore the possibility of PTLP activity by TEM.

The lack of lytic phages with activity against the two *C. difficile* strains in the tested sewage samples likely reflects low overall abundances in the environment. *C. difficile* is an obligate anaerobe that sporulates for survival in unfavourable conditions. When sporulated, the bacterium enters a state of metabolic dormancy which enables aerotolerance (43). The metabolically dormant state of the spore poses a challenge for strictly lytic phages as they require a metabolically active host for replication. Instead, a temperate lifecycle would provide an evolutionary advantage, as they can integrate and remain stable whether the host is in the sporulated or vegetative state (150, 243). Furthermore, higher concentrations of phage particles are exhibited in activated sludge liquor compared to raw sewage (244), thereby emphasizing the importance of an active replicating host in fostering phage abundance**.** Furthermore, lytic phages may not be able to attach to their hosts in their spore form, as sporulation causes changes to the exterior surface of the host, and the spore coat and exosporium may conceal surface attachment molecules, such as for attachment pili and S-layer protein A (245-247). The ability of *C. difficile* to form spores could potentially lead to a selection preference for temperate phages over strictly lytic phages. This, combined with the likelihood that *C. difficile* found in the environment will be in its spore form, may explain the difficulty in isolating free phages from wastewater samples.

*C. difficile* phage titres have been reported to rapidly decline, even within optimal storage media and buffer (149). Consequently, the duration of phages circulating within the sewage

system prior to sample may impact phage survival. However, it is not known how long the phages were circulating in the sewage system before the sewage samples were collected. Furthermore, factors like exposure to UV light, temperature and pH fluctuations, and the presence of pharmaceuticals, personal care products and household chemicals may exert adverse influences on phage abundance (248-252).

The sewage sample had undergone preliminary treatments and was not in its raw state. This was a precautionary measure taken to limit exposure to SARS-Cov-2(253), as the sample was collected during the CoVID-19 pandemic. Sewage treatment is likely to affect phage survival, and future studies should be done on raw sewage.

The use of two indicator strains was a limiting factor in this study. As the phages present in the sewage samples may have had a narrow host range, increasing the number of local indicator strains could provide a broader range of potential hosts. While such phages may not be ideal for monophage therapy, they can be evolved to broaden their host range. Additionally, the indicator strains may have had integrated phages providing superinfection immunity or active anti-phage defence systems inhibiting successful phage infection. Whole genome sequences of *C. difficile* isolates can be screened for the presence of integrated phages and defence mechanisms using bioinformatic tools. Furthermore, plaque formation may have been inhibited by the low salt concentration of top agar (239).

## CONCLUSION

The attempt to isolate *C. difficile* phages from treated sewage samples was largely unsuccessful and this is consistent with previous failed attempts by other authors. The absence of virulent phages may be partially explained by the spore-forming nature of *C. difficile* and the selection of temperate phages over strictly lytic phages. Therefore, we decided to examine the presence of integrated prophages within the genomes of a collection of *C. difficile* strains that were isolated from patients with CDI attending hospitals within the greater Cape Town area and the results of these analyses are presented in the next chapter.

# CHAPTER 3: BIOINFORMATIC PROPHAGE SEARCH

## SUMMARY

Many challenges are faced when inducing and isolating phages from their hosts. However, this task has become easier with the rise of metagenomics, as many tools have become available for identification of phage genomes from sequence data. The PHASTER webserver was used to screen the genomes of 58 *C. difficile* clinical isolates for prophages and the qualities of predicted genomes were assessed using CheckV. Prophage carriage was predicted for 54 isolates, with a total of 98 prophage genomes detected. However, CheckV showed that only 5 were high-quality complete genomes, while the remaining isolates were classified as genome-fragments. Distance trees based on Mash distances between phage genomes identified groups of similar phages and the functional modules for representatives from each group were investigated further by alignment with the closest reference phage identified by PHASTER. Finally, Defense-Finder was used to identify antiphage resistance mechanisms across the isolates (including M68 and CF5), and identified several systems, including CRISPR-Cas types I-B and VI-A, abortive infection systems and restriction modification systems, some of which have not previously been described in *C. difficile*.

## INTRODUCTION

The methods for isolating phages in the laboratory have remained mostly unchanged since their establishment in the 1900s, but the current era provides us with culture-independent approaches for identifying phages. Next generation sequencing (NGS) technology has made it possible to efficiently gather large amounts of genome data from a diverse range of samples (254). As a result, it is now possible to mine bacterial genomes and environmental metagenomes for evidence of prophage sequences. This has contributed enormously to the discovery of novel phages, resulting in a rapid expansion in the number of complete phage genomes in public databases (255).

Today there exist several tools dedicated to the identification of phages such as PHASTER (256), VirSorter (257), VirSorter2 (in early-access at the time this study was initiated) (258), PhiSpy (259), VirFinder (260), and Prophet (261). There are differences in the approaches used by these tools with some relying on genomic feature comparisons to reference databases, whereas others utilise machine-learning based methods to identify potentially novel phages.

In this chapter, the PHASTER (Phage Search Tool – Enhanced Release) webserver was used for phage identification (256). The webserver offers a user-friendly interface, and the upgrade from its predecessor, PHAST (262), has enabled rapid and accurate identification of phages, making it a popular choice for preliminary screening of bacterial genomes. When given raw sequence data, PHASTER uses methods built into its pipeline to employ rapid ORF prediction and translation, and genome annotation. PHASTER has a custom, self-updating database of phage protein sequences collected from the National Center for Biotechnology Information (NCBI) and the phage database established by Srividhya *et al*. (263). These databases are used to perform BLAST searches for protein identification and phage sequence identification. The pipeline uses tRNAscan-SE (264) and ARAGORN (265) to identify tRNA and tmRNA sites and extract information for attachment sites. PHASTER assigns a 'completeness' score to identified phage regions based on the proportion of phage genes and/or matches to known reference phage genomes. The scores are graded as intact (score > 90), questionable (score 60 - 90), or incomplete (less than 60). PHASTER also outputs the phages in the reference database that have the highest number of genes with similarity to the identified region.

To date, no phages with activity against South African *C. difficile* strains have been isolated or characterized. RT 017 has been reported as the predominant ribotype circulating in the Western Cape region of South Africa (98). M68 and CF5 are well described RT 017 reference strains, and there is evidence that the genomes of these isolates harbour prophage regions, but there is no evidence of any of these being inducible (266). There is a scarcity of information regarding *C. difficile* phages in South Africa. Therefore, a bioinformatic analysis was done to identify potential phages from the genomes of South African *C. difficile* strains.

## METHODS AND MATERIALS

### *C. difficile* WGS

A total of 58 draft genome assemblies for *C. difficile* isolates belonging to a collection of South African clinical samples were provided by B. Kullin (98). Of these, ten were assembled from data generated in a previously published study (267) [ENA Accession numbers: ERR340646-ERR340655] and the remaining 48 from data generated by Prof Thomas Riley and Dr Daniel Knight.

### *C. difficile* genome evaluation and quality control

Overall *C. difficile* genome assembly metrics were assessed using QUAST (Quality Assessment Tool for Genome Assemblies) (version 5.1.0rc1) (accessed 26 April 2022) (268). Genome completeness and contamination were estimated using CheckM (version 1.2.1) (269). To rule out assembly contamination with phiX174 sequences prior to any analyses, a BLAST (Basic Local Alignment Search Tool) analysis (270) was carried out using the phiX174 genome (NCBI accession number NC_001422.1).

### Prophage search of sequenced *C. difficile* isolates

The PHASTER webserver was used to determine the presence of putative prophages in the genomes of sequenced *C. difficile* strains (accessed 23 March 2021) (256, 262). Genome sequence data in FASTA format were used as input to the web server (https://phaster.ca, accessed 20 May 2022) and the option for multiple separate contigs was selected. *C. difficile* RT 017 strains CF5 and M68 (GenBank accession numbers NC_017173.1 and NC_017175.1, respectively) were included in the analysis.

All 58 *C. difficile* genomes were screened for prophages using VirSorter2 (258). Parameters were set to identify double-stranded DNA prophages with a minimal length of 5000 bp. The resulting sequences underwent quality control and host region trimming using CheckV. These trimmed sequences were then re-analysed with VirSorter2, and then annotated with DRAMv (version 2.2.4) for manual curation (271).

### Assessing the quality of prophage genomes

A quality control assessment was done on the 98 intact prophage genomes identified by PHASTER. CheckV (version 1.01) was used to assess the genome quality by estimating completeness and contamination scores of phages detected by PHASTER (272).

### Prophage and *C. difficile* genome analysis

Minimum hash distances between predicted phage genomes was determined using Mash (version 2.3) (273) and the resulting distance matrix visualised by hierarchal clustering. The tree was visualized using Interactive Tree of Life viewer v6 [date accessed: 5 February 2023] (274). Phage genomes clustering within groups were assigned a group number (Group 1 – 6). Representative phages were selected, and the genomes were aligned and mapped according to a reference phage. Phage genome organization maps were constructed using CLINKER

(version 0.0.27) (275). A conserved domain search (NCBI, accessed 19 February 2023) analysis was done to predict the function of genes that showed no similarity to the reference phage genome.

Core genome phylogeny was done on RT 017 isolates. Genes belonging to the core genome of RT 017 were identified using Panaroo (version 1.3.0) (276) and aligned using MAFFT (version 7.510) (277). To construct a maximum likelihood tree, the best fit models for each gene partition in the alignments was determined using ModelTest-Ng (278) and used to generate bootstrapped maximum likelihood trees (n = 1000) using IQ-Tree2 (version 2.2.0) (279).

Known anti-phage systems were screened for in all *C. difficile* genomes (including the reference M68 and CF5 genomes) using Defense-Finder (version 1.0.9) (280, 281).

### Statistical analyses

An Independent-Samples Kruskal-Wallis test was done to compare the genome sizes between the three completeness categories (intact, questionable, incomplete) and pairwise comparisons of completeness were done and adjusted using Bonferroni correction. All statistic calculations were done using IBM SPSS Statistics (version 28.0.1.1 (15)).

## RESULTS

### *C. difficile* genome assembly quality control

High levels of genome fragmentation and/or assembly incompleteness and contamination can affect the ability of phage prediction algorithms to detect prophages. Therefore, the qualities of *C. difficile* genome assemblies were assessed using QUAST and CheckM. Overall, genome assemblies showed a relatively low level of fragmentation (median N50 and L50 values of 183,615 bp and 8, respectively), with a median of 81 contigs of 500bp or larger (Table 8). All assemblies were classified by CheckM as having high completeness (≥90%) and low contamination (≤5%). No phiX174 contaminating sequence was detected in any of the assemblies. Thus, *C. difficile* draft genomes used in this study were of high quality.

Table 8. Metrics summary for 58 *C. difficile* genome assemblies

| Metric | Median (Range) |
|---|---|
| Genome size (bp) | 4,343,441 (4,118,749 - 4,403,195) |
| Number of contigs | 81 (48 - 349) |
| GC% | 28.66 (28.42 - 28.79) |
| N50 | 183,615 (26,385 - 239,860) |
| L50 | 8 (5 - 53) |
| Completeness (%) | 98.96 (97.68 - 98.96) |
| Contamination (%) | 0.6 (0.38 – 3.84) |

### PHASTER Screen of clinical *C. difficile* collection

The PHASTER webserver was used to identify putative prophages in the genomes of *C. difficile* isolates (**Table S1**). Almost all isolates (54/58, 93.1%) were predicted to harbour intact prophages. PHASTER did not detect prophages in isolates SA221002, SA231012, SA242006 and SA514119. Overall, a total of 98 intact, 75 questionable, and 456 incomplete prophage regions were identified (**Figure 7**). Most isolates were predicted to harbour either one (22/58, 37.93% of isolates) or two intact prophages (23/58, 39.66% of isolates) per genome with a median of two intact prophages per host. A further 12.07% of isolates (7/58) were predicted to harbour three intact prophages in their genomes, while the final two isolates were predicted to harbour four and five prophages, respectively.

PHASTER infers the most closely related phage in the reference database based on the highest number of gene hits. The most identified closest relative was phiCDHM19 (68/98, 69.39% of predicted intact prophages) and every time phiCDHM19 was detected, it was consistently the highest hit. Other commonly found best matching reference phages were phiMMP02 (14/98, 14.29% of predicted intact prophages) and phiMMP01 (7/98, 7.14% of predicted intact prophages), the latter is also present in the *C. difficile* M68 genome. The remaining predicted intact prophages were predicted to most closely match phiCDHM11 (5/98, 5.10%), phiCDHM14 (3/98, 3.10%) and phiCD27 (1/98, 1.02%) reference phages. *C. difficile* CF5 had one intact phage which shared the highest number of genes with the phiMMP03 reference phage.

PHASTER predicted genomes between 14 and 73kb, with a median of 27.2kb (Figure 8). Published *C. difficile* phage genome sizes typically range between 30 and 60kb in length (282) (Table S3). An independent samples Kruskal-Wallis test and pairwise comparisons were done to compare the phage genome sizes between the three completeness categories. Overall, there was a statistically significant difference between the genome sizes in the completeness categories ($p < 0.0001$, adjusted using Bonferroni correction for multiple tests).

### VirSorter2 comparison

VirSorter2 was used as a secondary tool used to confirm the presence of prophages in *C. difficile* genomes. Out of the 58 genomes screened, a total of 2078 prophage genomes were detected. All prophage genomes detected by PHASTER were also detected by VirSorter2, therefore, VirSorter2 supported PHASTER's results.

Table 9. Genome metrics of high-quality phages

| Phage | Length (bp) | Viral genes | Completeness (%) | Contamination (%) |
|---|---|---|---|---|
| SA05842529_phage1 | 55107 | 41 | 99.63 | 0 |
| SA06148880_phage1 | 57742 | 38 | 97.86 | 6.34 |
| SA122025_phage1 | 73413 | 42 | 100 | 20.18 |
| SA122025_phage2 | 66276 | 41 | 100 | 13.48 |
| SA424047_phage1 | 57981 | 39 | 98.29 | 6.32 |

### CheckV analysis

To estimate 'completeness' of the predicted phage genomes, intact prophage genomes detected by PHASTER were subjected to a viral genome quality assessment using CheckV, which assigns completeness and contamination scores and MIUViG classifications. Most prophage genomes (68/98, 69.39%) were classed as medium-quality genomes, predicted to be 50-90% complete. Only 5/98 (5.10%) (Table 9) of the predicted phage genomes were predicted to represent high-quality genomes (≥90% complete), while the remaining phage genomes (25/98, 25.51%) were classed as low-quality genomes (<50 complete). Of the 5 high-quality viral genomes, 4 had highest similarity to the reference phage phiMMP01, while SA122025 had highest similarity to phiCDHM19. Two high quality phages came from the same host (SA112025). Two high-quality genomes had 100% completeness scores, and the remaining 3 scored 97.86%, 98.29%, and 99.63%.

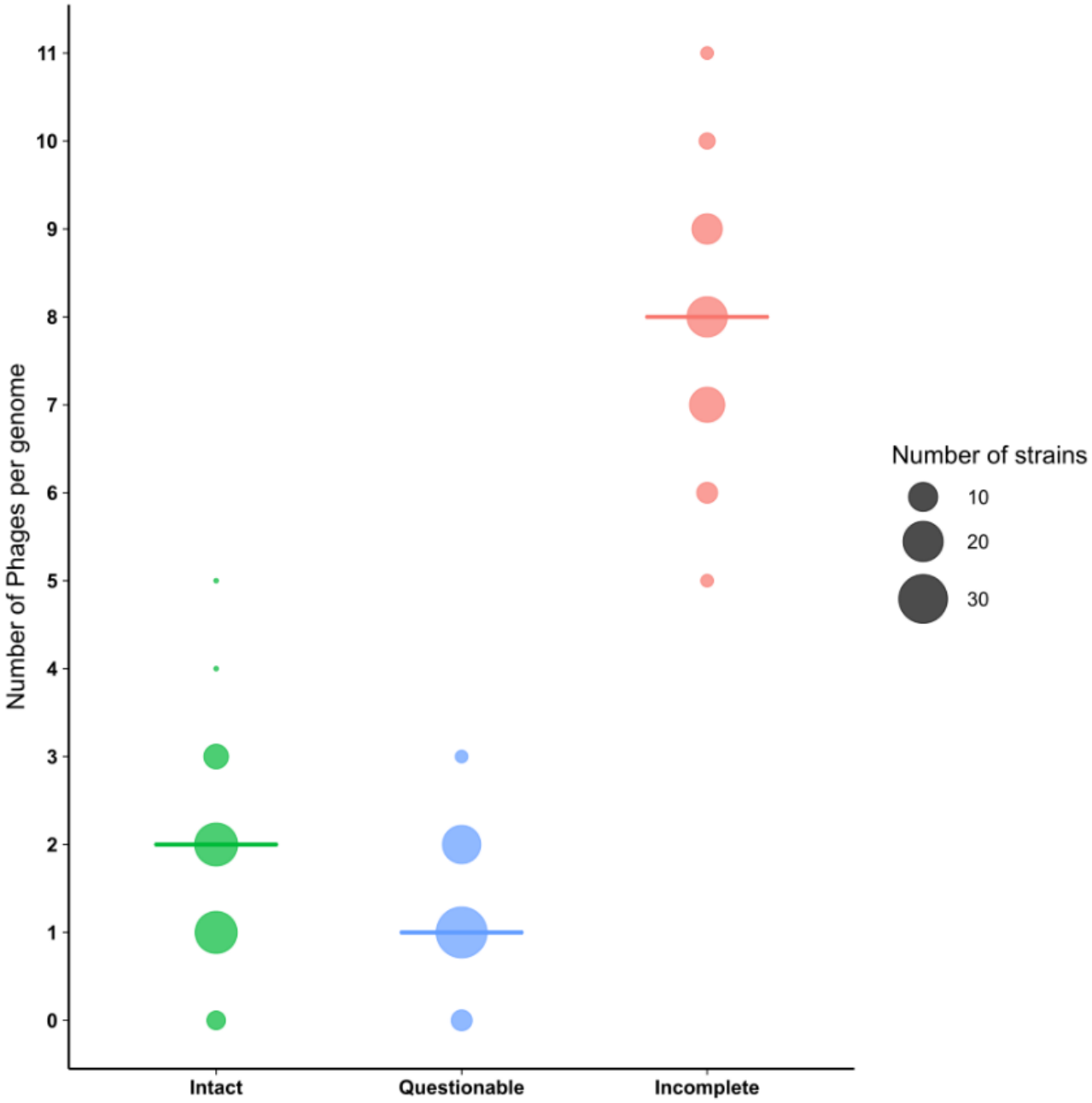



Figure 7. Number of intact, questionable and incomplete prophages per *C. difficile* genome. Bubbles represent PHASTER's estimate of the number of phages per genome. Horizontal bars indicate the median *C. difficile* isolates per host genome.

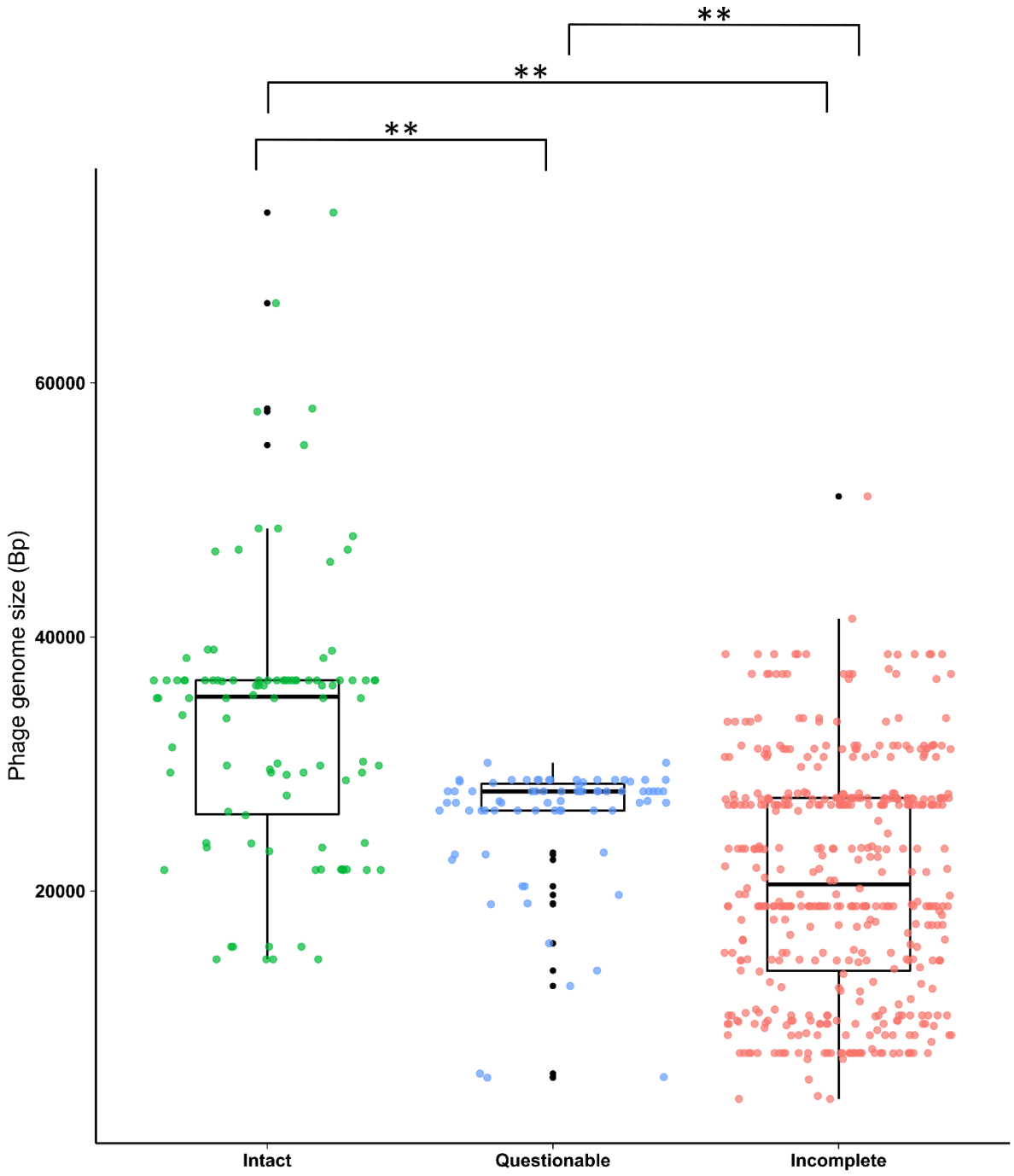



Figure 8. Genome size of 629 putative prophages regions identified in *C. difficile* genomes. Black points indicate the outliers. Upper and lower horizontal bars represent the 25$^{th}$ and 75$^{th}$ interquartile range; thick horizontal bars represent the median. Tips of vertical bars indicate the 10th and 90th percentile. Independent-Samples Kruskal Wallis test was done and pairwise comparisons between genome sizes and completeness categories, followed by Bonferroni correction where ** represent p-value of <0.0001.

## Phage genome organisation

Six phage groups (groups 1 – 6) were identified (**Figure 9**) and features of the genomes were mapped (**Figure S1**) based on genome annotations. Individual phages for each group are indicated in **Figure S2.**

Group 1 (**Figure S1 (A)**) consisted of a single phage entity (SA111045_phage1) which displayed some similarity to phiCDHM19. This group exhibited distinct functional modules pertaining to tail morphogenesis and lysogeny control. However, several open reading frames (ORFs) encoded proteins of unknown function. A conserved domain search was able to predict functions of numerous ORFs belonging to modules associated with DNA replication, DNA packing and capsid morphogenesis.

Group 2 phages (**Figure S1(B)**) consisted of three phages and were most closely related to phiMMP01 (NC_028883) by Mash analysis. Each predicted prophage exhibited various functional modules, including DNA packaging, capsid morphogenesis, tail morphogenesis, and lysis modules. Notably, the tail morphogenesis module contained five additional genes compared to the reference. A conserved domain search identified two of these genes to have functional domains associated with the terminase large subunit and the terminase ATPase subunit. Despite the absence of similarity between the putative phage genomes and the reference phage's DNA replication region, conserved domain searches successfully predicted the functional roles of several proteins involved in lysogeny, lysis, capsid morphogenesis, and DNA packaging. Furthermore, an abortive infection (Abi)-like protein was identified across all phages within this group. SA122025_phage2 and SA05842529_phage1 were identified within the middle portion of large contigs.

Group 3 (**Figure S1(C)**) consisted of three similar phages that were most closely related to phiMMP02 (NC_019421). DNA packaging and tail morphogenesis modules were largely similar to phiMMP02, except for an additional gene inserted into the tail morphogenesis module of all three predicted prophages. The functional role of this gene could not be determined. The putative phage capsid morphogenesis modules did not show similarity to the reference. Similarly, the DNA packaging module in the putative phages did not match the reference, but it did contain genes with predicted large terminase subunit domains. The phage genomes

lacked modules for lysis, lysogeny control, and majority of the DNA replication module, but the genomes were located at the end of contigs.

Group 4 (Figure S1(D)) had three similar phages with homology to phiCDHM11 (NC_029001). The DNA packing, capsid morphogenesis, and tail morphogenesis functional modules in the putative phages were highly similar to the reference phiCDHM11 phage, however the lysis and DNA replication modules were absent. While the genome of SA122111_phage2 spanned the entire contig length, SA231004_phage3 and SA122111_phage3 initiated at contig positions 188 bp and 153 bp, respectively, and extended until the respective contig termini. This observation raises the possibility that modules could have been overlooked at either extremity of the contig for all three phages.

Group 5 (Figure S1(E)) consisted of 10 phages, all of which were most closely related to phiCDHM19 (NC_028996). Predicted functional modules for DNA packing, capsid morphogenesis, and tail morphogenesis were present in all predicted prophages. However, in all 10 phages, many tail morphogenesis ORFs were absent, and it must be noted that this module was located at the end of the contig. Furthermore, lysis, lysogenic conversion, lysogeny control, and DNA replication modules were not present in all genomes. A conserved domain search predicted the function of several genes involved lysogeny control of SA122025_phage1. However, these genes were scattered and did not seem to form a distinct module. Furthermore, this module was not observed in the remaining nine isolates. Conserved domain searches predicted the functions of ORFs involved in DNA replication, and these modular regions were identified on the genomes of SA122025_phage1, SA714001_phage1, SA211016_phage3, and SA06148880. The genomes of SA122025_phage1 and SA714001_phage1 were more than twice the size of the other eight phage genomes. All Group 5 phages, except for SA122025_phage1, were located at the ends of contigs, while three phage genomes (SA242017_phage2, SA06148880_phage3, and SA714001_phage1) spanned the entire contig they were detected on.

Group 6 consisted of five phages of different sizes, that showed some similarity to phiCDHM19 (NC_028996). While there was some similarity in the predicted phage functional modules to the tail morphogenesis and capsid morphogenesis modules in the reference phage, there was

little similarity to the rest of the reference phage genome. However, possible modules with similar function were observed for the predicted phages. For example, a putative DNA packaging functional module, harboured a gene encoding a product with a terminase small subunit. A few genes of the lysogeny control module were found scattered and upstream to the reference, and there was similarity to a single ORF associated with the DNA replication module, which was also upstream. The lysogenic conversion and lysis functional modules were absent. Around half of the phage genomes had hypothetical proteins with unknown function. The phage genomes of SA122108_phage1 and SA05864722_phage1 covered nearly the full length of their respective contigs. However, for SA714043_phage2, the genome started at position 1856bp of the contig. While the genomes of SA122111_phage1 and SA211024_phage1 began at contig positions 9,473 bp and 10,072 bp, respectively, they extended until the contig's terminus. As a result, the possibility remains that certain modules could have been missed due to potential truncation.

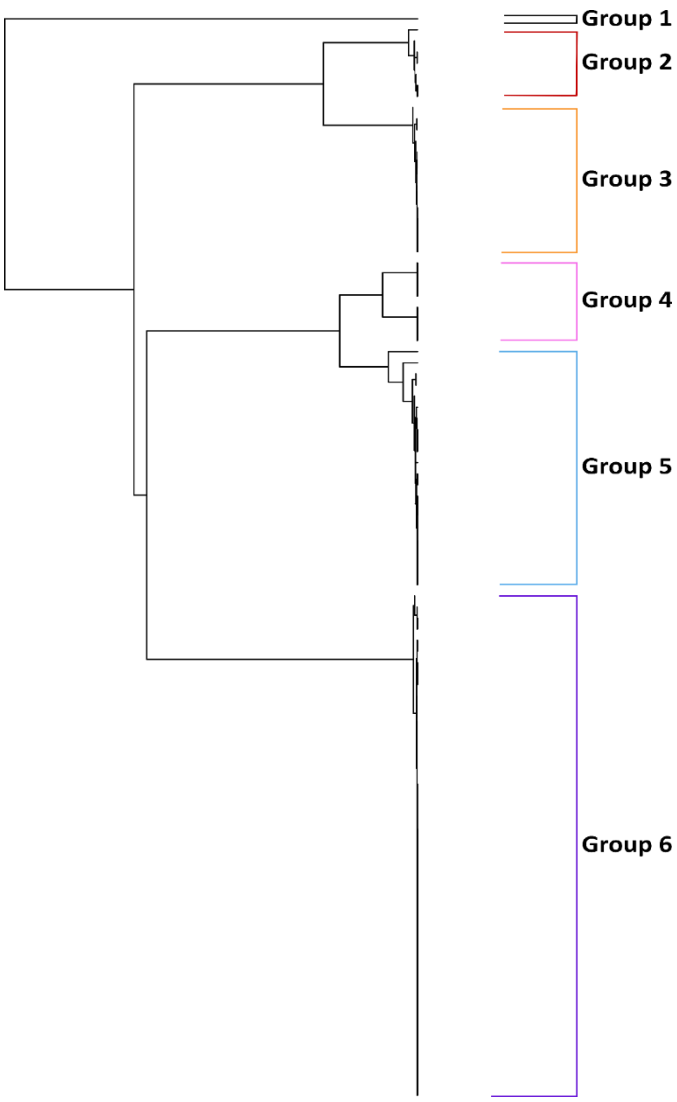



**Figure 9**. Phylogenetic analysis based on Mash distances of 96 phage sequences derived from 56 *C. difficile* isolates. The Mash distance tree revealed the genetic relatedness among the analysed prophages, grouping them into six distinct clusters (Groups 1 - 6). Group 1 (black), Group 2 (red), Group 3 (orange), Group 4 (pink), Group 5 (blue), Group 6 (purple). Group 1 contains one phage (SA111045_phage1), which did not cluster with other groups and, shows genetic divergence. Group 6 contains majority of the putative phages.

### Anti-phage systems

All 58 *C. difficile* genomes and two reference strains, M68 and CF5, were screened for anti-phage systems. In total, 10 different anti-phage systems were identified (**Figure 10**). All isolates contained AbiD, AbiU, types I-B and VI-A CRISPR-Cas systems, Lamassu, and RM types I and IV. All isolates contained one gene which was identified as potentially belonging to VI-A CRISPR-Cas systems, but other mandatory genes were absent. All isolates had low system wholeness and system scores of 0.333 and 1, respectively. The Gajiba system was only found in 1 isolate. Most isolates (68.33%) contained the PD-λ-1 system and 61.67% of isolates did not carry the Kiwa system. The Lamassu and RM type I systems were the only systems which had two copies, as was seen in 96.6% and 90% of isolates, respectively.

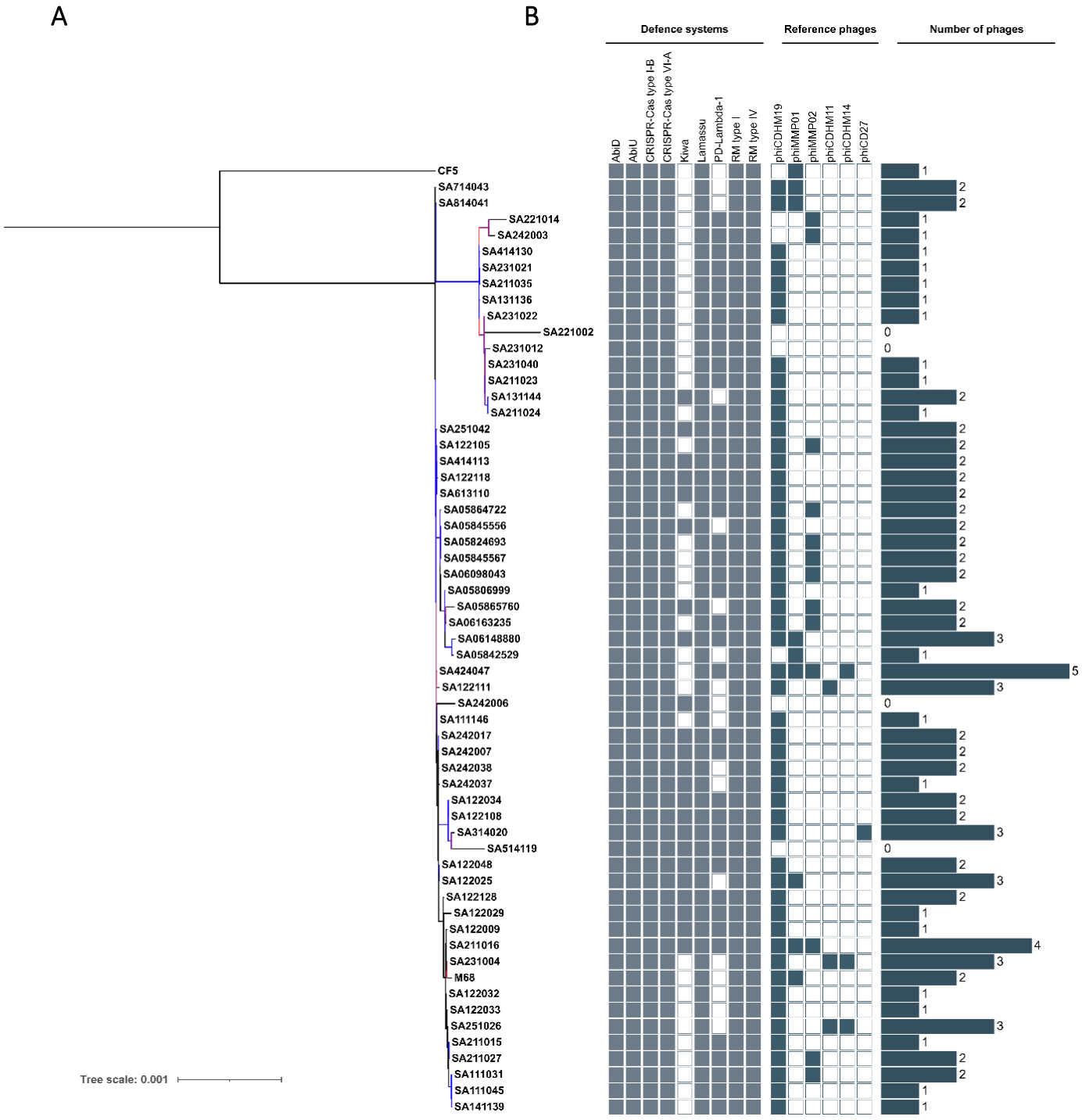



**Figure 10.** Core genome phylogeny and anti-phage defence systems of *C. difficile* RT 017 isolates. **A)** Maximum - likelihood phylogenetic tree of *C. difficile* core genome. Bootstrap values are indicated by colour, ranging from 70% (red) to 100% (blue). **B)** Coloured blocks represent the presence (■) or absence (☐) of anti-phage systems, predicted reference phages, and number of phages.

## DISCUSSION

The first step towards phage therapy is to identify potential lysogens and the phages which they carry. Since phages are highly specific to their hosts, and since the distribution of strains may vary depending on the geographical area, it is therefore important to understand the phage carriage among circulating strains. Currently, South Africa lacks knowledge regarding phages infecting the local *C. difficile* population. This chapter aimed to address this knowledge gap by screening a collection of South African RT 017 clinical isolates for evidence of prophage carriage.

### *C. difficile* genome assembly and phage genome quality

QUAST determined N50 and L50 statistics, which informs about the length of the shortest contig found at 50% of the total assembly length, and the number of contigs that cover 50% of the assembly, respectively. All *C. difficile* genome assemblies were roughly the same length and ranged between 4,118,749 – 4,403,195 bp, with a median size of 4,343,441 bp, and are therefore comparable in size. *C. difficile* typically has a genome size ranging between 4 Mb and 4.4 Mb, while the two well-characterized RT 017 reference strains CF5 (GenBank accession no. NC_017173.1) and M68 (GenBank accession no. NC_017175.1) have genome sizes of 4,159,517 bp and 4,308,325 bp, respectively. Therefore, all *C. difficile* genome sizes were within the expected size range and were of high-quality for further analyses.

The quality of the genome assembly affects phage identification, as bioinformatics tools and algorithms used for phage detection may be compromised by genome fragmentation. Prophages which lie at the edge of a contig may be missed, incorrectly labelled as "incomplete", or parts of the putative phage genome may not be identified. Although high-quality *C. difficile* genomes were used, eight assemblies had N50 values smaller than 132kb, which is smaller than the largest *C. difficile* phage genome. Given that this means that 50% of the genomic data is present on contigs of ≤ 132kb, it is possible that phages have been missed or labelled as incomplete in these genomes. Interestingly, 4/8 of the isolates with N50 ≤ 132kb were detected as carrying no intact phages, though there is insufficient evidence to confirm whether the small contig sizes resulted in PHASTER missing potential intact phages. Resequencing and/or generation of closed genomes using hybrid long-read sequencing approaches could potentially improve phage detection for these assemblies but were,

however, beyond the scope of the study. Genome incompleteness and the presence of contaminating sequences from other organisms could lead to under- and overestimation of the number of prophages, respectively. CheckM has been developed to assess the quality of metagenome-assembled contigs and estimates the bacterial genome completeness and contamination based on the copy number of marker genes, which it defines as single-copy genes identified in ≥97% of genomes. All assemblies in the current analysis were above the thresholds for high quality draft genomes suggesting that incompleteness and contamination were unlikely to have affected the ability of PHASTER to detect prophages in the dataset.

### Phage detection

According to PHASTER analysis, almost all *C. difficile* strains were predicted to carry prophages. However, it is uncommon to find such a high proportion of prophage carriers in a set of *C. difficile* isolates. Phothichaisri *et al*. investigated 73 clinical *C. difficile* isolates for the presence of phages by detecting the phage holin gene via PCR and found 66 isolates to carry the holin gene (170). However, phage induction (using mitomycin C) of these lysogens were successful in only 32.8% of isolates. This indicates that majority of these isolates were not carrying intact phages. Similarly, efforts made by Nale *et al*. to induce prophages from 91 *C. difficile* isolates were successful in only seven isolates (167), despite testing two antibiotics and varying concentrations.

PHASTER predicted multiple phage carriage in 59% of isolates, however, many of these may be non-functional, as carriage of more than one prophage is uncommon. Although PHASTER reported these phages as intact, it may be that a large proportion of them were not intact but rather defective phages. Defective phages, which differ from phage tail-like particles as they have a capsid head containing the viral genome, are widely distributed inducible particles that are unable to infect the target bacterium (167, 283-285). It should also be mentioned that it has been reported that PHASTER has a higher rate of false positive compared to its predecessor, PHAST (262). Given this, it is possible that PHASTER identified defective phages and reported a large proportion of them as intact phages, resulting in an overrepresentation of intact phages. Previously, *Nale et al*. observed dual phage carriage in only one isolate out of the set of 91 (167). Additionally, three isolates carried both intact and defective phages. Therefore, although carriage of multiple prophages (intact or otherwise) is known to exist, our

study supports previous studies that suggest that carriage of multiple intact phages is relatively uncommon in *C. difficile* isolates.

Given that multiple small prophages per genome were found, it is possible that PHASTER had separated larger prophages into multiple smaller prophages, as this is a known limitation (262). CheckV found that the majority (69.4%) of the medium-quality phage genomes are smaller than 50 kb and they were reported as genome fragments. It is then possible these are fragments of larger phage genomes that had been split. Although high-quality *C. difficile* draft genomes were used, an additional analysis using a phage search tool such as VirSorter (updated to VirSorter2) would be useful (257, 258). These tools were designed for phage detection in draft genomes, and they operate especially well using fragmented data, too. A known limitation of PHASTER is that the DBSCAN algorithm assumes an even distribution of phage genes, which is not true in practise, further affecting results (285, 286). To minimize the amount of overreported intact phages, analyses should be done using multiple phage detection tools, with the intersecting results being considered for further bioinformatic analyses.

### Prophage search of reference strains

The PHASTER analysis performed on CF5 and M68 found one intact phage and two intact phages, respectively. The phages with the highest hits for CF5 and M68 were phiMMP03, and phiMMP01 and phiCDHM19, respectively. At present, though there is evidence of prophage carriage (266), the literature contains no data on these strains producing inducible, functional phages.

### CheckV: Genome sizes of intact phages

PHASTER reported the genome sizes of the 98 intact phages to range from 14,620 and 73,412 bp, with a median value of 27,221 bp. This indicates more than half were below 30 kb in size, which is interesting considering they are below the size of the smallest *C. difficile* phage genome of 31,394 bp (184). CheckV was used to assess the phage genome quality. CheckV is a bioinformatic tool designed to identify metagenomic viral contigs. CheckV identifies and removes host regions, and subsequently estimates genome completeness by comparing average amino acid identity (AAI) or hidden Markov Model (HMM) profiles against a database of complete viral genomes (version 0.6). CheckV assigns an in-house quality score which aligns with the Minimum Information about an Uncultivated Virus Genome (MIUViG) standards

(287). Completeness scores were estimated and used to assign a corresponding quality score which align with the MIUViG classification (287). High-quality CheckV completeness scores (≥ 90%) aligns with high-quality MIUViG genomes. CheckV completeness values scoring below 90% are designated as genome fragments by MIUViG standards. CheckV found 25 low-quality genomes with sizes between 14,620 and 26,246 bp, 68 were medium quality with genome sizes from 25,967 to 48,540 bp, and 5 were high-quality with genome sizes from 55,107 and 73,413 bp. The phages identified by CheckV as high-quality phage genomes were SA05842529_phage1, SA06148880_phage1, SA122025_phage1, SA122025_phage2, and SA424047_phage1, all having completeness scores ranging between 97.86 and 100%, strongly suggesting these are intact phage genomes. This shows a trend of larger phage genomes likely producing intact phages. However, at the time of writing, the smallest *C. difficile* phage genome belongs to phiCD08011, with a size of 31,394 bp (184). In fact, there are eight other published phages with genome sizes < 40kb. Typically, *C. difficile* phage genome sizes range from ~30 – ~60 kb(282), except for the big phages, phiCD211 (phiCDIF129T), phiCD5763 and phiCD2955, which have genome sizes of 131.3 kb, 132.5 kb, and 131.6 kb, respectively (179, 180, 191) (Table S3). This suggests that *C. difficile* phages with a genome size below 30 kb are not likely to be complete.

### Genomic organisation of phages

Using Mash distances of predicted phage genomes, a tree was constructed, and the phages were clustered into six groups. The reference phage for each group was assigned and the genomes were mapped based on gene annotations. Overall, phage genomes matched the reference genomes, but there were several absent modules which commonly occurred towards the end of the contig. The genomes used for PHASTER contained several short contigs, and therefore it is possible that not all genes were identified due to assembly truncation. Alternatively, this could reflect genuine loss of these modules in the putative prophages. However, using the present sequence data, it could not be confirmed whether assembly truncation occurred or if the functional modules were lost.

Group 1 contained one phage (SA111045_phage1) which did not cluster with any other phages, and it showed little similarity to the reference phage. This suggests genetic divergence and a possibility of a unique phage.

All phages within Group 2 had high-quality genomes, as confirmed by CheckV predictions.

Notably, in comparison to the reference phages, these phages harboured five additional ORFs in the tail morphogenesis module. While the function of three ORFs could not be determined, the remaining two displayed domains associated with tail morphogenesis. The presence of an Abi-like protein suggests its role in preventing superinfection of other phages. Considering that SA122025_phage2 and SA05842529_phage1 were identified within the middle portion of large contigs, it becomes less likely (though not entirely ruled out) that the absent modules were prematurely truncated during the assembly process.

Group 3, 4 and 5 phages had absent modules, but several of the genomes were located at the end of contigs. Additionally, Group 5 and 6 contained phages whereby the entire contig was identified as a phage. This suggests a possibility of assembly truncation, but it cannot be confirmed due to the limited data. In contrast, one phage (SA122025_phage1), which belonged to Group 5, did not occur at the contig edge. This suggests the possibility that the absent modules were genuinely lost and could potentially result in a non-functional phage.

### *C. difficile* genome analyses

Bacteria have an arsenal of tactics to defend against phage attacks, most notably is the adaptive immunity provided by CRISPR-Cas systems. While all analysed C. difficile isolates were predicted to harbour type I-B CRISPR-Cas systems, the presence of VI-A systems were not convincing. Not only did these isolates lack mandatory genes, but they yielded notably low system and wholeness scores. Furthermore, the published literature does not support this prediction, as type VI-A systems has not been reported to be carried in C. difficile, only type I-B CRISPR-Cas systems have been reported (173, 174). Taken together, this suggests that the type VI-A system was likely misidentified.

CRISPR-Cas systems are divided into two classes based on their effector modules. Class 1 systems have intricate complexes with multiple Cas protein subunits, while Class 2 systems have a single, large, multidomain and multifunctional effector protein (288). Class 1 includes types I, III, IV, and Class 2 includes types II, V and VI, each of which can be further classified into various subtypes (289). The presence of Class 1 type I-B in all isolates is consistent with previous studies (173, 174, 290, 291). Single copies of this CRISPR array were found, except for isolate SA714001 which carried two copies of Class 1 subtype I-B. This finding contrasts with previous reports of *C. difficile* isolates having multiple sets of CRISPR arrays (290, 291). Additionally, class 2 subtype VI-A systems were identified in all isolates. However, each system was missing

important functional components and is unlikely to be functional. Up until now, subtype VI-A CRISPR-Cas system has not been described in *C. difficile*. Briefly, type I loci have the signature *cas3* gene which encodes a large protein with a helicase and the I-B effector complex is encoded by the *cas8b1, cas7* and *cas5* genes. Cas6, which is found upstream of the effector operon, is involved in pre-crRNA processing into mature crRNAs, which then interacts with the effector complex to locate the protospacer (foreign DNA sequence complementary to the spacer) (292, 293). Cas8 recognizes the protospacer adjacent motif (PAM) and induces binding of the effector complex to its target, forming an R-loop between crRNA and dsDNA. The Cas3 nuclease then cleaves target region downstream of PAM and the opposite strand is degraded (292, 293). On the other hand, the subtype VI-A system has effector protein (Cas13a) is encoded by *cas13a*, has two HEPN (Higher Eukaryotes and Prokaryotes Nucleotide-binding) domains which enables targeting and cleavage of ssRNA phages (289, 294). Unlike other subtypes of type VI systems, VI-A has genes for both *cas1* and *cas2* (292). Interestingly, Cas13a has co-lateral RNase activity, resulting in non-specific mRNA cleave, which impairs growth of the bacteria and leads to cell death or dormancy (295). *In silico* studies have identified CRISPR arrays in *C. difficile* and *C. difficile* prophages, often carrying multiple copies of the array (174). Biological experiments have not been done to confirm the functionality of these findings.

AbiD and AbiU systems were present in majority of isolates, including M68 and CF5. Abi is a common resistance strategy against phage infection which is reserved as a last line of defence if other resistance mechanisms fail (201). The system is tightly regulated, and the mechanism for cell death is only activated when sensor proteins detect phage infection (296). Complete Abi systems have not been identified in *C. difficile*. An AbiF protein has been identified on the ϕC2 genome and has been speculated to be involved in preventing superinfection, modulating ϕC2 replication, and may play a role in low levels of free phage isolation (181, 214). Additionally, an Abi-like protein has been identified in ϕ027, which is present on in RT 027 isolates (297).

Restriction modification (RM) systems Type I and Type IV were identified. RM systems modifies and protects self-DNA through the action of methyltransferase (MTase) which recognizes specific sequences known as recognition sites and adds a methyl group. Unmethylated recognition sites, such as those found in foreign DNA, will be recognized and cleaved by

restriction endonucleases (REases) (298). CdiCD6II RM systems have been identified in *C. difficile*, but experimental evidence of the activity is lacking (299).

The Lamassu, Kiwa, and Gajiba systems were identified; they are recently discovered systems that are not yet fully characterised, nor have they been identified in published studies. Lamassu is a two-gene system consisting of a structural maintenance of chromosomes (SMC) gene (*lmuB*) and *lmuA* encoding an effector protein having an N-terminal Cap4 dsDNA endonuclease domain (300, 301). It has been found in 1.3% of sequenced genomes, and showed protection against siphoviruses in *Bacillus cereus* VD014 and *Bacillus sp.* NIO-1130 (302). The Kiwa system has two genes, *kwaA* encoding a transmembrane protein and *kwaB* encoding a protein with the DUF4868 domain. The Kiwa system has shown protection against certain siphoviruses in *Escherichia coli* O55:H7 RM12579. The Gabija system consists of *gajA* which encodes a sequence-specific DNA nicking endonuclease protein (303), and *gajB* which encodes a helicase (302). Gabija systems has been found in 8.5% of sequenced bacterial and archaeal genomes, and shows protection against myoviruses in *B. cereus* VD045 and *B. cereus* Hub5-5 (302). The PD-Lambda-1 (PD-λ-1) system was detected in most isolates and was absent in CF5 and M68. The PD-λ-1 system operon has been predicted to encode proteins that contain a transmembrane domain, DUF4041 domain, and a GIY-YIG-nuclease domain. PD-λ-1 confers robust protection against lambda phages in *Escherichia coli*, and up until now, this resistance mechanism had not been observed in *C. difficile* (304), and requires further investigation.

Various mechanisms enable phages to infect bacteria even when anti-phage defence systems are carried. These include anti-CRISPR proteins [304], phage-encoded CRISPR-Cas arrays [165, 184], gene mutations that prevent Abi system activation [305], RM evasion strategies such as base modification or decreasing the number of restriction sites [306, 307], and more. Since all bacterial isolates in this study carried numerous defence systems, it suggests that the success of phage infections may rely more on the phage's ability to evade host defences rather than their absence. The current literature offers limited insight into *C. difficile* antiphage systems and most *in silico* predictions lack experimental validation. Therefore, further studies are crucial to validate the predictions presented here and confirm the functionality of these potential antiphage systems. Furthermore, future research must be done to understand the interplay between host defence systems and phage evasion strategies, and their implications for phage therapy.

## CONCLUSION

Using a bioinformatic approach, a diverse population of phages was identified in majority of *C. difficile* isolates. However, only a small portion of phages were predicted to have high-quality genomes. Mash values were used to construct a distance tree, whereby six distinct groups of phages were established, and a genetically divergent phage was identified, indicating a potentially unique phage. Genomic features, based on genome annotations, were mapped for each phage group. This study also investigated *C. difficile* isolates for potential defence systems and reported a wide repertoire of anti-phage systems. Future studies utilising *in silico* phage detection methods should employ bioinformatic tools better equipped for screening draft genomes. Furthermore, phage genomes should be analysed to identify genes involved in host immune evasion strategies and superinfection immunity. Subsequent biological experiments are required to validate the bioinformatic predictions.

# CHAPTER 4: PROPHAGE INDUCTION AND OPTIMIZATION

### SUMMARY

Prophages are bacteriophages that have integrated into the host bacterial genome. These prophages can be triggered to produce viable phages, resulting in the lysis of its host. Temperate phages have been reported in *C. difficile* and have been shown to be induced by various agents. An *in silico* analysis, in Chapter 3, described the presence of prophages in a set of clinical isolates investigated in this study. In this chapter, the *in silico* observations were validated by inducing phages from a subset of South African *C. difficile* isolates by means of exposure mitomycin C, a DNA-damaging agent. A subset of 20 isolates were induced in liquid assays, and the phage lysates of 29 isolates were tested in double agar overlay and spot test

assays. PCR assays were performed to classify phages as myoviruses or siphoviruses. Four isolates (20%) were induced in liquid assays, while 7 different isolates produced plaques (24.1%). Subsequent conventional PCR assays successfully characterised the lysate of one isolate (SA06148880) as a myovirus.

## INTRODUCTION

### Phage life cycle

Temperate phages can exist in either a lytic or both lytic and lysogenic lifestyles (139). Phage infection occurs by binding to surface receptors on the target bacterium, followed by insertion of phage genetic material (139). Lysogenic phages integrate their genetic material into the host genome and the phage genetic material, now known as a prophage, replicates with each host cell replication cycle (305). This phase persists until phage induction occurs, leading to production of viral particles and bacterial cell lysis. In the lytic cycle, phages take over the host replication machinery to produce the next generation of virions, which lyse the bacterial cell to release mature viral particles. Phages capable of replicating in both the lysogenic and lytic phase are known as temperate phages. Switching from the lysogenic to lytic phase is known as phage induction (139).

*C. difficile* harbours temperate phages belonging to the class *Caudoviricetes*. At the time of writing, with the new taxonomic guidelines, *C. difficile* temperate phages have not been assigned into families, previously they were assigned into the now obsolete families *Myoviridae* (having contractile tails) and *Siphoviridae* (having long, non-contractile tails). Further morphological characterisation under transmission electron microscopy (TEM) has identified different morphotypes for each family (167, 169). Although *C. difficile* phages are usually isolated following induction (commonly done using mitomycin C, norfloxacin, and UV irradiation), free phages have been isolated from faecal samples of CDI patients, indicating *in vivo* induction (189). *In vivo* prophage induction demonstrates the possible effect phages play on the microbial ecosystem within the gut and its potential implication in disease related to dysbiosis of the gut, a hypothesis which requires further exploration.

Isolation of temperate phages from host bacteria involves induction of prophages by non-lethal DNA damage, typically using antibiotics (167), ultraviolet light irradiation (306), or mitomycin C (307). The subsequent host lysis can be observed as plaque formation in double agar overlays

or as clearing in liquid media. However, this process presents challenges, as phage induction varies on the inducing agent and its concentration (167), which could also differ between different host strains. Therefore, identifying the preferred inducing agent and its optimal concentration beforehand is crucial for successful isolation of temperate phages.

In South Africa, RT 017 is the dominant strain circulating in hospitals, particularly in TB hospitals, yet no knowledge is available relating to the phages infecting *C. difficile*. This research will contribute to the current pool of knowledge regarding phages and will provide foundational material for further research into South African phages.

## METHODS AND MATERIALS

### Clinical *C. difficile* sample collection, growth media and culturing conditions

#### *C. difficile* clinical samples used.

A collection of 58 *C. difficile* clinical isolates, hosted in the Division of Medical Microbiology and collected in a previous study from patients with CDI (HREC ref 310/2008) was used. Isolates within this collection have been previously characterised. Metadata associated with the deidentified isolates were provided by the research project co-supervisor, who generated data during post-doctoral studies. This consisted of antimicrobial susceptibility profiles, whole genome sequence data, *tcdA*, *tcdB* and *cdtAB* toxin profiles, ribotype and multilocus variable-number tandem-repeat analysis (MVLA) data. These data have been reported in previous publications (97, 98, 226). A manageable subset of 29 isolates was used for double agar overlay plaque assay and spot test assays. All isolates included in the subset were revived from spore stocks using brain heart infusion agar (BHI) (Thermo Fisher Scientific, Massachusetts, USA), then stored in BHI broth containing 50% glycerol at -80°C. Gram stains were done to confirm pure cultures of Gram-positive rods.

### Culture media and growth conditions

Culturing, media used, and culture conditions were as described in Chapter 2 (Bacterial culture media and growth conditions).

### Prophage induction using Mitomycin C

#### Mitomycin C concentration titration

Titration experiments were done to determine the sublethal concentration of mitomycin C (MMC) against the *C. difficile* isolates used in this study. The concentrations of MMC used were

1, 2, 4, and 5 µl/mg. Briefly, isolates were streaked from frozen stocks and grown as described in Chapter 2 (Bacterial culture media and growth conditions). Overnight liquid cultures were diluted 1:10 into BHIS broth and grown for 3 hours to reach mid-log phase. Thereafter, 500µl culture and MMC (at final concentrations of 1, 2, 4, and 5µl/mg, respectively) added to separate screw-cap glass tubes containing 5ml molten 0.8% BHIS agar, which was kept molten at 50°C in a water bath. This molten mixture was poured onto a 1.5% BHI plate and allowed to set before being incubated overnight under anaerobic conditions (described in Chapter 2).

### Mitomycin C induction

Liquid MMC phage induction assays was performed using a previously described protocol (239). Briefly, *C. difficile* isolates were revived from frozen stocks and cultured as described in Chapter 2. To obtain lysates, overnight cultures were diluted (1:10 dilution) into fresh BHIS broth and grown for 3 - 4 hours to reach mid-log phase (239). MMC was added to achieve a final concentration of 3µg/ml (167, 189) and cultures incubated for another 16-18 hours (overnight) at 37°C in an anaerobic jar containing an anaerobic sachet (AnaeroPack, Mitsubishi Gas Chemical, Japan). The following day, lysates were centrifuged at 3000 × g for 10 minutes (Centrifuge 5420, Eppendorf), whereafter supernatants were filtered through 0.22 µm syringe filters (MF-Millipore) to remove unlysed host cells or debris. Phage lysates were stored at 4°C in both original culture media (containing 10 mM each of $MgCl_2$ and $CaCl_2$) and SM buffer (100mM NaCl, 8mM $MgSO_4$, 50mM Tris-Cl (pH 7.6), 0.001 % (w/v) gelatine) (166, 239).

### Phage amplification and precipitation

To enrich for possible phages, two host strains SA221002 and SA05842529 were exposed to crude phage filtrates.

Briefly, 1ml overnight cultures of host strains were diluted 1:10 into 10ml BHIS broth and incubated for 3 hours anaerobically at 37°C. These log phase cultures were exposed to 100µl crude phage filtrates; incubation was done overnight (16 h) at 37°C, in anaerobic jars with AnaeroPacks® (Mitsubishi Gas Chemical, Japan). Cultures were centrifuged at 20,000×g (Centrifuge 5420, Eppendorf), filtered using syringe and 0.45µm membrane filters (MilliporeSigma, Darmstadt, Germany). Two aliquots of filtrates were stored: One aliquot was stored the original media at 4°C (308), a second aliquot was precipitated and stored in SM buffer (100mM NaCl, 8mM $MgSO_4$, 50mM Tris-Cl (pH 7.6), 0.001 % (w/v) gelatine).

Phages were precipitated using a modification of a previously described method (309). For this, 1.5ml crude phage filtrate was transferred to a clean 2ml Eppendorf tube and mixed with 375µl of a PEG solution (20% (w/v) polyethylene glycol (PEG)-8000, 2.5M NaCl). This was mixed gently by inversion and incubated at 4°C overnight to improve phage precipitation. The following day, the phage solution was centrifuged at 13,000×g (Centrifuge 5420, Eppendorf) for 10 minutes at 4°C. The supernatant fraction was removed except for 100µl; the latter was resuspended in 500µl BHI broth. Residual PEG was removed by a second centrifugation at 13,000×g for 10 minutes at 22°C. The supernatant was carefully transferred to a clean 2ml Eppendorf tube and topped up to 500µl BHI. To increase stability of phages in culture media, $MgCl_2$ and $CaCl_2$ were added to a final concentration of 10mM each, and the phage lysate was stored at 4°C.

### Double agar overlay

Ten millilitres of overnight liquid cultures were diluted 1:10 into fresh BHIS broth and grown for 2 hours to reach early log phase. Thereafter, 200µl concentrated phage solution was added and incubated for one hour to allow phages to adhere to the host cells. After incubation, 1ml was transferred into 5ml 0.8% BHIS molten agar and poured over 1.5% BHI agar plates. Overlays were solidified on the bench at room temperature for 10 minutes and incubated 20 – 24 hours anaerobically at 37°C.

### Double agar overlay of serially diluted phage-bacteria solutions

Briefly, 200µl phage solution was added to 10ml early log phase *C. difficile* culture grown in BHIS broth. The solution was incubated for 1h to allow the phages to adhere to host cells. Thereafter, 200µl was serially diluted into 5 glass bottles containing 9ml BHIS broth. To each tube, a further 800µl aliquot of cultured *C. difficile* was added, giving a final volume of 10ml. From this, 1ml was used in double agar overlay assays (as described above) and incubated for 24h in an anaerobic chamber (BugBox®, Baker Ruskinn) under an atmosphere of nitrogen (85%), $CO_2$ (10%), and hydrogen (5%), at 37°C and 60% relative humidity.

### Phage spot test assay

Spot test assays were done to test if prepared lysates contained viable phages that are capable of forming plaques. Growth conditions, media and double agar overlays preparation was done as described in Chapter 2. Briefly, overnight liquid cultures were diluted 1:10 into fresh BHIS broth and incubated for 3 hours. After incubation, 1ml was transferred into 5ml 0.8% BHIS agar, which was kept molten in a water bath at 50°C and poured over 1.5% BHI agar plates.

Overlays were allowed to set on the bench at room temperature for 10 minutes, then 5µl phage solution was spotted onto the plate. When dry, plates were inverted and incubated 20 – 24 hours anaerobically at 37°C.

### Phage lysate media and processing

Following phage double agar overlay or spot test assays, plaque were picked with sterile pipette tip, and added to 1ml BHIS liquid media or SM buffer (100mM NaCl, 8mM $MgSO_4$, 50mM Tris-Cl (pH 7.6), 0.001 % (w/v) gelatine). Phage plaque agar plugs were crushed and allowed to diffuse out of the media at room temperature overnight (236). To remove unlysed bacterial cells and debris, the solution was centrifuged at 4000×g for 1 minute and filtered through 0.22 µm syringe filters (MF-Millipore). Filtrates were stored at 4°C until used for other procedures (phage precipitation, induction experiments, PCR verification of phage DNA and classification).

### Monitoring optical density following high-throughput prophage induction

#### Isolate selection

A subset of clinical *C. difficile* isolates was used for a high throughput mitomycin C induction, with a total of twenty isolates (Table S2). The isolates selected included four prophage-free isolates (as determined by PHASTER in Chapter 3), sixteen isolates carrying prophages which were selected based on *in silico* confirmed prophage diversity, and two RT 017 reference strains, M68 and CD630 (GenBank accession numbers NC_017175.1 and NC_009089, respectively)

#### Establishing *C. difficile* growth curve in microtitre plates

To determine when early-log phase was reached, growth curves were established for each *C. difficile* isolate. Growth was initiated by inoculation of single colonies into BHI broth, in 96-well plates. Cultures were incubated overnight in an anaerobic workstation (Concept Dual 1000 Anaerobic Workstation, Baker Ruskinn), under an atmosphere of nitrogen (85%), $CO_2$ (10%), and hydrogen (5%), at 37°C and 60% relative humidity. On Day 2, flat bottomed, 96-well plates were set up as shown in Figure 11, with 180µl and 190µl BHI aliquoted into wells of each half of the plate. Aliquots of 20µl and 10µl of overnight cultures were inoculated into wells containing 180µl and 190µl BHI, respectively, generating 1:10 and 1:20 dilution starter cultures. Growth determinations were done in duplicate. The 96-well plate was placed in a

plate reader (Stratus, Cerillo, Charlottesville, USA), generating optical density values read at a 600nm wavelength ($OD_{600}$) every three minutes. The mean $OD_{600}$ values (from 2 replicates) and standard deviations were plotted using Excel (Microsoft Corporation, version 2306).

### High-throughput mitomycin C induction in liquid culture

A high-throughput induction with mitomycin was done as described by Phothichaisri *et al* (170), with modifications as follows: Cultures were set up as described in the above section. The plate was incubated for 195 minutes in an anaerobic chamber, whereafter 5µl of 123µg/ml mitomycin C (3µg/ml final concentration) was added to all experimental replicates and 5µl BHI

**A**

| | Replicate 1 | | | Replicate 2 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| A | S1 | S9 | S17 | S1 | S9 | S17 | | | | | | |
| B | S2 | S10 | S18 | S2 | S10 | S18 | | | | | | |
| C | S3 | S11 | S19 | S3 | S11 | S19 | | | | | | |
| D | S4 | S12 | S20 | S4 | S12 | S20 | | | | | | |
| E | S5 | S13 | - | S5 | S13 | - | | | | | | |
| F | S6 | S14 | - | S6 | S14 | - | | | | | | |
| G | S7 | S15 | - | S7 | S15 | - | | | | | | |
| H | S8 | S16 | - | S8 | S16 | - | | | | | | |

**B**

| | Induced | | | | | | Uninduced | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| A | S1 | S9 | S17 | S1 | S9 | S17 | S1 | S9 | S17 | S1 | S9 | S17 |
| B | S2 | S10 | S18 | S2 | S10 | S18 | S2 | S10 | S18 | S2 | S10 | S18 |
| C | S3 | S11 | S19 | S3 | S11 | S19 | S3 | S11 | S19 | S3 | S11 | S19 |
| D | S4 | S12 | S20 | S4 | S12 | S20 | S4 | S12 | S20 | S4 | S12 | S20 |
| E | S5 | S13 | - | S5 | S13 | - | S5 | S13 | - | S5 | S13 | - |
| F | S6 | S14 | - | S6 | S14 | - | S6 | S14 | - | S6 | S14 | - |
| G | S7 | S15 | - | S7 | S15 | - | S7 | S15 | - | S7 | S15 | - |
| H | S8 | S16 | - | S8 | S16 | - | S8 | S16 | - | S8 | S16 | - |

**Figure 11**. Set up of 96-well plate for growth curves. **(A)** Pre-experimental set up, to establish baseline growth parameters prior to induction. **(B)** Experimental set up for prophage induction, done in duplicate, including uninduced controls. Blanks indicated by minus sign.

to replicates of the uninduced controls. The plate was incubated overnight under anaerobic conditions and the growth was recorded by measurement of the optical density at 600nm wavelength ($OD_{600}$) using a plate reader (Stratus, Cerillo, Charlottesville, USA). OD values were recorded at three-minute intervals.

### Phage DNA purification

Residual bacterial DNA was removed from lysates by DNAse I treatment. Reaction mixtures contained 450µl phage filtrate, 50µl DNase I buffer (10X, New England BioLabs), 1µl DNase I (1U/ml, New England BioLabs), and 1µl RNase A (20mg/ml, New England BioLabs). Samples

were incubated at 37°C for 1.5h on a heat block. DNAse I and RNAse A were inactivated with 20µl of 0.5M EDTA (20 mM final concentration) and heated for 10 minutes at 75°C. To obtain phage DNA, viral capsid heads were digested using 1.25µl of 20mg/ml proteinase K, incubated at 56°C for 1.5 hours, without shaking. Proteinase K was heat inactivated at 90°C for 10 minutes. Crude phage DNA solutions were stored at 4°C.

### Conventional PCR assays to verify absence of residual bacterial DNA in phage lysate

To identify host *C. difficile* DNA from phage lysate samples, amplification of the bacterial 16s rRNA gene PCR was done using previously published primer pairs: forward primer, 27F (5'-AGAGTTTGATCCTGGCTCAG-3'), and reverse primer, 1525R (5'-AAGGAGGTGWTCCARCC-3') (237). PCRs were carried out in 2720 Thermal-Cycler (Applied Biosystems, California, USA). Reaction mixtures (25µl total volume per reaction) contained 5µl 5X Go Taq Buffer (Promega), 2mM $MgCl_2$ (Promega), 0.2mM dNTPs (New England Biolabs), 0.2µM each primer, and Taq (Promega) at a concentration of 0.026U/µl. PCRs were run with an initial denaturation at 94°C for 5 minutes, followed by 35 cycles of 30s denaturation at 94°C, 30s annealing at 56°C, and 45s elongation at 72°C, and a final elongation step of 5 minutes at 72°C. PCR products were run on 2% agarose gel (SeaKem LE Agarose, Lonza, Basel, Switzerland) (stained with ethidium bromide (0.5µg/ml)) electrophoresis for 1h at 100V and visualized under UV transillumination (GelDoc XR+ Molecular Imager, Bio-Rad, Hercules, California, USA) using Image Lab (version 6.1.0 build 7). The QuickLoad 1 kb DNA ladder (New England BioLabs) was used to establish band size.

### Conventional PCR assays to classify obtained phages

PCRs targeting the myovirus and siphovirus holin genes were performed using published primers (169) were done as described in Chapter 2. PCR products were assessed and visualized agarose gel electrophoreses and UV transillumination as described above. A 100 bp DNA ladder (Promega) was used to determine band size.

PCR products were subjected to commercial Sanger sequencing (InqabaBiotech, Pretoria, South Africa) from both ends using the same primers used for amplification. The results of the sequences were analysed using the NCBI database using the BLAST algorithm (270).

# RESULTS

## Optimization of prophage induction using mitomycin C

### Titration of Mitomycin C in double agar overlays

Although published data exists regarding the concentration of MMC to use against *C. difficile*, local isolates may differ in sensitivity. Thus, to determine the optimal sub-lethal concentration of MMC to be used for induction, a titration of MMC was done using concentrations 1, 2, 4, and 5 µg/ml. At concentrations of 1µg/ml, bacterial lawns were present, but when the concentration was increased to 2µg/ml, the lawns thinned out. At 5µg/ml, the bacterial growth was too poor to be used for plaque assays. Henceforth, induction was done using a MMC concentration of 3µg/ml, as the dose did not severely affect *C. difficile* growth and this concentration is routinely used for *C. difficile* prophage induction (167, 169, 170).

### Optimization of double agar overlay and prophage induction.

To obtain the most confluent bacterial lawn required for plaque observation in overlays, various methods were tested. To determine the optimal growth phase to start prophage induction, a 1:10 dilution of overnight culture was grown for 3, 4, 6 and 8 hours prior to addition of mitomycin C. No visible difference was observed in the lawn thickness, and therefore induction was started after 3 – 4 hours of incubation, as the cells would be in the early log phase. To determine the optimal time required for induction, isolates were induced for 6h, overnight, and 48 hours. No observable difference was noted among the different induction times and thus overnight induction was employed in subsequent experiments.

Following this, different amounts of culture and soft agar, using 0.5ml culture in 3ml soft agar, 0.5ml soft agar in 5ml soft agar, or 10ml culture in 100ml soft agar, or 15ml in 100ml soft agar. Overall, it was found that optimal growth (more confluent lawns) was seen when 10 -15ml liquid culture was added to 100ml soft agar. This mixture was poured and plaque spot assays done. This was time saving and minimized oxygen exposure time.

### Plaque detection by double agar overlays and spot test assays

Double agar overlay (DAO) and spot test assays were done. Two and seven isolates produced plaques using SA05842529 and SA221002 as host strains, respectively (Table 10). Spot test produced large (2-5mm) zones of lysis while DAO produced small 1mm plaques. Phages induced from lysogen SA122111 produced plaques in both host strains.

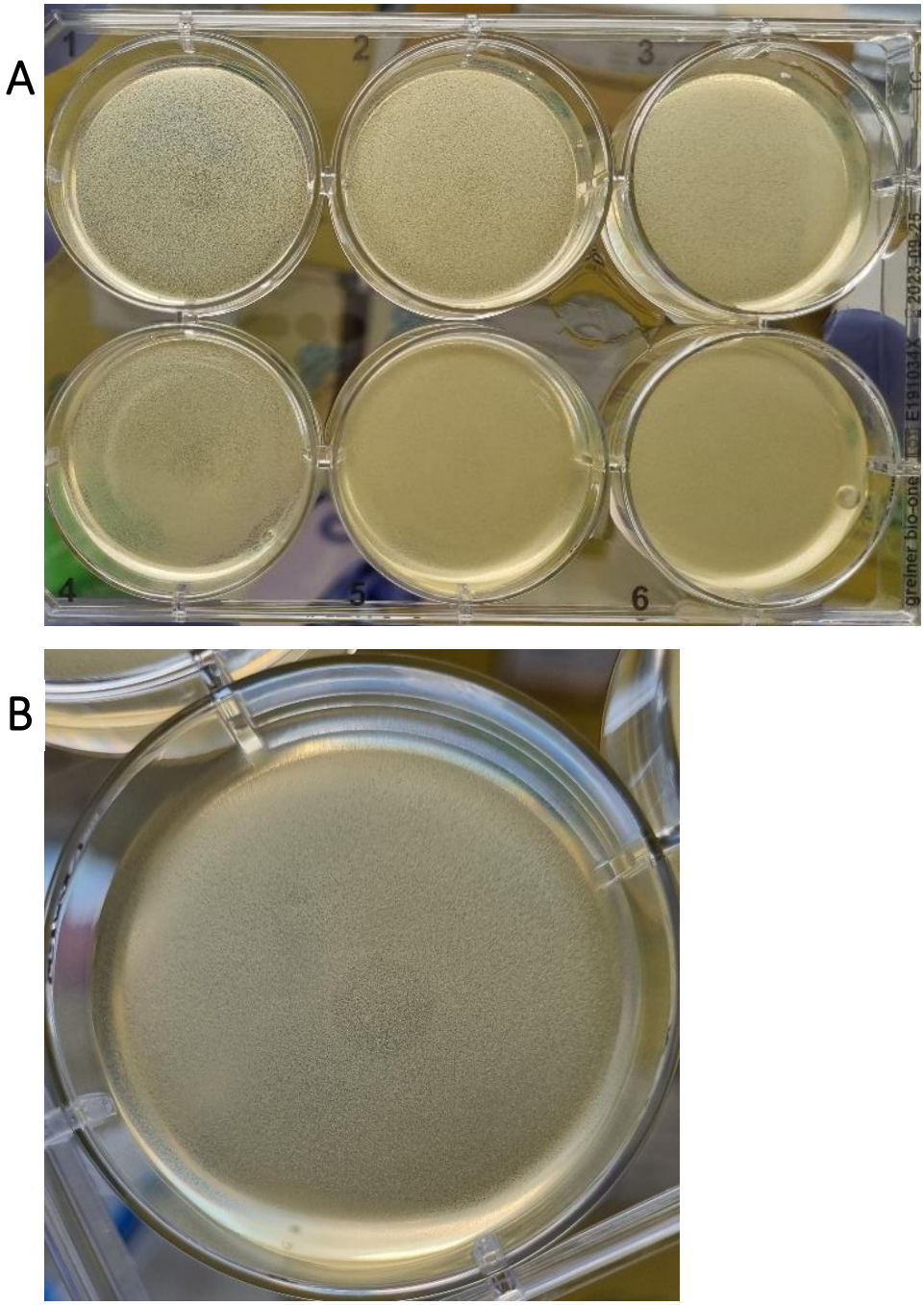



Figure 12. Turbid plaques identified on spot test assays and double agar overlay assays. A) Wells 1, 2 and 4 show large turbid plaques produced on host strain SA221002. Phage lysate for wells 1, 2, and 4 originated from induction of lysogens SA211016, SA231022, SA242003, respectively. B) Large (4mm) zone of lysis observed following spotting of phage lysate from lysogen SA05824693 onto host SA221002.

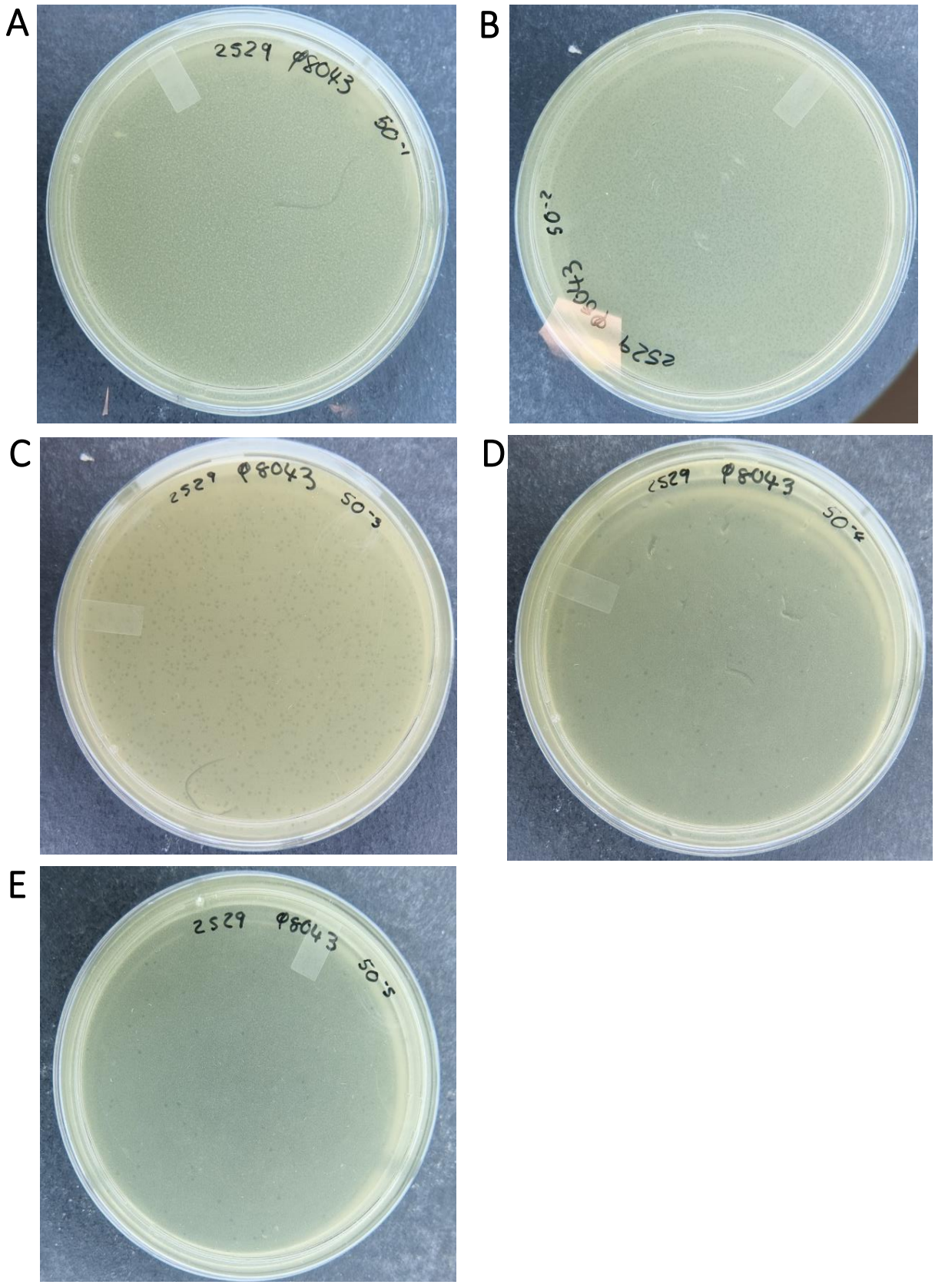
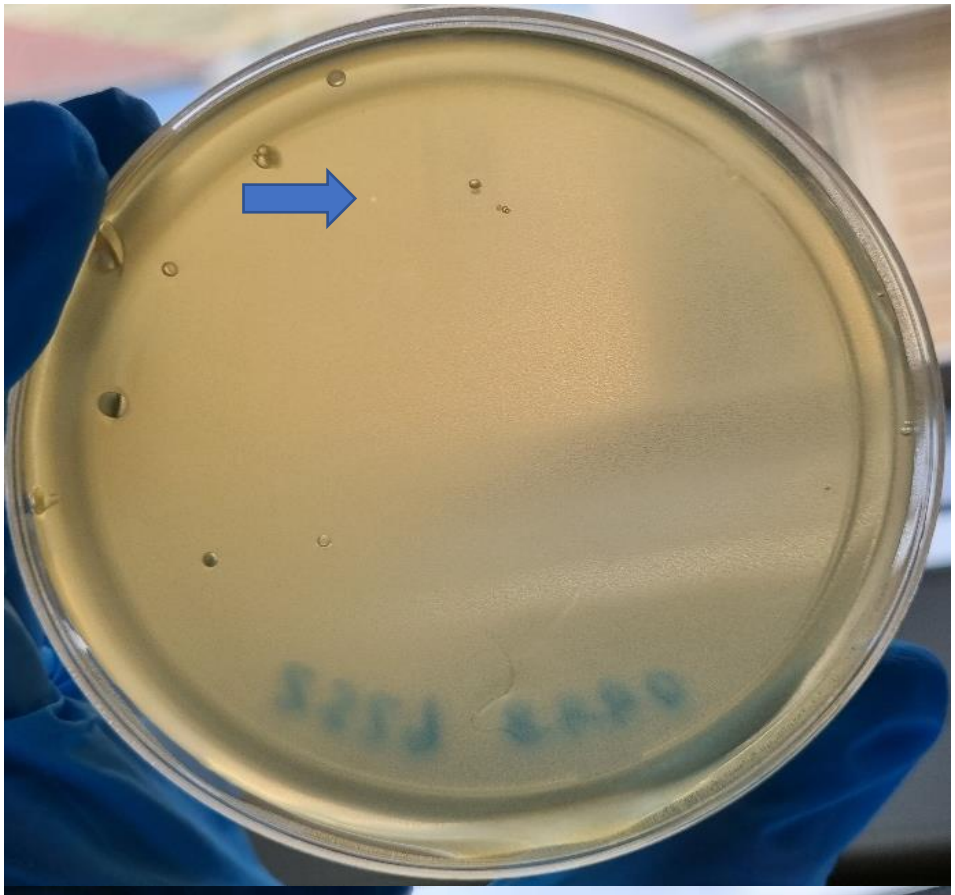



**Figure 13.** Phage titres were quantified by performing a fifty-fold dilution series on double agar overlays from $50^{-1}$ to $50^{-5}$ dilutions. Host strain SA05842529 was infected with a phage solution originating from induction of lysogen SA06098043. **A – C)** High phage titres caused excessive amounts of bacterial lysis on plates. **D** and **E)** Phage titres for $50^{4}$ and $50^{5}$ plates were established to be 5.5x108 and 1.6x1010 PFU/ml, respectively.

A



B

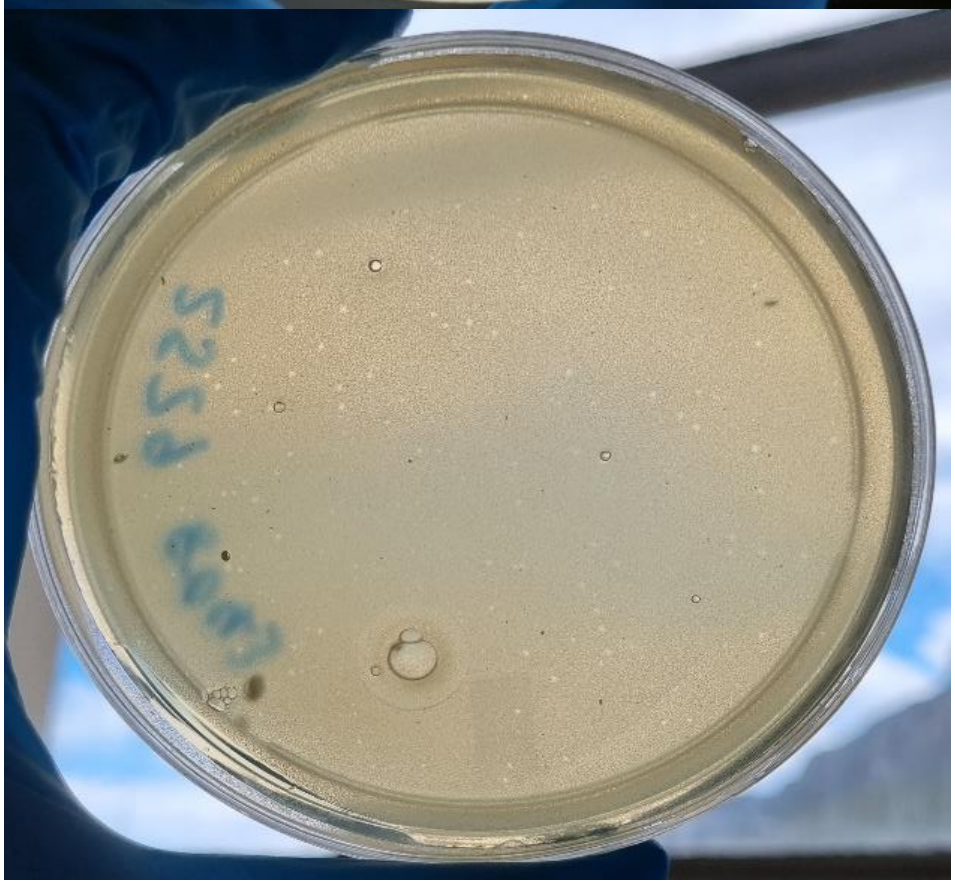


**Figure 14.** Plaque presence on double agar overlay assays using host strain SA05842529 **A)** Host strain was infected with phage ϕSA06148880 and produced a single plaque (blue arrow). **B)** The host strain was infected with phage ϕSA06098043, resulting in a high plaque formation.

Table 10. Plaque production in double agar overlay and spot test assays.

| | Host | |
|---|---|---|
| Phage lysate origin | SA05842529 | SA221002 |
| SA05806999 | - | - |
| SA05824693 | - | + |
| SA05842529 | - | - |
| SA05845556 | - | - |
| SA05845567 | - | - |
| SA05864722 | - | - |
| SA05865760 | - | - |
| SA06098043 | + | + |
| SA06148880 | + | - |
| SA122105 | - | + |
| SA231004 | - | - |
| SA242003 | - | + |
| SA122108 | - | - |
| SA242007 | - | - |
| SA613110 | - | - |
| SA221014 | - | - |
| SA414113 | - | - |
| SA122118 | - | - |
| SA211015 | - | - |
| SA211016 | - | + |
| SA231021 | - | - |
| SA231022 | - | + |
| SA242017 | - | - |
| SA314020 | - | - |
| SA122111 | + | + |
| SA221002 | - | - |

+, plaques observed

-, no plaques observed

Plaques produced in *C. difficile* host strains SA05842529 and SA221002 infected with phage lysate solutions induced from 26 *C. difficile* isolates.

## Observing prophage induction by monitoring host cell density

### Establishing *C. difficile* growth curve

A high-throughput MMC prophage induction was done to determine if the results obtained in Chapter 2 can be experimentally proven. The list of isolates used is provided in Supplementary material (**Table S2**). The average of the $OD_{600}$ values for each replicate were plotted (**Figure 15**). Out of the twenty isolates, four (SA05806999, SA06098043, SA06148880, and SA122111) evidence of induction.

Isolate SA06148880 had $OD_{600}$ = 0.32855 at the point of induction, and its highest OD was 0.42045 which was reached at 300 minutes, and slowly dropped to 0.3266 at 600 minutes. At 300 minutes, the SA06148880 uninduced control had OD600 of 0.8673, and reached its highest point at 360 minutes with 0.9092 $OD_{600}$.

SA06098043 had OD = 0.15305 at 195 minutes, maximum OD was found at 420 minutes 0.36135, and dropped to 0.30205 at 600 minutes. SA06098043 uninduced control had lower OD at 195 minutes, 0.0455 $OD_{600}$, but had a higher maximum OD of 0.50035 at point 510 minutes. At 420 minutes, the control had already surpassed SA06098043, with OD 0.4829.

Isolate SA122111 and SA122111 uninduced control had similar OD at the point of MMC addition, 0.3229 and 0.32805, respectively. 285 minutes SA122111 had its highest OD of 0.45015, at this time the control had OD 0.58155, which increased to reach a maximum of 0.58985 at 360 minutes.

At 195 minutes, SA05806999 had an OD of 0.1438, and its respective uninduced control had OD = 0.05555. By 330 minutes SA05806999 had reached maximum optical density of 0.31155, where it plateaued, while the control had a lower OD of 0.2935 at that time. The SA05806999 uninduced control spiked and reached a maximum OD of 0.56075 at 450 minutes.

Most isolates had a lower lag phase $OD_{600}$ than the uninduced control, but these differences in OD do not indicate induction of isolates.

### Detecting conserved holin genes in *Myoviridae* and *Siphoviridae* using conventional PCR

Conventional PCR was done to identify phages from the isolates showing phage induction and reported in **Table 12**. The PCR primers target *C. difficile* myovirus and siphovirus holin genes, with expected amplicon sizes of 227 bp and 150bp, respectively (169). A positive PCR for a myovirus was obtained from one isolate, SA06148880, with an expected amplicon size of 227 bp (**Figure 16**) Myovirus PCR on phage lysates derived from isolates SA06098043, SA122111

and SA05806999 were negative. All isolates were negative for holin genes from the siphoviruses. Genomic DNA from SA221002 was used as the positive control, as it contains the PCR target gene as determined by *in silico* PCR analysis. Myovirus PCR on genomic DNA showed a clear band, and siphovirus showed non-specific binding, with bands of varying size. The absence of bacterial DNA was confirmed by absence of 16S rRNA PCR amplicons (Figure 17)

Myovirus holin-specific PCR products of SA06148880 and SA221002 lysate solutions were subjected to Sanger sequencing and the generated sequences were analysed using BLAST analysis of the NCBI database (270, 310). BLAST analysis revealed strong homology to *C. difficile* phage holin proteins (between 86.89% and 82.32% identity) on *C. difficile* genomes and phage genomes.

Table 11. Detection of phage induction via plaque assays and liquid induction assays.

| | Plaque assay | | Liquid induction |
|---|---|---|---|
| Induced isolate | SA05842529 | SA221002 | |
| SA05806999 | - | - | + |
| SA05824693 | - | + | - |
| **SA06098043** | **+** | **+** | **+** |
| **SA06148880** | **+** | **-** | **+** |
| SA122105 | - | + | - |
| SA242003 | - | + | - |
| SA211016 | - | + | - |
| SA231022 | - | + | - |
| **SA122111** | **+** | **+** | **+** |

-, holin gene not detected;

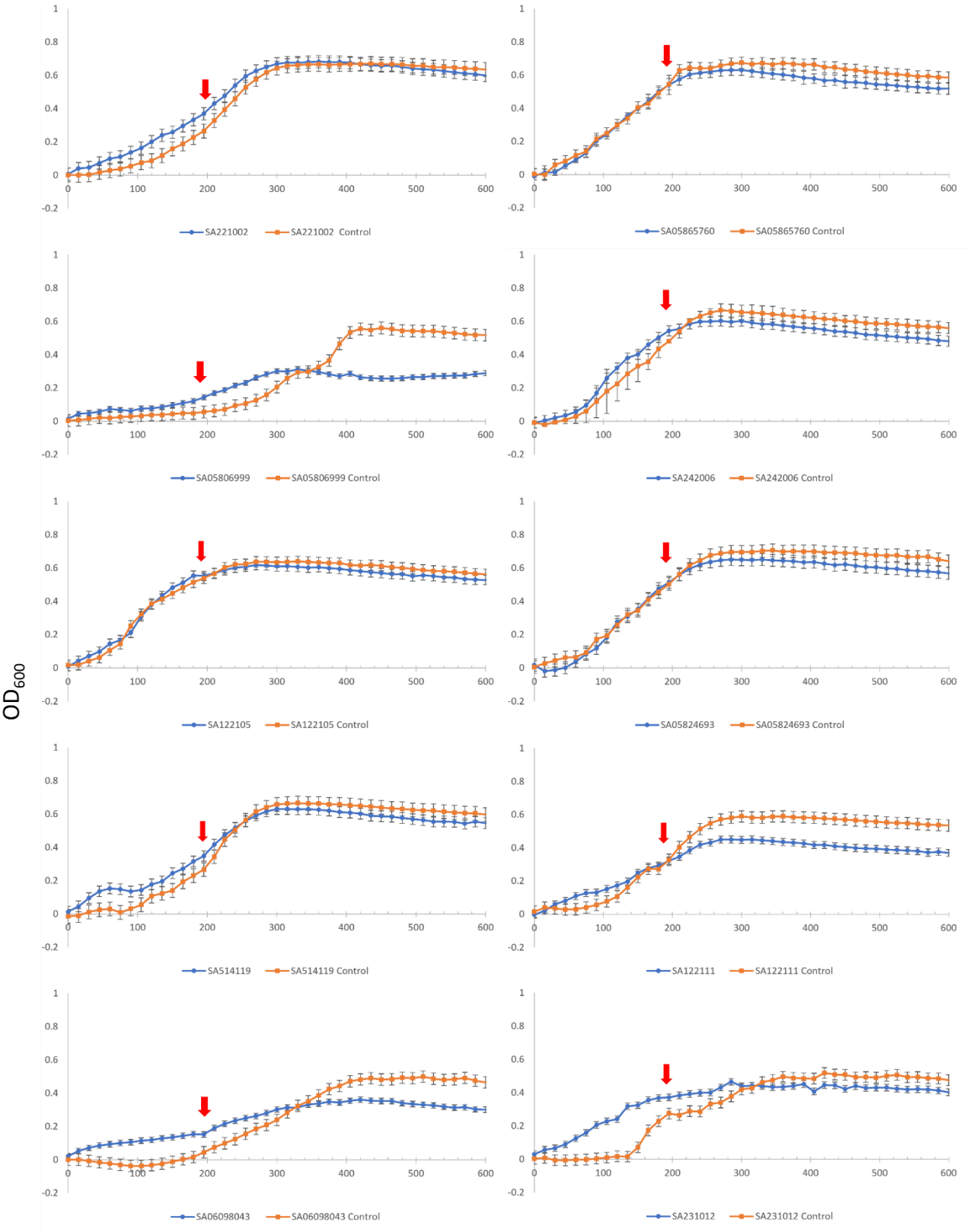
+, holin gene detected.

Table 12. PCR amplification of *C. difficile Myoviridae* and *Siphoviridae* holin genes in plaque-producing phage lysates.

| Isolate | *Myoviridae* | *Siphoviridae* |
|---|---|---|
| SA122111 | - | - |
| SA05806999 | - | - |
| SA06098043 | - | - |
| SA06148880 | + | - |

-, holin gene not detected;

+, holin gene detected.


OD600
Time (Minutes)
SA221002
SA221002 Control
SA05865760
SA05865760 Control
SA05806999
SA05806999 Control
SA242006
SA242006 Control
SA122105
SA122105 Control
SA05824693
SA05824693 Control
SA514119
SA514119 Control
SA122111
SA122111 Control
SA06098043
SA06098043 Control
SA231012
SA231012 Control

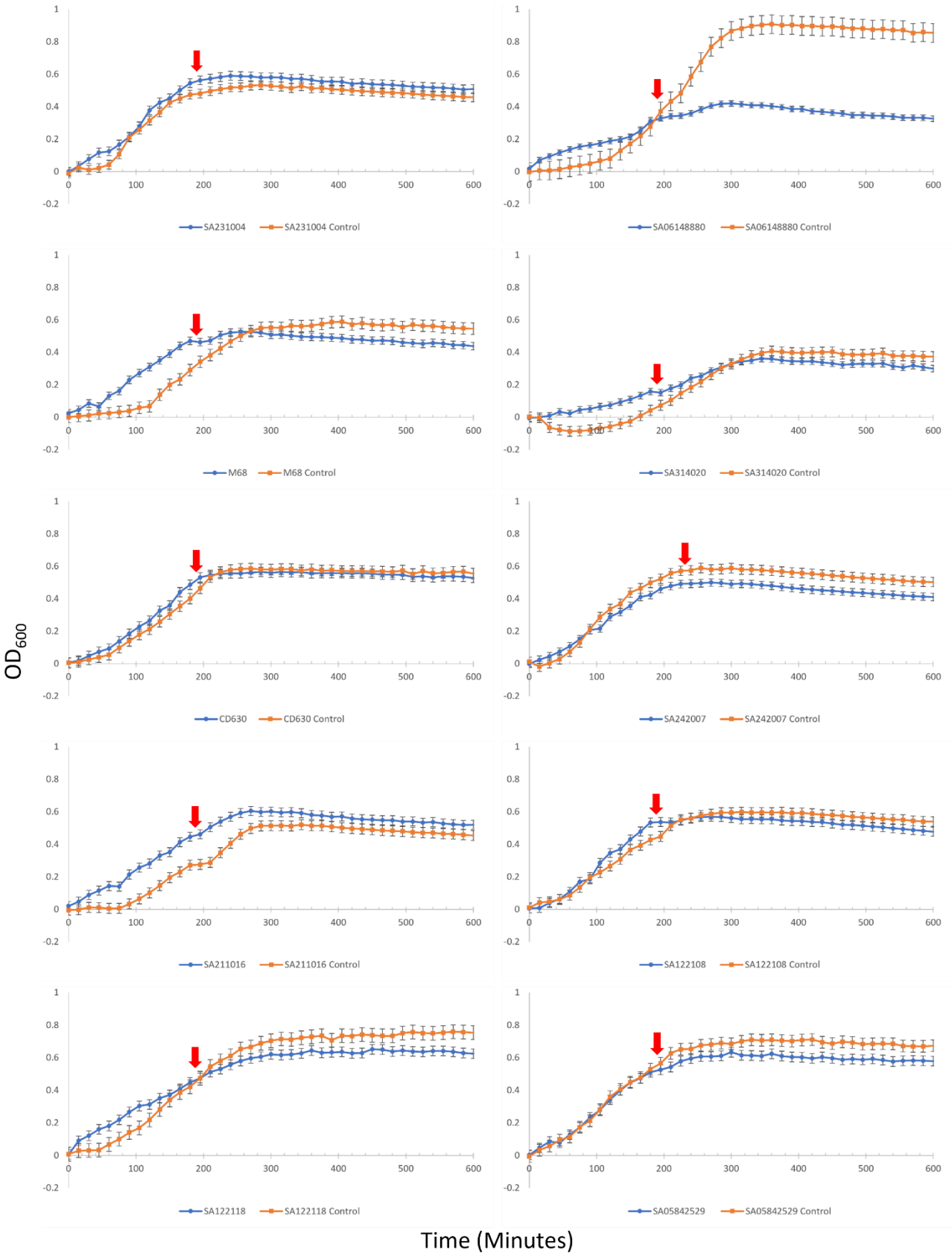



**Figure 15.** *C. difficile* growth after phage induction in liquid cultures. Optical density, representing bacterial cell density measured at 600nm wavelength ($OD_{600}$) were monitored before and after induction using mitomycin C (3µg/ml). Blue curves represent induced isolate; orange curves represent uninduced controls. Growth curves plotted using the mean $OD_{600}$. Error bars represent the standard deviation, which was determined using the $OD_{600}$ for each replicate. Arrows indicate time point at which mitomycin C was added (190 minutes).

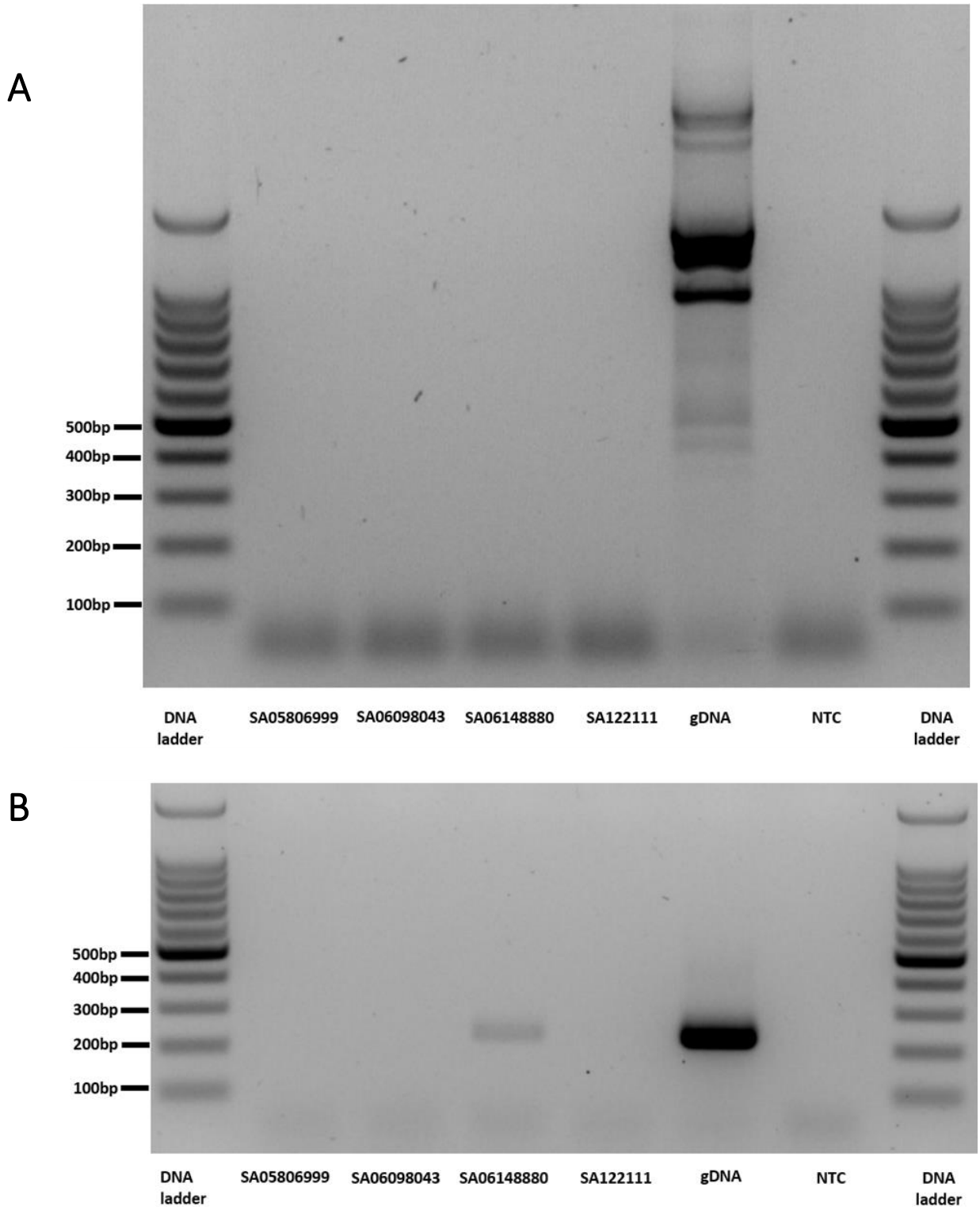



Figure 16. Agarose gels of phage holin gene PCR amplicons of *C. difficile* clinical isolates induced with mitomycin C. A) PCR for siphovirus-specific holin gene; no bands were observed of the expected 150 bp size; gDNA, SA221002 genomic DNA showed non-specific bands; DNA ladder, 100 bp molecular weight marker; NTC, non-template control. B) PCR for myovirus-specific holin gene; *C. difficile* SA06148880 showed a visible band with expected size of 227 bp; gDNA, *C. difficile* SA221002 genomic DNA, had a visible band of 227 bp.

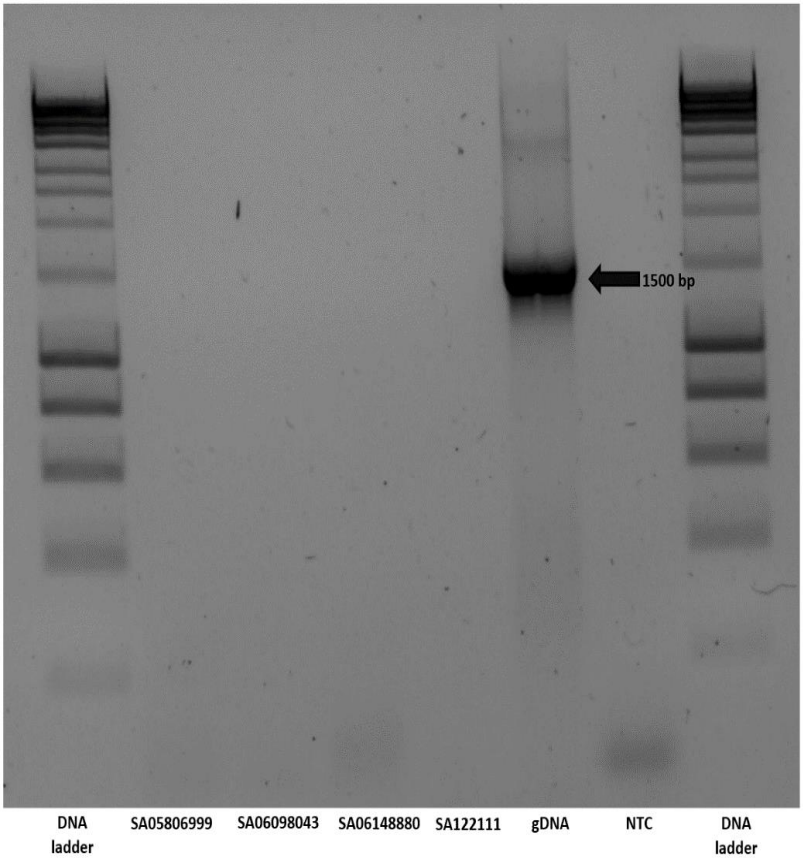



Figure 17. Confirmation of the absence of bacterial DNA. Agarose gel showing lack of 16s rRNA DNA amplification in DNAse I treated phage lysates. gDNA, *C. difficile* SA221002 genomic DNA as positive control; NTC, non-template control; DNA ladder, 1kb molecular weight marker, black arrow indicates region of expected band size of 1500 bp.

## DISCUSSION

In this study, phage induction using Mitomycin C generated plaques in one third of *C. difficile* isolates. Two different plaque morphologies were noted: small clear plaques and large opaque plaques. However, as only one inducing agent was tested, isolates that did not produce plaques should not be dismissed as non-phage carriers, as previous studies demonstrated that different inducing agents and varied concentrations of the same reagent may induce different phages in *C. difficile* (167, 169). Therefore, isolates tested in this study might produce phages if induced by a different agent, or different concentrations of the inducing agent. Besides mitomycin C, ultraviolet (UV) irradiation at 302nm and norfloxacin are other agents used in phage induction studies (167, 169, 187, 188, 239). Nale *et al.* (167) demonstrated the impact of induction efficacy with mitomycin C and norfloxacin at varied concentrations (0.3, 1, 3, 6 and 9 µg/ml). Although one isolate produced the highest abundance of phages at 6µg/ml, they reported

3µg/ml MMC was the optimal concentration required for induction. It was also observed that 61% of prophages were induced upon exposure to both MMC and norfloxacin, whereas 31% were only induced by induction with norfloxacin, and a further 8% from only mitomycin C exposure. Additionally, norfloxacin exposure resulted in production of phages with different morphologies (167). This highlights the impact of different inducing agents on phage induction.

A limiting factor in this study is the number of hosts used for phage induction, as only two test strains were used as suitable as hosts. This is relevant as *C. difficile* phages are reported as being strain specific (182, 311, 312). Using two test isolates could therefore have resulted in selection bias for the phages infecting the test strains. Further investigation must be done using a greater range of indicator isolates and using other inducing agents. When phages are induced from a set of strains, extensive screening for a suitable host must be done, and therefore the isolated phages are not an accurate representation of the true phage population (165, 211).

### Phage induction and plaque morphology

Zones of lysis observed in DAO and spot test assays were variable; either large (2-5mm) and turbid (**Figure 12**), or small (~1mm) and clear (**Figure 13 - 14**). Turbid plaques were observed in both DAO and spot test assays, while clear plaques were only seen in DAO assays. It was postulated previously (313) that clear plaques are typically produced by virulent phages, whilst turbid plaques are generally produced by temperate phages. However, *C. difficile* studies typically produced temperate phages (165-167, 169, 179, 186, 187, 193, 239). When turbid plaques previously were observed in studies, phage tail-like particles (PTLPs) were isolated from plaques (240). PTLPs are a class of bacteriocins, observed in various bacteria including *C. difficile* (166, 167, 240-242, 314). Two types of bacteriocins exist, R-type and F-type are known and encoded by the bacterium and resemble phage tails from myoviruses and siphoviruses (314). PTLPs are induced using the same methods as phage induction (166, 167, 169, 242). To date, *C. difficile* phage-tail like R-type bacteriocins, called diffocins, have been reported to be induced from most strains of *C. difficile* (167, 240) and shown to exhibit bactericidal activity against other strains of the same species (242). Diffocins have been shown to produce both partial and complete zones of clearing in non-producer *C. difficile* isolates. Interestingly, when non-producer *C. difficile* strains were challenged with diffocins, partial or total zones of clearing on bacterial lawns were seen (240). In our study, the two host isolates did not produce plaques

when infected with their self-derived lysates. This raises the possibility that the turbid plaques were produced by the activity of PTLPs.

### Liquid induction and identification of phages using holin-specific PCR assays

Phage induction was done in liquid cultures, using a subset of 20 isolates. Four isolates (SA05806999, SA06098043, SA06148880, SA122111) exhibited reduction in optical density following prophage induction, suggesting cell lysis and induction. Of these isolates, three (SA06098043, SA06148880, SA122111) also produced plaques previously. Isolates showing evidence of phage activity were subjected to conventional PCR targeting the holin genes which discriminates between the myovirus and siphovirus holin genes. Only one isolate (SA06148880) produced a positive PCR band corresponding to the amplicon size of a myovirus-specific holin gene. However, no amplicons were detected in the other isolates. A possible explanation for this could be that lysis was due to PTLPs, as the DNA encoding these proteins often lack holin genes (242, 314).

A limitation of liquid induction was the OD at which MMC was introduced. Induction was initiated at 195 minutes, but not all isolates had the same OD, and it ranged from 0.1 to 0.6. This can be improved by standardising the initial inoculum, to ensure all isolates are at the same OD at the timepoint when MMC is added. Furthermore, induction should be done at an OD of 0.15 – 0.1 to yield optimal results (239), otherwise the observed induction may be low or absent. Furthermore, plaque assays and spot tests should have been done as a confirmatory experiment for liquid inductions to validate the induction results. Future work should include MIC assays of different MMC concentrations to determine the optimal concentration for each isolate.

### *C. difficile* strain relatedness

*C. difficile* strains used in this study were characterised previously (97, 98, 267). From earlier MLVA data and unpublished work (provided by B Kullin via personal communication), two clusters of isolates which showed evidence of induction were identified as clonal (98, 267, 315). One cluster contains three isolates (SA122105, SA05806999, and SA06098043) and another contains two (SA231022 and SA242003). All isolates within the clusters produced the same induction results as other isolates within the cluster. Given that they are clonal, it is plausible that the induced phages (or PTLPs) are identical. In contrast, isolates SA08524693, SA06148880, SA122111, and SA211016 did not cluster with other induction-positive isolates,

and each clinical sample was collected from a separate hospital. Of these, only SA06148880 generated a phage PCR band. Although the genomes of SA06148880 and SA122111 indicate close relationship, these isolates produced different results; whereby only SA06148880 produced a PCR band for a myovirus. This divergence in outcomes suggests that genetic variation between isolates may contribute to the observed differences in phage carriage. Future work should be done to verify that of the nine isolates showing lytic activity when induced, five were genetically distinct, and potentially harbour five different temperate phages.

## CONCLUSION

Among the *C. difficile* isolates tested, one-third displayed MMC-induced phage activity. Of those, most isolates showed induction in either the plaque or liquid induction assay, while only a few exhibited activity in both assays. Subsequent PCR assays could only successfully verify and characterise one phage as a myovirus, originating from the lysate of SA06148880. While other isolates showed evidence of phage activity, additional confirmatory testing is needed. Additionally, the remaining phages could not be characterised. Furthermore, the presence of PTLPs was not ruled out. Future studies should include TEM to verify the presence of phages in lysates showing lytic activity, while also using additional inducing agents and a wider selection of host isolates.

# CHAPTER 5: GENERAL CONCLUSIONS

The rise in antibiotic resistance necessitates the development of newer drugs or alternative therapies such as phage therapy. Phages characterised by strict virulence are often considered ideal for therapy due to their assured lytic activity and the decreased chance of transduction of virulence or antibiotic resistance genes (316). However, it is worth noting that many *C. difficile* phages are known to be temperate (317). Despite this, temperate phages have demonstrated efficacy in clearing CDI in both monophage therapy (215, 216) and phage

cocktails (183, 196). Though efficacious in gut and animal models (183, 215, 217), more research must be done to advance to human clinical trials.

Treatment with temperate phages requires further research to understand the relationship between phages and their human host, bacterial host specificity, the potential for transfer of genetic material, and *in vivo* conditions that modulate the switch between lysogeny and lytic phases. The approach of engineering temperate phages to become lytic has shown success and could be adapted for *C. difficile* phages to treat CDI (318). Notably, studies must be done to understand the human immune response to phages and potential mass lysis of bacterial cells; such studies are essential to ensure safety in use.

In the context of *C. difficile*, several areas still require deeper understanding. Firstly, disparities observed owing to the geographical distribution of *C. difficile* ribotypes, combined with the limited phages host range, imply that locally sourced phages would likely provide the most effective approach for CDI treatment in a specific population. Achieving this would require studies to be done across various geographic regions to understand the phage-host distribution, diversity, and specificity of isolates within the locale. This study sought to bridge this information gap, focusing specifically on the sub-Saharan region, particularly South Africa.

The primary goal of Chapter 2 was to ascertain the presence of phages within local wastewater systems, potentially serving as a local phage repository. This was attempted through analysis of a single wastewater sample, although it's important to acknowledge that this was not the central focus of this study. A single plaque was observed, and while attempts were made to replicate it, efforts were unsuccessful. Furthermore, PCR validations of the plaque yielded negative results. The absence of lytic phages observed in this study is in line with the existing body of knowledge, given that *C. difficile* reported in the current published literature are temperate. The anaerobic and spore-forming characteristics of *C. difficile* are inclined to favour lysogenic cycles over lytic ones, thus likely promoting the prevalence of temperate phages (186, 216, 319). Considering *C. difficile*'s propensity for sporulation, the isolation of free phages from the environment becomes less likely. Additionally, the study utilised only two *C. difficile* host isolates, a limitation that potentially introduced a bias toward these hosts, possibly impeding phage detection. While our investigation of sewage samples offered preliminary indications of phage activity in local sewage systems; these findings were inconclusive and require validation. Consequently, additional studies are required, using raw sewage collected

at different points and incorporating a larger pool of host isolates. TEM should be included to confirm that plaque production resulted from the activity of phages. Moreover, incorporating multiple rounds of amplification before doing overlay assays and utilizing rolling circle amplification for PCR assays.

A high-throughput bioinformatic approach was employed to identify integrated phages within the genomic sequences of a representative set of South African clinical *C. difficile* isolates (n=58). Using PHASTER, prophages were identified in majority of isolates (n=54). Cross-validation of PHASTER's predictions against VirSorter2 yielded complete concordance, corroborating the presence of all prophages identified by PHASTER.

Multiple bioinformatic tools were utilised to increase robustness of bioinformatic predictions. PHASTER employs a signature-based approach which and compares genomic features to a database of known prophages (256). VirSorter2, designed for draft genomes and fragmented metagenomes, excels at identifying diverse viral groups, including novel types potentially missed by PHASTER's database dependence (258). Its multi-classifier and modular design allow for adaptation and expansion (258), while PHASTER relies on a single classifier and a fixed database. This illustrates the value of a combined approach ideally using complementary tools that have different underlying detection methods to determine the presence of prophages within bacterial hosts.

Predicted phage genomes were clustered into six distinct groups. Notable variations were evident, with certain functional modules lacking in some groups. Overall, only modules for DNA packaging, capsid morphogenesis, and tail morphogenesis were universally present. When compared with completed genome maps of *C. difficile* phages, all genes can be categorized into functional modules (packing, structural, attachment, lysis, lysogenic conversion or lysogeny control, DNA replication) (180, 185, 193, 266, 320). However, due to limited sequence data, it could not be determined whether the absent modules resulted from assembly truncation, or if they were genuinely absent in specific phage groups. Challenges such as these may arise when working with draft genomes. While closed genomes are ideal, their creation demands additional resources and labour. Nevertheless, ongoing progress in sequencing technologies and phage identification tools is being made, and such limitations will present as less of a challenge in future.

*C. difficile* isolates were shown to display a diverse repertoire of anti-phage systems, including the adaptive immunity conferred by CRISPR-Cas type I-B which has been previously reported (174, 290). Additional systems were identified, including the AbiD and AbiU, RM type I and IV systems, alongside other recently discovered systems, namely Kiwa, PD-λ-1, Lamasu, and Gabija systems. However, many of the systems predicted have not been experimentally confirmed in *C. difficile*. The two host isolates displayed distinct repertoires of anti-phage systems, potentially contributing to differences in sensitivity. Further research is needed to validate the outcomes of the bioinformatic analysis, and to understand the underlying mechanisms of these systems, given their potential to influence susceptibility to phage infection. In addition to this, screening the phage genomes for the presence of bacterial countermeasures would provide insight on the interplay between host and phage. These may include anti-CRISPR proteins (321), phage-encoded CRISPR-Cas arrays (173, 210), gene mutations that prevent Abi system activation (322), RM evasion strategies such as base modification or decreasing the number of restriction sites (323, 324), among others. Since all bacterial isolates carried defence systems, it is possible that the infection of the host strain is more dependent on the phage's ability to evade host immunity, rather than the host's lack of defence systems, but this possibility requires investigation.

Biological experiments were conducted to validate the outcomes of the bioinformatic analysis. To this end, subsets of isolates were utilised in plaque assays and liquid induction assays. Among the induced isolates, the majority displayed evidence of phage induction in either of the two assays, with a smaller portion exhibiting induction in both plaque and liquid induction assays. Two distinct plaque morphologies were observed: small clear plaques and large opaque plaques.

Although the study originally aimed to employ TEM for phage morphological characterisation, sufficient phage titres were not obtained, and TEM was therefore abandoned. Despite this, future investigations might still consider TEM for exploration of morphological features, but also the potential presence of PTLPs.

While conventional PCR targeting the holin genes of myoviruses and siphoviruses was employed to confirm phage presence in plaques and lysates, conclusive characterisation was only achieved for one isolate, identified as a myovirus. The limited success of PCR amplification in other isolates requires further investigation. Lack of primer homology with target regions in

these isolates and potentially insufficient phage DNA could be possible explanations. Notably, the absence of a positive control for siphoviruses hinders definitive conclusions. Without confirmation of their presence, the possibility of siphoviruses being present cannot be excluded, but neither confirmed. Moving forward, including a siphovirus positive control, optimizing primers, exploring alternative targets, and employing multiple rounds of phage amplification offer more comprehensive results.

The biological experiments aligned with PHASTER's predictions of the isolates deemed phage-free. However, most of the isolates which PHASTER identified as prophage carriers did not show evidence of induction. Nonetheless, this does not discredit PHASTER's accuracy; as only a one inducing agent and two host strains were used, which potentially limited the outcome of biological experiments. The narrow host range has been reported before, where Phothichaisri *et al* screened four phages against a panel of 92 *C. difficile* test strains, and found the six susceptible isolates (6.5%) (170), Sekulovic *et al* had tested the host range of ΦCD38-2 on 207 *C. difficile* isolates, and found that 48% was susceptible to infection (188). Similar findings regarding the host ranges have been reported in other studies (166, 167, 169, 192).

Further experiments are needed to show whether these phages are induced using other inducing agents, such as UV irradiation or antibiotic treatment. Liquid induction experiments should be conducted with a lower starting OD of 0.1 - 0.15. Furthermore, due to the narrow host range of phages, number of test isolates should be expanded to include a diverse range of isolates belonging to RT 017.

The work presented in this thesis allows for future research into the use of phages and phage-based therapeutic interventions to treat CDI. It informs on the potential of treated sewage as a phage reservoir. It also identified a diverse phage population of within South African *C. difficile* RT 017 isolates, and clustered them into distinct genetic groups, and explored the potential mosaic structure of each group. Furthermore, the anti-phage systems present in *C. difficile* isolates were investigated and shown to display a diverse arsenal of defence systems. Additionally, this study provides preliminary evidence for a possible novel phage as well. Future studies should explore this potential novel phage, as well as the bacterial defence systems observed in local isolates. Furthermore, this study reports on MMC for phage induction using double agar over assays, spot test assays, and liquid induction assays. This would contribute

knowledge towards development of phage collections, with characterised phages potentially valuable for CDI phage therapy.

## SUPPLEMENTARY MATERIAL

**Table S1.** *C. difficile* strains regions containing identified prophage with respective reference phages.

| Strain | Region | Start position | End position | Phage (No. matching genes) | NCBI Accession number |
|---|---|---|---|---|---|
| SA111031 | 9 | 9473 | 45656 | phiCDHM19(15) | NC_028996 |
| | | | | phiCD119(11) | NC_007917 |
| | 10 | 51 | 29616 | phiMMP02(23) | NC_019421 |
| | | | | phiCD27(22) | NC_011398 |
| | | | | phiCD505(21) | NC_028764 |
| | | | | CDKM9(20) | NC_048642 |
| | | | | CDKM15(19) | NC_048643 |
| SA111045 | 8 | 9971 | 46552 | phiCDHM19(15) | NC_028996 |
| | | | | phiCD119(12) | NC_007917 |
| SA111146 | 9 | 1 | 36582 | phiCDHM19(15) | NC_028996 |
| | | | | phiCD119(12) | NC_007917 |
| SA122009 | 9 | 87 | 21742 | phiCDHM19(26) | NC_028996 |
| | | | | phiCD119(21) | NC_007917 |
| SA122025 | 3 | 318580 | 391992 | phiCDHM19(53) | NC_028996 |
| | | | | phiCD119(35) | NC_007917 |
| | 6 | 61329 | 127604 | phiMMP01(31) | NC_028883 |
| | | | | phiC2(29) | NC_009231 |
| | | | | phiMMP03(27) | NC_028959 |
| | | | | CDMH1(25) | NC_024144 |
| | 10 | 9473 | 46054 | phiCDHM19(15) | NC_028996 |
| | | | | phiCD119(12) | NC_007917 |
| SA122029 | 8 | 216 | 21910 | phiCDHM19(28) | NC_028996 |
| | | | | phiCD119(21) | NC_007917 |
| SA122032 | 8 | 9971 | 46552 | phiCDHM19(15) | NC_028996 |
| | | | | phiCD119(12) | NC_007917 |
| SA122033 | 9 | 9971 | 46552 | phiCDHM19(15) | NC_028996 |
| | | | | phiCD119(12) | NC_007917 |
| SA122034 | 8 | 731 | 47599 | phiCDHM19(17) | NC_028996 |
| | | | | phiCD119(13) | NC_007917 |
| | 10 | 233 | 21888 | phiCDHM19(28) | NC_028996 |
| | | | | phiCD119(21) | NC_007917 |
| SA122048 | 8 | 8465 | 45046 | phiCDHM19(15) | NC_028996 |
| | | | | phiCD119(12) | NC_007917 |
| | 10 | 216 | 21910 | phiCDHM19(28) | NC_028996 |
| | | | | phiCD119(21) | NC_007917 |
| SA122105 | 9 | 9698 | 48039 | phiCDHM19(16) | NC_028996 |
| | | | | phiCD119(11) | NC_007917 |

| Strain | Region | Start position | End position | Phage (No. matching genes) | NCBI Accession number |
|---|---|---|---|---|---|
| | 10 | 154 | 30031 | phiMMP02(22) | NC_019421 |
| | | | | phiCD27(21) | NC_011398 |
| | | | | phiCD505(20) | NC_028764 |
| | | | | CDKM9(20) | NC_048642 |
| | | | | CDKM15(18) | NC_048643 |
| SA122108 | 9 | 731 | 47599 | phiCDHM19(17) | NC_028996 |
| | | | | phiCD119(13) | NC_007917 |
| | 11 | 216 | 21910 | phiCDHM19(28) | NC_028996 |
| | | | | phiCD119(21) | NC_007917 |
| SA122111 | 8 | 9473 | 48388 | phiCDHM19(16) | NC_028996 |
| | | | | phiCD119 (11) | NC_007917 |
| | 11 | 153 | 15768 | phiCDHM11(18) | NC_029001 |
| | | | | phiCDHM14(18) | NC_048665 |
| | | | | phiCD506(15) | NC_028838 |
| | | | | phiCDHM13(15) | NC_029116 |
| | | | | phiMMP04(14) | NC_019422 |
| | 12 | 3 | 14622 | phiCDHM11(19) | NC_029001 |
| | | | | phiCDHM14(19) | NC_048665 |
| | | | | phiCD506(16) | NC_028838 |
| | | | | phiCD481_1(15) | NC_028951 |
| | | | | phiCDHM13(14) | NC_029116 |
| SA122118 | 9 | 8682 | 44868 | phiCDHM19(15) | NC_028996 |
| | | | | phiCD119(11) | NC_007917 |
| | 12 | 216 | 24000 | phiCDHM19(27) | NC_028996 |
| | | | | phiCD119(22) | NC_007917 |
| SA122128 | 9 | 8465 | 45046 | phiCDHM19(15) | NC_028996 |
| | | | | phiCD119(12) | NC_007917 |
| | 11 | 233 | 21888 | phiCDHM19(28) | NC_028996 |
| | | | | phiCD119(21) | NC_007917 |
| SA131136 | 9 | 9473 | 46054 | phiCDHM19(15) | NC_028996 |
| | | | | phiCD119(12) | NC_007917 |
| SA131144 | 8 | 9473 | 46054 | phiCDHM19(15) | NC_028996 |
| | | | | phiCD119(12) | NC_007917 |
| | 10 | 233 | 21888 | phiCDHM19(28) | NC_028996 |
| | | | | phiCD119(21) | NC_007917 |
| SA141139 | 9 | 9443 | 46024 | phiCDHM19(15) | NC_028996 |
| | | | | phiCD119(12) | NC_007917 |
| SA211015 | 9 | 1 | 36582 | phiCDHM19(15) | NC_028996 |
| | | | | phiCD119(12) | NC_007917 |
| SA211016 | 9 | 449 | 36635 | phiCDHM19(15) | NC_028996 |
| | | | | phiCD119(11) | NC_007917 |
| | 11 | 653 | 34246 | phiMMP02(24) | NC_019421 |
| | | | | phiCD27(23) | NC_011398 |
| | | | | phiCD505(23) | NC_028764 |
| | | | | CDKM15(22) | NC_048643 |
| | | | | CDKM9(20) | NC_048642 |
| | | | | CDMH1(13) | NC_024144 |
| | 12 | 188 | 34039 | phiMMP02(24) | NC_019421 |
| | | | | phiCD27(23) | NC_011398 |
| | | | | phiCD505(23) | NC_028764 |
| | | | | CDKM15(22) | NC_048643 |
| | | | | CDKM9(20) | NC_048642 |

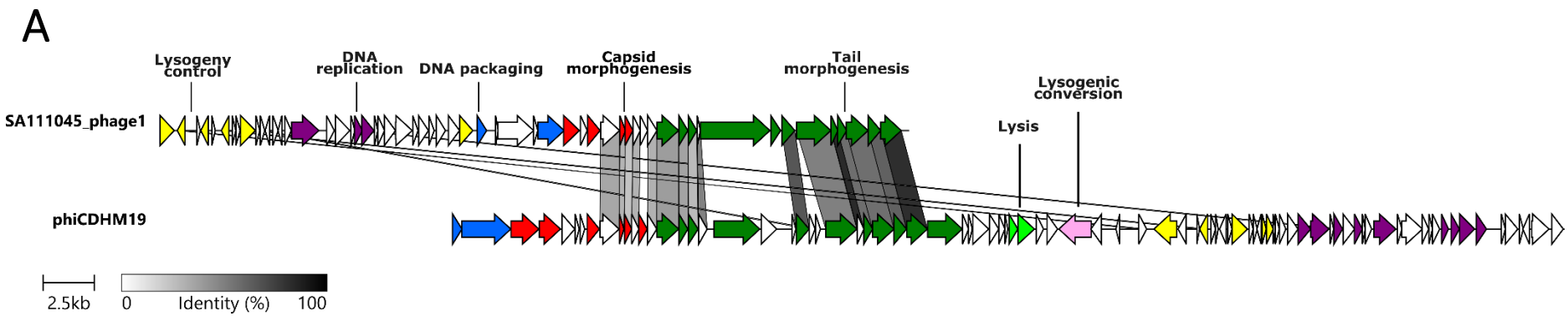
| Strain | Region | Start position | End position | Phage (No. matching genes) | NCBI Accession number |
|---|---|---|---|---|---|
| | | | | CDMH1(13) | NC_024144 |
| | 13 | 4310 | 31824 | phiMMP01(24) | NC_028883 |
| | | | | phiMMP03(23) | NC_028959 |
| | | | | phiC2(23) | NC_009231 |
| | | | | CDMH1(22) | NC_024144 |
| SA211023 | 9 | 1 | 36582 | phiCDHM19(15) | NC_028996 |
| | | | | phiCD119(12) | NC_007917 |
| SA211024 | 9 | 10072 | 46653 | phiCDHM19(15) | NC_028996 |
| | | | | phiCD119(12) | NC_007917 |
| SA211027 | 9 | 731 | 35915 | phiCDHM19(14) | NC_028996 |
| | | | | phiCD119(11) | NC_007917 |
| | 10 | 154 | 30031 | phiMMP02(22) | NC_019421 |
| | | | | phiCD27(21) | NC_011398 |
| | | | | CDKM9(20) | NC_048642 |
| | | | | phiCD505(20) | NC_028764 |
| | | | | CDKM15(18) | NC_048643 |
| SA211035 | 9 | 9473 | 46054 | phiCDHM19(15) | NC_028996 |
| | | | | phiCD119(12) | NC_007917 |
| SA221002 | | | | No intact phages | |
| SA221014 | 5 | 51 | 30233 | phiMMP02(23) | NC_019421 |
| | | | | phiCD27(22) | NC_011398 |
| | | | | phiCD505(21) | NC_028764 |
| | | | | CDKM9(20) | NC_048642 |
| | | | | CDKM15(19) | NC_048643 |
| SA231004 | 8 | 93 | 39100 | phiCDHM19(16) | NC_028996 |
| | | | | phiCD119(11) | NC_007917 |
| | 11 | 3 | 15618 | phiCDHM14(18) | NC_048665 |
| | | | | phiCDHM11(18) | NC_029001 |
| | | | | phiCD506(15) | NC_028838 |
| | | | | phiCDHM13(15) | NC_029116 |
| | | | | phiMMP04(14) | NC_019422 |
| | | | | phiCD481_1(12) | NC_028951 |
| | 12 | 188 | 14807 | phiCDHM14(19) | NC_048665 |
| | | | | phiCDHM11(19) | NC_029001 |
| | | | | phiCD506(16) | NC_028838 |
| | | | | phiCD481_1(15) | NC_028951 |
| | | | | phiCDHM13(14) | NC_029116 |
| SA231012 | | | | No intact phages | |
| SA231021 | 9 | 1 | 36582 | phiCDHM19(15) | NC_028996 |
| | | | | phiCD119(12) | NC_007917 |
| SA231022 | 9 | 1 | 36582 | phiCDHM19(15) | NC_028996 |
| | | | | phiCD119(12) | NC_007917 |
| SA231040 | 9 | 10713 | 47294 | phiCDHM19(15) | NC_028996 |
| | | | | phiCD119(12) | NC_007917 |
| SA242003 | 7 | 3 | 30038 | phiMMP02(23) | NC_019421 |
| | | | | phiCD27(22) | NC_011398 |
| | | | | phiCD505(21) | NC_028764 |
| | | | | CDKM9(20) | NC_048642 |
| | | | | CDKM15(19) | NC_048643 |
| SA242006 | | | | No intact phages | |
| SA242007 | 9 | 9852 | 46433 | phiCDHM19(15) | NC_028996 |
| | | | | phiCD119(12) | NC_007917 |

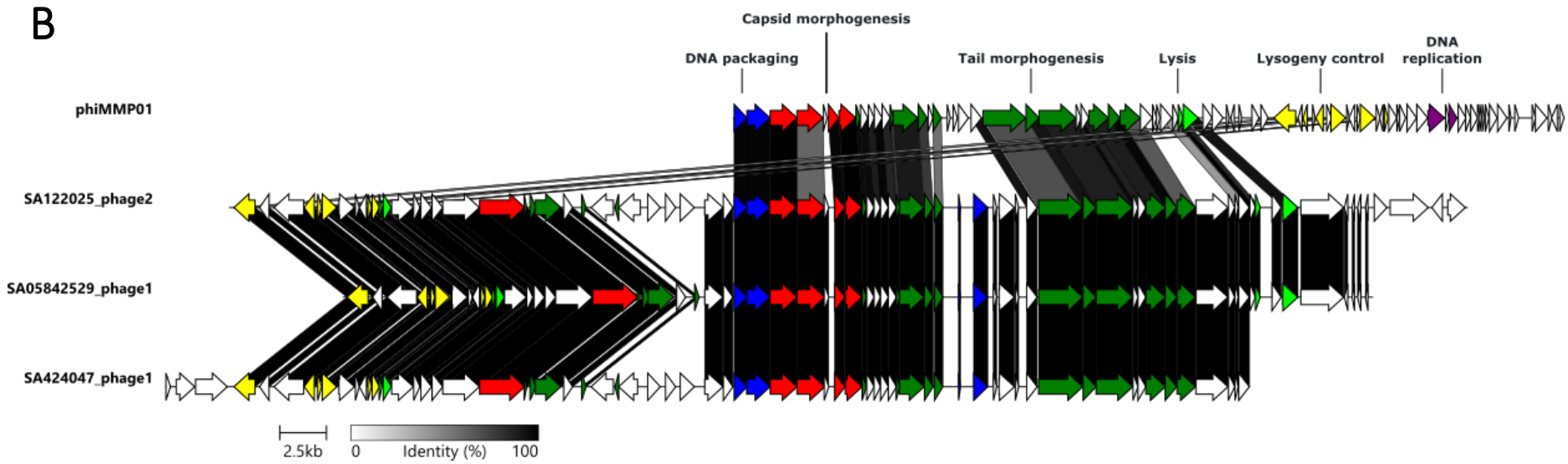
| Strain | Region | Start position | End position | Phage (No. matching genes) | NCBI Accession number |
|---|---|---|---|---|---|
| | 11 | 233 | 21888 | phiCDHM19(28) | NC_028996 |
| | | | | phiCD119(21) | NC_007917 |
| SA242017 | 9 | 9473 | 46054 | phiCDHM19(15) | NC_028996 |
| | | | | phiCD119(12) | NC_007917 |
| | 12 | 1 | 26246 | phiCDHM19(34) | NC_028996 |
| | | | | phiCD119(25 | NC_007917 |
| SA242037 | 4 | 233 | 28947 | phiCDHM19(35) | NC_028996 |
| | | | | phiCD119(27) | NC_007917 |
| SA242038 | 8 | 1 | 36582 | phiCDHM19(15) | NC_028996 |
| | | | | phiCD119(12) | NC_007917 |
| | 10 | 233 | 23663 | phiCDHM19(30) | NC_028996 |
| | | | | phiCD119(23) | NC_007917 |
| SA251026 | 8 | 93 | 39100 | phiCDHM19(16) | NC_028996 |
| | | | | phiCD119(11) | NC_007917 |
| | 11 | 153 | 15768 | phiCDHM14(18) | NC_048665 |
| | | | | phiCDHM11(18) | NC_029001 |
| | | | | phiCD506(15) | NC_028838 |
| | | | | phiCDHM13(15) | NC_029116 |
| | | | | phiMMP04(14) | NC_019422 |
| | | | | phiCD481_1(12) | NC_028951 |
| | 12 | 3 | 14622 | phiCDHM11(19) | NC_029001 |
| | | | | phiCDHM14(19) | NC_048665 |
| | | | | phiCD506(16) | NC_028838 |
| | | | | phiCD481_1(15) | NC_028951 |
| | | | | phiCDHM13(14) | NC_029116 |
| SA251042 | 10 | 9473 | 46054 | phiCDHM19(15) | NC_028996 |
| | | | | phiCD119(12) | NC_007917 |
| | 11 | 485 | 23613 | phiCDHM19(27) | NC_028996 |
| | | | | phiCD119(21) | NC_007917 |
| SA314020 | 9 | 9473 | 46054 | phiCDHM19(15) | NC_028996 |
| | | | | phiCD119(12) | NC_007917 |
| | 10 | 764 | 32071 | phiCD27(24) | NC_011398 |
| | | | | phiMMP02(22) | NC_019421 |
| | | | | CDKM9(22) | NC_048642 |
| | | | | phiCD505(22) | NC_028764 |
| | | | | CDKM15(20) | NC_048643 |
| | 12 | 371 | 23792 | phiCDHM19(30) | NC_028996 |
| | | | | phiCD119(23) | NC_007917 |
| SA414113 | 10 | 731 | 35915 | phiCDHM19(14) | NC_028996 |
| | | | | phiCD119(11) | NC_007917 |
| | 13 | 216 | 24000 | phiCDHM19(27) | NC_028996 |
| | | | | phiCD119(22) | NC_007917 |
| SA414130 | 10 | 9443 | 46024 | phiCDHM19(15) | NC_028996 |
| | | | | phiCD119(12) | NC_007917 |
| SA424047 | 1 | 93 | 58073 | phiMMP01(30) | NC_028883 |
| | | | | phiMMP03(27) | NC_028959 |
| | | | | phiC2(27) | NC_009231 |
| | | | | CDMH1(24) | NC_024144 |
| | 8 | 449 | 48382 | phiCDHM19(18) | NC_028996 |
| | | | | phiCD119(13) | NC_007917 |
| | 11 | 653 | 36072 | phiMMP02(26) | NC_019421 |
| | | | | CDKM15(25) | NC_048643 |

| Strain | Region | Start position | End position | Phage (No. matching genes) | NCBI Accession number |
|---|---|---|---|---|---|
| | | | | phiCD505(24) | NC_028764 |
| | | | | phiCD27(23) | NC_011398 |
| | | | | CDKM9(20) | NC_048642 |
| | 13 | 3 | 15618 | phiCDHM11(18) | NC_029001 |
| | | | | phiCDHM14(18) | NC_048665 |
| | | | | phiCD506(15) | NC_028838 |
| | | | | phiCDHM13(15) | NC_029116 |
| | | | | phiMMP04(14) | NC_019422 |
| | | | | phiCD481_1(12) | NC_018951 |
| | 14 | 188 | 14807 | phiCDHM11(19) | NC_029001 |
| | | | | phiCDHM14(19) | NC_048665 |
| | | | | phiCD506(16) | NC_028838 |
| | | | | phiCD481_1(15) | NC_018951 |
| | | | | phiCDHM13(14) | NC_029116 |
| SA514119 | | | | No intact phages | |
| SA613110 | 10 | 1 | 36582 | phiCDHM19(15) | NC_028996 |
| | | | | phiCD119(12) | NC_007917 |
| | 12 | 620 | 24365 | phiCDHM19(27) | NC_028996 |
| | | | | phiCD119(22) | NC_007917 |
| SA714001 | 6 | 1 | 45912 | phiCDHM19(32) | NC_028996 |
| | | | | phiCD119(25) | NC_007917 |
| | | | | CDKM15(19) | NC_048643 |
| | | | | phiMMP03(18) | NC_028959 |
| SA714043 | 2 | 574405 | 622944 | phiMMP01(30) | NC_028883 |
| | | | | phiC2(27) | NC_009231 |
| | | | | phiMMP03(27) | NC_028959 |
| | | | | CDMH1(24) | NC_024144 |
| | 8 | 1856 | 48586 | phiCDHM19(17) | NC_028996 |
| | | | | phiCD119(14) | NC_007917 |
| SA814041 | 4 | 196750 | 245289 | phiMMP01(30) | NC_028883 |
| | | | | phiMMP03(27) | NC_028959 |
| | | | | phiC2(27) | NC_009231 |
| | | | | CDMH1(24) | NC_024144 |
| | 8 | 1 | 36582 | phiCDHM19(15) | NC_028996 |
| | | | | phiCD119(12) | NC_007917 |
| SA05806999 | 9 | 528 | 35712 | phiCDHM19(14) | NC_028996 |
| | | | | phiCD119(11) | NC_007917 |
| SA05824693 | 8 | 334 | 35518 | phiCDHM19(14) | NC_028996 |
| | | | | phiCD119(11) | NC_007917 |
| | 9 | 3 | 29328 | phiMMP02(23) | NC_019421 |
| | | | | phiCD27(22) | NC_011398 |
| | | | | phiCD505(21) | NC_028764 |
| | | | | CDKM9(20) | NC_048642 |
| | | | | CDKM15(19) | NC_048643 |
| SA05842529 | 3 | 196618 | 251724 | phiMMP01(31) | NC_028883 |
| | | | | phiC2(29) | NC_009231 |
| | | | | phiMMP03(27) | NC_028959 |
| | | | | CDMH1(25) | NC_024144 |
| SA05845556 | 7 | 56 | 36242 | phiCDHM19(15) | NC_028996 |
| | | | | phiCD119(11) | NC_007917 |
| | 10 | 144 | 21838 | phiCDHM19(28) | NC_028996 |
| | | | | phiCD119(21) | NC_007917 |

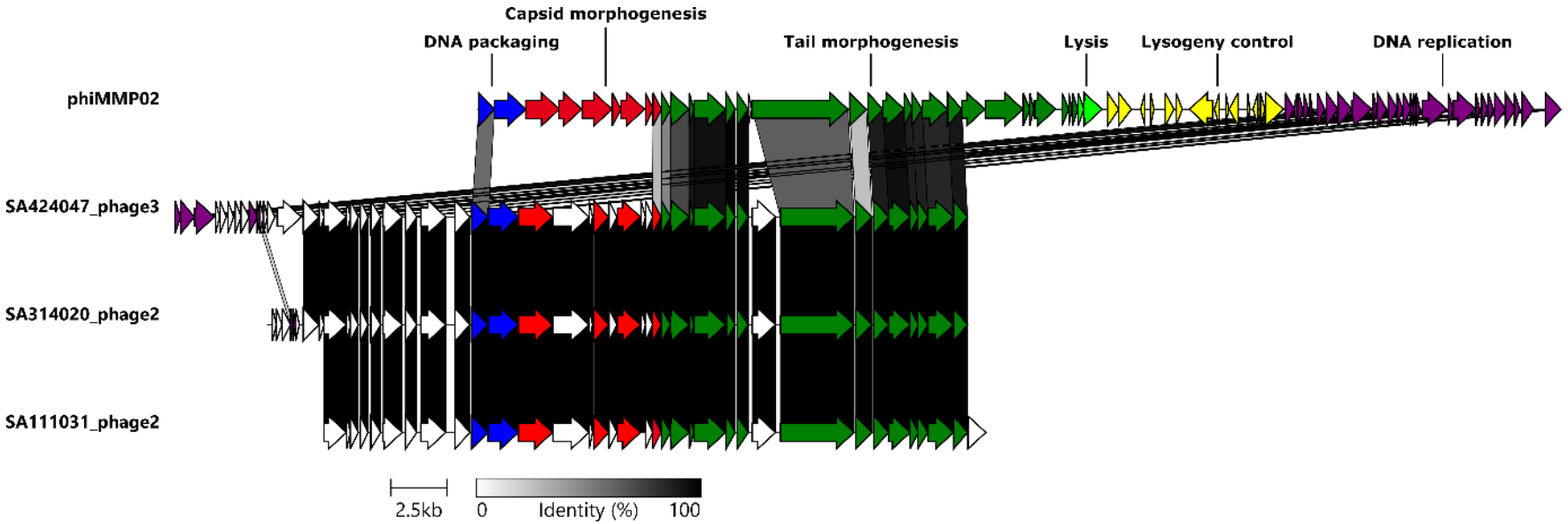
| Strain | Region | Start position | End position | Phage (No. matching genes) | NCBI Accession number |
|---|---|---|---|---|---|
| SA05845567 | 7 | 528 | 35712 | phiCDHM19(14) | NC_028996 |
| | | | | phiCD119(11) | NC_007917 |
| | 8 | 1 | 29326 | phiMMP02(22) | NC_019421 |
| | | | | phiCD27(21) | NC_011398 |
| | | | | phiCD505(21) | NC_028764 |
| | | | | CDKM9(20) | NC_048642 |
| | | | | CDKM15(18) | NC_048643 |
| SA05864722 | 7 | 528 | 35712 | phiCDHM19(14) | NC_028996 |
| | | | | phiCD119(11) | NC_007917 |
| | 9 | 3 | 29328 | phiMMP02(23) | NC_019421 |
| | | | | phiCD27(22) | NC_011398 |
| SA05865760 | 7 | 334 | 36843 | phiCDHM19(15) | NC_028996 |
| | | | | phiCD119(12 | NC_007917 |
| | 8 | 2 | 25968 | phiMMP02(21) | NC_019421 |
| | | | | CDKM9(19) | NC_048642 |
| | | | | phiCD27(19) | NC_011398 |
| | | | | phiCD505(18) | NC_028764 |
| | | | | CDKM15(15) | NC_048643 |
| | 9 | 144 | 21838 | phiCDHM19(28) | NC_028996 |
| | | | | phiCD119(21) | NC_007917 |
| SA06098043 | 8 | 9238 | 44422 | phiCDHM19(14) | NC_028996 |
| | | | | phiCD119(11) | NC_007917 |
| | 9 | 82 | 29959 | phiMMP02(22) | NC_019421 |
| | | | | phiCD27(21) | NC_011398 |
| | | | | phiCD505(20) | NC_028764 |
| | | | | CDKM9(20) | NC_048642 |
| | | | | CDKM15(18) | NC_048643 |
| SA06148880 | 4 | 570577 | 628318 | phiMMP01(29) | NC_028883 |
| | | | | phiC2(27) | NC_009231 |
| | | | | phiMMP03(27) | NC_028959 |
| | | | | CDMH1(24) | NC_024144 |
| | 9 | 334 | 38672 | phiCDHM19(16) | NC_028996 |
| | | | | phiCD119(11) | NC_007917 |
| | 10 | 2 | 29150 | phiCDHM19(36) | NC_028996 |
| | | | | phiCD119(26) | NC_007917 |
| | | | | CDMH1(11) | NC_024144 |
| SA06163235 | 8 | 334 | 36520 | phiCDHM19(15) | NC_028996 |
| | | | | phiCD119(11) | NC_007917 |
| | | | | phiCD506(7) | NC_028838 |
| | 9 | 3 | 29328 | phiMMP02(23) | NC_019421 |
| | | | | phiCD27(22) | NC_011398 |
| | | | | phiCD505(21) | NC_028764 |
| | | | | CDKM9(20) | NC_048642 |
| | | | | CDKM15(19) | NC_048643 |
| CF5 | 4 | 1707633 | 1763916 | phiMMP03(30) | NC_028959 |
| | | | | phiC2(30) | NC_009231 |
| | | | | phiMMP01(30) | NC_028883 |
| | | | | CDMH1(29) | NC_024144 |
| M68 | 4 | 1686755 | 1741861 | phiMMP01(32) | NC_028883 |
| | | | | phiC2(29) | NC_009231 |
| | | | | phiMMP03(28) | NC_028959 |
| | | | | CDMH1(27) | NC_024144 |

| Strain | Region | Start position | End position | Phage (No. matching genes) | NCBI Accession number |
|---|---|---|---|---|---|
| | 9 | 4244938 | 4303101 | phiCDHM19(21) | NC_028996 |
| | | | | phiCD119(16) | NC_007917 |

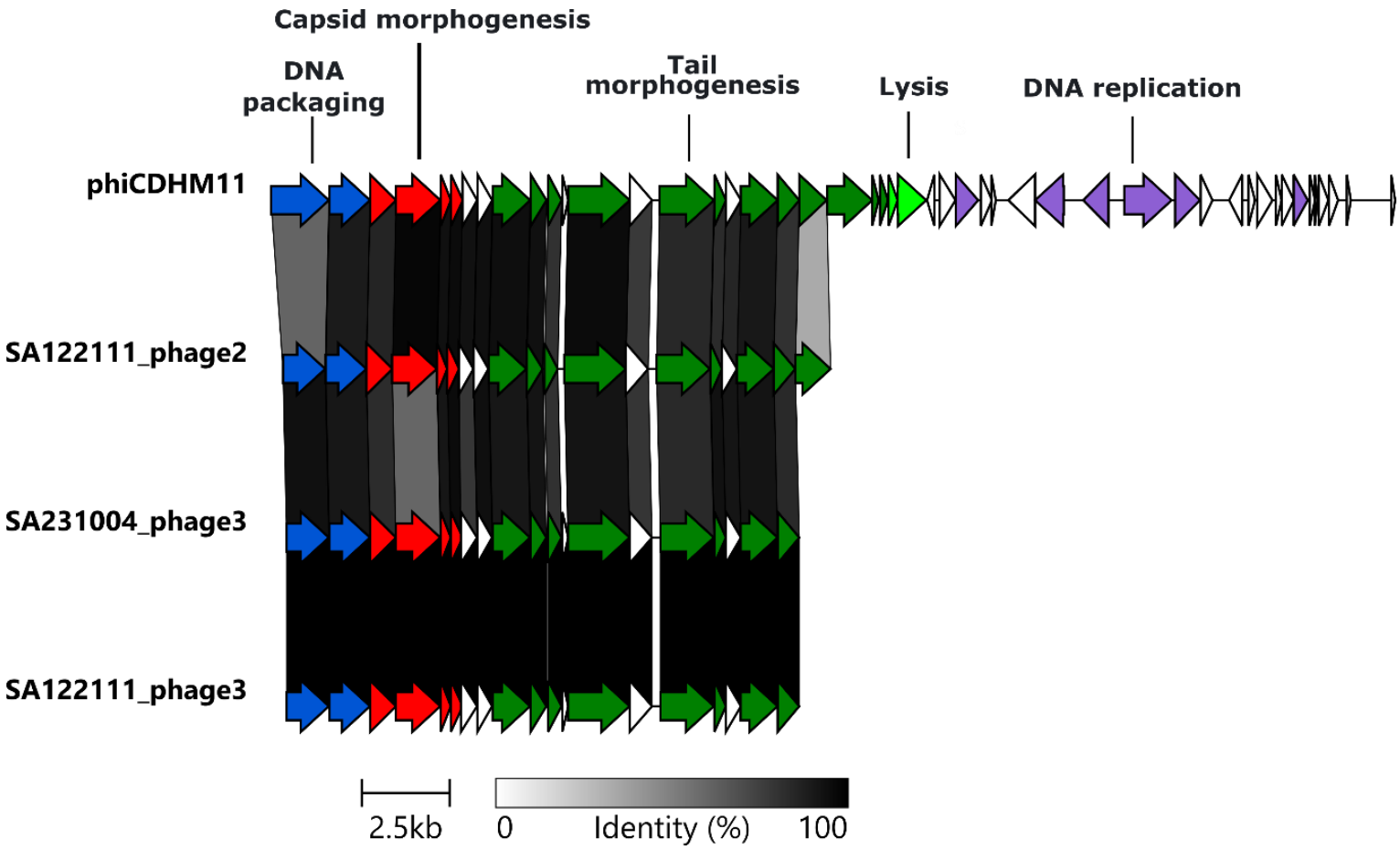

A
Lysogeny control
DNA replication
DNA packaging
Capsid morphogenesis
Tail morphogenesis
Lysogenic conversion
Lysis
SA111045_phage1
phiCDHM19
2.5kb
0
Identity (%)
100


B
Capsid morphogenesis
DNA packaging
Tail morphogenesis
Lysis
Lysogeny control
DNA replication
phiMMP01
SA122025_phage2
SA05842529_phage1
SA424047_phage1
2.5kb
0
Identity (%)
100

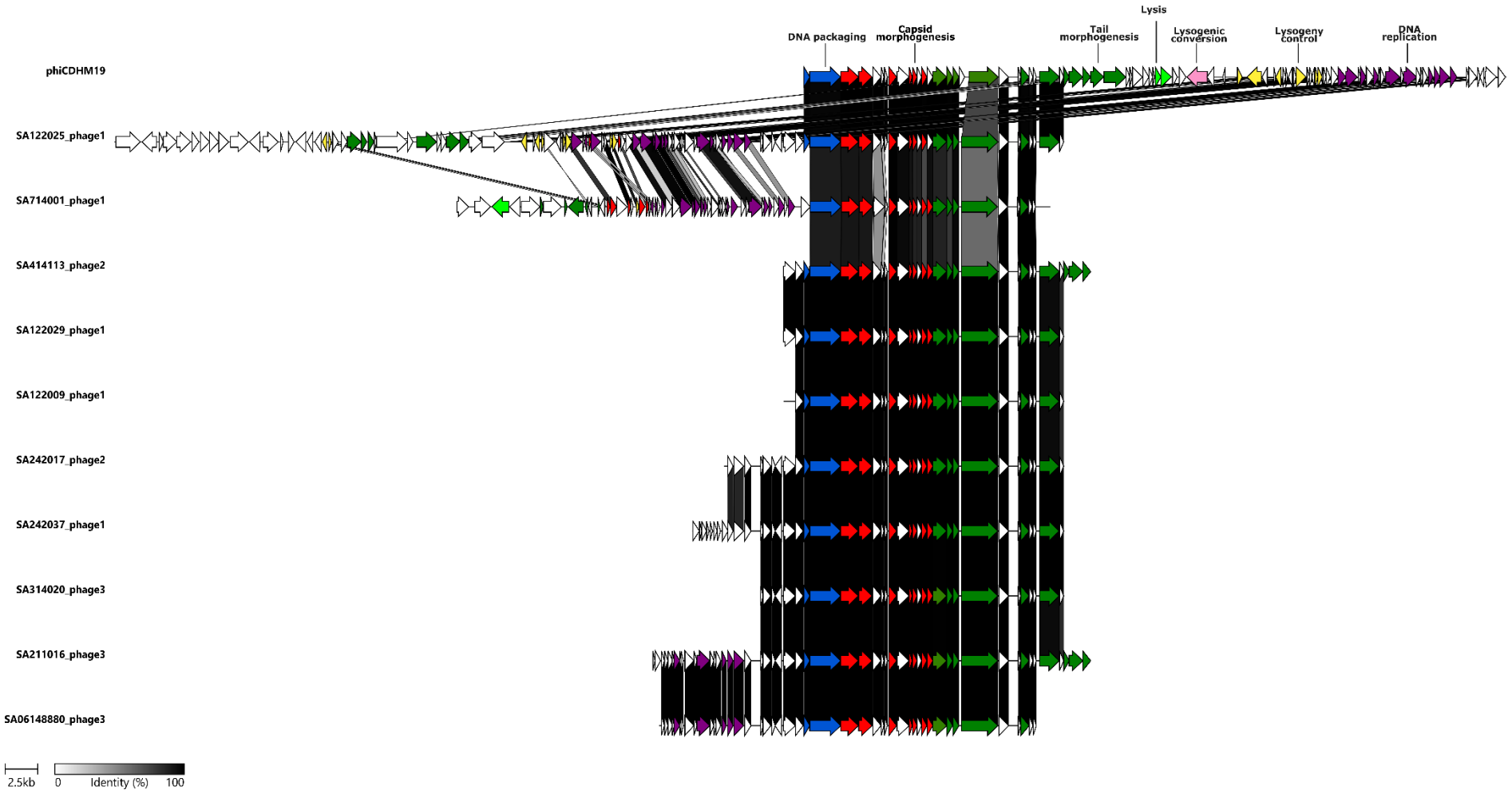
C


Capsid morphogenesis
DNA packaging
Tail morphogenesis
Lysis
Lysogeny control
DNA replication
phiMMP02
SA424047_phage3
SA314020_phage2
SA111031_phage2
2.5kb
0
Identity (%)
100

D


Capsid morphogenesis
DNA packaging
Tail morphogenesis
Lysis
DNA replication
phiCDHM11
SA122111_phage2
SA231004_phage3
SA122111_phage3
2.5kb
0
Identity (%)
100

E


DNA packaging
Capsid morphogenesis
Tail morphogenesis
Lysis
Lysogenic conversion
Lysogeny control
DNA replication
phiCDHM19
SA122025_phage1
SA714001_phage1
SA414113_phage2
SA122029_phage1
SA122009_phage1
SA242017_phage2
SA242037_phage1
SA314020_phage3
SA211016_phage3
SA06148880_phage3
2.5kb
0
Identity (%)
100

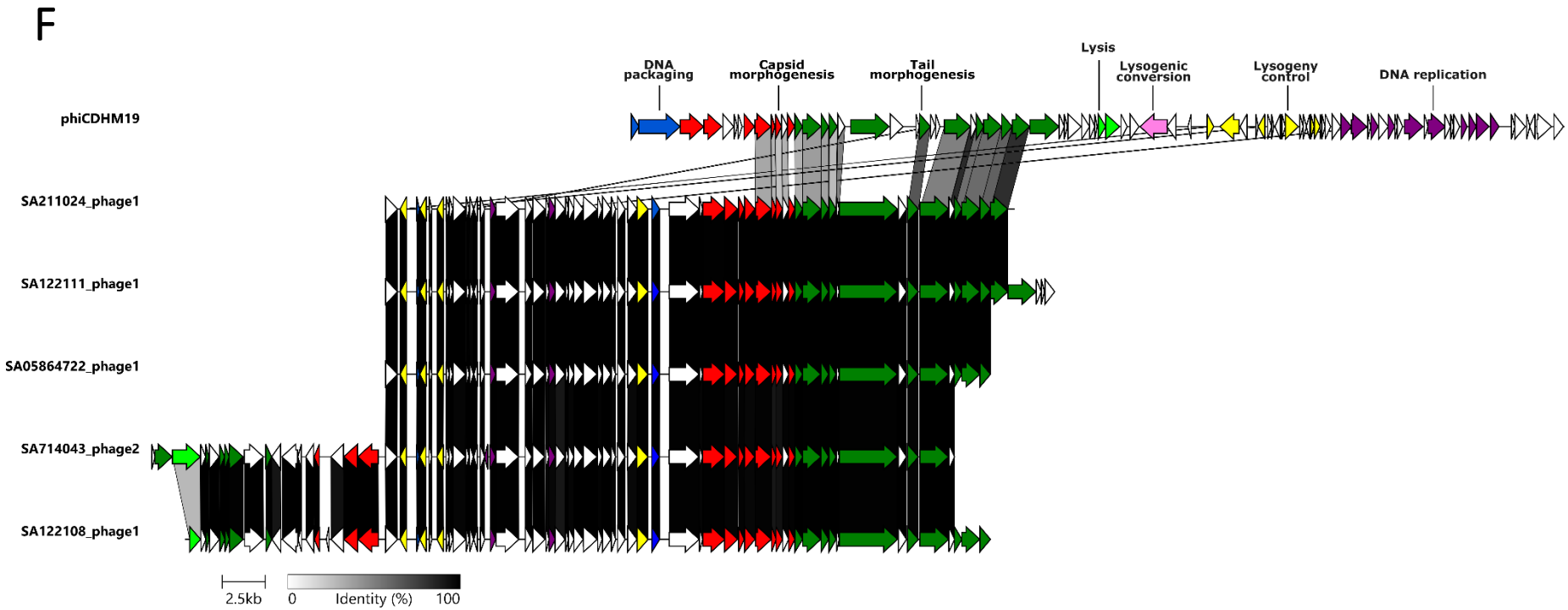



Figure S1. Genomic organisation of phages. Representative phages were selected based on diversity within groups of similar phages. Reference phages were selected based on PHASTER's best hit reference phage. Functional modules colour-coded as DNA packaging (blue), capsid morphogenesis (red), tail morphogenesis (green), lysis (light green), lysogenic conversion (pink), lysogeny control (yellow), DNA replication (purple), hypothetical proteins and protein of unknown function indicated in white, non-homologous genes indicated in grey. Gene direction indicated by arrow orientation. A) Group 1 displayed no similarity to other putative phages; B) Group 2, C) Group 3, D) Group 4, E) Group 5, F) Group 6.


**Figure S2**. Similarity of 98 prophage genomes based of MinHash values. Phage groups colour-coded accordingly: Group 1 (black), group 2 (red), group 3 (orange), group 4 (pink), group 5 (blue), and group 6 (purple). Group 1 consists of only SA111045_phage1.

Table S2. *C. difficile* isolates used in high throughput mitomycin C induction.

| Isolate | MMC induction |
|---|---|
| SA221002* | - |
| SA242006* | - |
| SA514119* | - |
| SA231012* | - |
| M68 | - |
| CD630 | - |
| SA211016 | - |
| SA122118 | - |
| SA05865760 | - |
| SA122105 | - |
| SA122111 | + |
| SA231004 | - |
| SA314020 | - |
| SA122108 | - |
| SA242007 | - |
| SA05842529 | - |
| SA05806999 | + |
| SA05824693 | - |
| SA06098043 | + |
| SA06148880 | + |

*Phage-free isolate (as detected by PHASTER)

Table S3. Genome sizes of *C. difficile* phages and predicted prophages

| Phage name | Family* | Genome size (bp) | GenBank accession No. | Reference |
|---|---|---|---|---|
| phiCD119 | *Myoviridae* | 53,325 | NC_007917.1 | (185) |
| phiC2 | *Myoviridae* | 56,538 | NC_009231 | (181) |
| PhiC630-1 | *Myoviridae* | 55,850 | NS | (181, 325) |
| phiC630-2 | *Myoviridae* | 49,178 | NS | (181, 325) |
| phiCD27 | *Myoviridae* | 50,930 | NC_011398 | (182) |
| **phiCD6356** | *Siphoviridae* | 37,664 | GU949551.1 | (165) |
| **phiCD6365** | *Siphoviridae* | 50,000 | NS | (165) |
| **phiCD38-2** | *Siphoviridae* | 41,090 | NC_015568.1 | (188) |
| phiMMP02 | *Myoviridae* | 48,396 | NC_019421 | (189) |
| phiMMP04 | *Myoviridae* | 31,674 | NC_019422 | (189) |
| phiCDHM1 | *Myoviridae* | 54,279 | HG531805 | (320) |
| **phiCD211†** | *Siphoviridae* | 131,326 | NC_029048.1 | (180) |
| **phiCDIF129T†** | *Siphoviridae* | 131,326 | CP011970.1 | (179) |
| CDKM15 | *Myoviridae* | 50,606 | NS | (190) |
| CDKM9 | *Myoviridae* | 49,822 | KX228399 | (190) |
| **phiCD5763** | *Siphoviridae* | 132,500 | NS | (191) |
| phiCD2955 | *Myoviridae* | 131,600 | NS | (191) |
| JD032 | *Myoviridae* | 35,109 | MK473382 | (192) |
| phiCDKH01 | *Myoviridae* | 45,089 | MN718463 | (193) |
| phiCD1801 | *Myoviridae* | 44,363 | MW512570 | (184) |
| phiCD08011 | *Myoviridae* | 31,394 | MW512572 | (184) |
| phiCD418 | *Myoviridae* | 53,311 | MW512573 | (184) |
| phiCD2301 | *Myoviridae* | 38,695 | MW512571 | (184) |
| phiCDHM13 | *Myoviridae* | 33,596 | HG796225.1 | (266) |
| phiCDHM14 | *Myoviridae* | 33,596 | LK985321.1 | (266) |
| phiCDHM19 | *Myoviridae* | 54,295 | LK985322.1 | (266) |
| **phiCD24-1** | *Siphoviridae* | 44,129 | LN681534.1 | Unpublished |
| **phiCD111** | *Siphoviridae* | 41,560 | LN681535.1 | Unpublished |
| **phiCD146** | *Siphoviridae* | 41,507 | LN681536.1 | Unpublished |
| phiMMP01 | *Myoviridae* | 44,461 | LN681541.1 | Unpublished |
| phiMMP03 | *Myoviridae* | 52,261 | LN681542.1 | Unpublished |
| phiCD481-1 | *Myoviridae* | 32,846 | LN681538 | Unpublished |
| phiCD505 | *Myoviridae* | 49,316 | LN681539.1 | Unpublished |
| phiCD506 | *Myoviridae* | 33,274 | LN681540.1 | Unpublished |
| phiCDHM11 | *Myoviridae* | 32,000 | HG798901.1 | Unpublished |
| phiSemix9P1 | NS | 56,606 | KX905163.1 | Unpublished |

NS; not specified

Boldface indicates siphovirus morphology

†phiCD211 and phiCDIF129T are identical phages

*Morphology-based taxonomic classification, before taxonomic re-classification

# LITERATURE CITED


1. Control CfD. Antibiotic Resistance: The Global Threat. wwwcdcgov/drugresistance/pdf/antibiotic_resistant_fspdf.
2. Organization WH. 2013. Antibiotic resistance - a threat to global health security. https://wwwwhoint/drugresistance/activities/wha66_side_event/en/.
3. Freeman J, Bauer MP, Baines SD, Corver J, Fawley WN, Goorhuis B, Kuijper EJ, Wilcox MH. 2010. The changing epidemiology of Clostridium difficile infections. Clin Microbiol Rev 23:529-49.
4. Na X, Kelly C. 2011. Probiotics in clostridium difficile Infection. J Clin Gastroenterol 45 Suppl:S154-8.
5. Kelly CP. 2012. Can we identify patients at high risk of recurrent Clostridium difficile infection? Clin Microbiol Infect 18 Suppl 6:21-7.
6. Wang S, Xu M, Wang W, Cao X, Piao M, Khan S, Yan F, Cao H, Wang B. 2016. Systematic Review: Adverse Events of Fecal Microbiota Transplantation. PLoS One 11:e0161174.
7. Marcella C, Cui B, Kelly CR, Ianiro G, Cammarota G, Zhang F. 2021. Systematic review: the global incidence of faecal microbiota transplantation-related adverse events from 2000 to 2020. Aliment Pharmacol Ther 53:33-42.
8. Moelling K, Broecker F, Willy C. 2018. A Wake-Up Call: We Need Phage Therapy Now. Viruses 10.
9. Anonymous. 2019. Antibiotic resistance threats in the United States, 2019.
10. Rajabally N, Kullin B, Ebrahim K, Brock T, Weintraub A, Whitelaw A, Bamford C, Watermeyer G, Thomson S, Abratt V, Reid S. 2016. A comparison of Clostridium difficile diagnostic methods for identification of local strains in a South African centre. J Med Microbiol 65:320-327.
11. Legenza L, Barnett S, Rose W, Bianchini M, Safdar N, Coetzee R. 2018. Epidemiology and outcomes of Clostridium difficile infection among hospitalised patients: results of a multicentre retrospective study in South Africa. BMJ Glob Health 3:e000889.
12. Hall IC, O'Toole E. 1935. Intestinal Flora in New-Born Infants: With a Description of a New Pathogenic Anaerobe, Bacillus difficilis. American Journal of Diseases of Children 49:390-402.
13. Smits WK, Lyras D, Lacy DB, Wilcox MH, Kuijper EJ. 2016. Clostridium difficile infection. Nat Rev Dis Primers 2:16020.
14. Larson HE, Price AB, Honour P, Borriello SP. 1978. Clostridium difficile and the aetiology of pseudomembranous colitis. Lancet 1:1063-6.
15. Alyousef AA. 2018. Clostridium difficile: Epidemiology, Pathogenicity, and an Update on the Limitations of and Challenges in Its Diagnosis. J AOAC Int 101:1119-1126.
16. Lawson PA, Citron DM, Tyrrell KL, Finegold SM. 2016. Reclassification of Clostridium difficile as Clostridioides difficile (Hall and O'Toole 1935) Prevot 1938. Anaerobe 40:95-9.
17. Oren A, Rupnik M. 2018. Clostridium difficile and Clostridioides difficile: Two validly published and correct names. Anaerobe 52:125-126.
18. Kachrimanidou M, Malisiovas N. 2011. Clostridium difficile infection: a comprehensive review. Crit Rev Microbiol 37:178-87.
19. Fordtran JS. 2006. Colitis due to Clostridium difficile toxins: underdiagnosed, highly virulent, and nosocomial. Proc (Bayl Univ Med Cent) 19:3-12.
20. Hull MW, Beck PL. 2004. Clostridium difficile-associated colitis. Canadian Family Physician 50:1536.
21. Kelly CP, Pothoulakis C, LaMont JT. 1994. Clostridium difficile colitis. N Engl J Med 330:257-62.
22. Johal SS, Hammond J, Solomon K, James PD, Mahida YR. 2004. Clostridium difficile associated diarrhoea in hospitalised patients: onset in the community and hospital and role of flexible sigmoidoscopy. Gut 53:673-7.

23. Sailhamer EA, Carson K, Chang Y, Zacharias N, Spaniolas K, Tabbara M, Alam HB, DeMoya MA, Velmahos GC. 2009. Fulminant Clostridium difficile colitis: patterns of care and predictors of mortality. Arch Surg 144:433-9; discussion 439-40.
24. Marshak RH, Lester LJ. 1950. Megacolon a complication of ulcerative colitis. Gastroenterology 16:768-72.
25. Earhart MM. 2008. The identification and treatment of toxic megacolon secondary to pseudomembranous colitis. Dimens Crit Care Nurs 27:249-54.
26. Autenrieth DM, Baumgart DC. 2012. Toxic megacolon. Inflamm Bowel Dis 18:584-91.
27. Levine CD. 1999. Toxic megacolon: diagnosis and treatment challenges. AACN Clin Issues 10:492-9.
28. Sayedy L, Kothari D, Richards RJ. 2010. Toxic megacolon associated Clostridium difficile colitis. World J Gastrointest Endosc 2:293-7.
29. Bartlett JG. 2006. Narrative review: the new epidemic of Clostridium difficile-associated enteric disease. Ann Intern Med 145:758-64.
30. Thornton CS, Rubin JE, Greninger AL, Peirano G, Chiu CY, Pillai DR. 2018. Epidemiological and genomic characterization of community-acquired Clostridium difficile infections. BMC Infectious Diseases 18:443.
31. Du T, Choi KB, Silva A, Golding GR, Pelude L, Hizon R, Al-Rawahi GN, Brooks J, Chow B, Collet JC, Comeau JL, Davis I, Evans GA, Frenette C, Han G, Johnstone J, Kibsey P, Katz KC, Langley JM, Lee BE, Longtin Y, Mertz D, Minion J, Science M, Srigley JA, Stagg P, Suh KN, Thampi N, Wong A, Hota SS. 2022. Characterization of Healthcare-Associated and Community-Associated Clostridioides difficile Infections among Adults, Canada, 2015-2019. Emerg Infect Dis 28:1128-1136.
32. Lim SC, Knight DR, Riley TV. 2020. Clostridium difficile and One Health. Clin Microbiol Infect 26:857-863.
33. Zanichelli V, Garenc C, Villeneuve J, Moisan D, Frenette C, Loo V, Longtin Y. 2020. Increased Community-Associated Clostridioides difficile Infections in Quebec, Canada, 2008-2015(1). Emerg Infect Dis 26:1291-1294.
34. Sheth PM, Douchant K, Uyanwune Y, Larocque M, Anantharajah A, Borgundvaag E, Dales L, McCreight L, McNaught L, Moore C. 2019. Evidence of transmission of Clostridium difficile in asymptomatic patients following admission screening in a tertiary care hospital. PLoS One 14:e0207138.
35. Halstead F, Ravi A, Thomson N, Nuur M, Hughes K, Brailey M, Oppenheim B. 2019. Whole genome sequencing of toxigenic Clostridium difficile in asymptomatic carriers: insights into possible role in transmission. Journal of Hospital Infection 102:125-134.
36. Adams DJ, Barone JB, Nylund CM. 2021. Community-Associated Clostridioides difficile Infection in Children: A Review of Recent Literature. Journal of the Pediatric Infectious Diseases Society 10:S22-S26.
37. Hensgens MP, Keessen EC, Squire MM, Riley TV, Koene MG, de Boer E, Lipman LJ, Kuijper EJ. 2012. Clostridium difficile infection in the community: a zoonotic disease? Clin Microbiol Infect 18:635-45.
38. Tschudin-Sutter S, Tamma PD, Naegeli AN, Speck KA, Milstone AM, Perl TM. 2013. Distinguishing Community-Associated From Hospital-Associated Clostridium difficile Infections in Children: Implications for Public Health Surveillance. Clinical Infectious Diseases 57:1665-1672.
39. Dubberke ER, Reske KA, Noble-Wang J, Thompson A, Killgore G, Mayfield J, Camins B, Woeltje K, McDonald JR, McDonald LC, Fraser VJ. 2007. Prevalence of Clostridium difficile environmental contamination and strain variability in multiple health care facilities. Am J Infect Control 35:315-8.
40. Kelly CP, LaMont JT. 2008. Clostridium difficile--more difficult than ever. N Engl J Med 359:1932-40.

41. Baines SD, O'Connor R, Saxton K, Freeman J, Wilcox MH. 2009. Activity of vancomycin against epidemic Clostridium difficile strains in a human gut model. J Antimicrob Chemother 63:520-5.
42. Ali S, Moore G, Wilson AP. 2011. Spread and persistence of Clostridium difficile spores during and after cleaning with sporicidal disinfectants. J Hosp Infect 79:97-8.
43. Paredes-Sabja D, Shen A, Sorg JA. 2014. Clostridium difficile spore biology: sporulation, germination, and spore structural proteins. Trends Microbiol 22:406-16.
44. Claro T, Daniels S, Humphreys H. 2014. Detecting Clostridium difficile spores from inanimate surfaces of the hospital environment: which method is best? J Clin Microbiol 52:3426-8.
45. Buffie CG, Jarchum I, Equinda M, Lipuma L, Gobourne A, Viale A, Ubeda C, Xavier J, Pamer EG. 2012. Profound alterations of intestinal microbiota following a single dose of clindamycin results in sustained susceptibility to Clostridium difficile-induced colitis. Infect Immun 80:62-73.
46. Bassis CM, Theriot CM, Young VB. 2014. Alteration of the murine gastrointestinal microbiota by tigecycline leads to increased susceptibility to Clostridium difficile infection. Antimicrob Agents Chemother 58:2767-74.
47. Sorg JA, Sonenshein AL. 2008. Bile salts and glycine as cogerminants for Clostridium difficile spores. J Bacteriol 190:2505-12.
48. Francis MB, Allen CA, Shrestha R, Sorg JA. 2013. Bile acid recognition by the Clostridium difficile germinant receptor, CspC, is important for establishing infection. PLoS Pathog 9:e1003356.
49. Chandrasekaran R, Lacy DB. 2017. The role of toxins in Clostridium difficile infection. FEMS Microbiol Rev 41:723-750.
50. Leffler DA, Lamont JT. 2015. Clostridium difficile infection. N Engl J Med 372:1539-48.
51. Mehner-Breitfeld D, Rathmann C, Riedel T, Just I, Gerhard R, Overmann J, Brüser T. 2018. Evidence for an Adaptation of a Phage-Derived Holin/Endolysin System to Toxin Transport in Clostridioides difficile. Front Microbiol 9:2446.
52. Zhang X, Li J, Chen C, Liu YJ, Cui Q, Hong W, Chen Z, Feng Y, Cui G. 2023. Molecular Basis of TcdR-Dependent Promoter Activity for Toxin Production by Clostridioides difficile Studied by a Heterologous Reporter System. Toxins (Basel) 15.
53. Hundsberger T, Braun V, Weidmann M, Leukel P, Sauerborn M, von Eichel-Streiber C. 1997. Transcription analysis of the genes tcdA-E of the pathogenicity locus of Clostridium difficile. Eur J Biochem 244:735-42.
54. He R, Peng J, Yuan P, Yang J, Wu X, Wang Y, Wei W. 2017. Glucosyltransferase Activity of Clostridium difficile Toxin B Triggers Autophagy-mediated Cell Growth Arrest. Scientific Reports 7:10532.
55. Landenberger M, Nieland J, Roeder M, Nørgaard K, Papatheodorou P, Ernst K, Barth H. 2021. The cytotoxic effect of Clostridioides difficile pore-forming toxin CDTb. Biochimica et Biophysica Acta (BBA) - Biomembranes 1863:183603.
56. Martínez-Meléndez A, Cruz-López F, Morfin-Otero R, Maldonado-Garza HJ, Garza-González E. 2022. An Update on Clostridioides difficile Binary Toxin. Toxins 14:305.
57. Papatheodorou P, Minton NP, Aktories K, Barth H. 2024. An Updated View on the Cellular Uptake and Mode-of-Action of Clostridioides difficile Toxins. Adv Exp Med Biol 1435:219-247.
58. Wang IN, Deaton J, Young R. 2003. Sizing the holin lesion with an endolysin-beta-galactosidase fusion. J Bacteriol 185:779-87.
59. Carter GP, Lyras D, Allen DL, Mackin KE, Howarth PM, O'Connor JR, Rood JI. 2007. Binary toxin production in Clostridium difficile is regulated by CdtR, a LytTR family response regulator. J Bacteriol 189:7290-301.
60. Gonçalves C, Decré D, Barbut F, Burghoffer B, Petit JC. 2004. Prevalence and characterization of a binary toxin (actin-specific ADP-ribosyltransferase) from Clostridium difficile. J Clin Microbiol 42:1933-9.

61. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A 95:3140-5.
62. Knetsch CW, Lawley TD, Hensgens MP, Corver J, Wilcox MW, Kuijper EJ. 2013. Current application and future perspectives of molecular typing methods to study Clostridium difficile infections. Euro Surveill 18:20381.
63. Killgore G, Thompson A, Johnson S, Brazier J, Kuijper E, Pepin J, Frost EH, Savelkoul P, Nicholson B, van den Berg RJ, Kato H, Sambol SP, Zukowski W, Woods C, Limbago B, Gerding DN, McDonald LC. 2008. Comparison of seven techniques for typing international epidemic strains of Clostridium difficile: restriction endonuclease analysis, pulsed-field gel electrophoresis, PCR-ribotyping, multilocus sequence typing, multilocus variable-number tandem-repeat analysis, amplified fragment length polymorphism, and surface layer protein A gene sequence typing. J Clin Microbiol 46:431-7.
64. Clabots CR, Johnson S, Bettin KM, Mathie PA, Mulligan ME, Schaberg DR, Peterson LR, Gerding DN. 1993. Development of a rapid and efficient restriction endonuclease analysis typing system for Clostridium difficile and correlation with other typing systems. J Clin Microbiol 31:1870-5.
65. Rupnik M. 2008. Heterogeneity of large clostridial toxins: importance of Clostridium difficile toxinotypes. FEMS Microbiol Rev 32:541-55.
66. Kuijper EJ, van den Berg RJ, Brazier JS. 2009. Comparison of molecular typing methods applied to Clostridium difficile. Methods Mol Biol 551:159-71.
67. Huber CA, Foster NF, Riley TV, Paterson DL. 2013. Challenges for standardization of Clostridium difficile typing methods. J Clin Microbiol 51:2810-4.
68. MacCannell DR, Louie TJ, Gregson DB, Laverdiere M, Labbe AC, Laing F, Henwick S. 2006. Molecular analysis of Clostridium difficile PCR ribotype 027 isolates from Eastern and Western Canada. J Clin Microbiol 44:2147-52.
69. O'Connor JR, Johnson S, Gerding DN. 2009. Clostridium difficile Infection Caused by the Epidemic BI/NAP1/027 Strain. Gastroenterology 136:1913-1924.
70. Kuijper EJ, van den Berg RJ, Debast S, Visser CE, Veenendaal D, Troelstra A, van der Kooi T, van den Hof S, Notermans DW. 2006. Clostridium difficile ribotype 027, toxinotype III, the Netherlands. Emerg Infect Dis 12:827-30.
71. Oleastro M, Coelho M, Gião M, Coutinho S, Mota S, Santos A, Rodrigues J, Faria D. 2014. Outbreak of Clostridium difficile PCR ribotype 027--the recent experience of a regional hospital. BMC Infect Dis 14:209.
72. Jin H, Ni K, Wei L, Shen L, Xu H, Kong Q, Ni X. 2016. Identification of Clostridium difficile RT078 From Patients and Environmental Surfaces in Zhejiang Province, China. Infect Control Hosp Epidemiol 37:745-6.
73. van Dorp SM, de Greeff SC, Harmanus C, Sanders I, Dekkers OM, Knetsch CW, Kampinga GA, Notermans DW, Kuijper EJ. 2017. Ribotype 078 Clostridium difficile infection incidence in Dutch hospitals is not associated with provincial pig farming: Results from a national sentinel surveillance, 2009-2015. PLoS One 12:e0189183.
74. Goorhuis A, Bakker D, Corver J, Debast SB, Harmanus C, Notermans DW, Bergwerff AA, Dekker FW, Kuijper EJ. 2008. Emergence of Clostridium difficile infection due to a new hypervirulent strain, polymerase chain reaction ribotype 078. Clin Infect Dis 47:1162-70.
75. Keel K, Brazier JS, Post KW, Weese S, Songer JG. 2007. Prevalence of PCR ribotypes among Clostridium difficile isolates from pigs, calves, and other species. J Clin Microbiol 45:1963-4.
76. Rupnik M, Kato N, Grabnar M, Kato H. 2003. New types of toxin A-negative, toxin B-positive strains among Clostridium difficile isolates from Asia. J Clin Microbiol 41:1118-25.
77. Pituch H, van den Braak N, van Leeuwen W, van Belkum A, Martirosian G, Obuch-Woszczatyński P, Łuczak M, Meisel-Mikołajczyk F. 2001. Clonal dissemination of a toxin-A-

negative/toxin-B-positive Clostridium difficile strain from patients with antibiotic-associated diarrhea in Poland. Clin Microbiol Infect 7:442-6.
78. Kullin B, Meggersee R, D'Alton J, Galvao B, Rajabally N, Whitelaw A, Bamford C, Reid SJ, Abratt VR. 2015. Prevalence of gastrointestinal pathogenic bacteria in patients with diarrhoea attending Groote Schuur Hospital, Cape Town, South Africa. S Afr Med J 105:121-5.
79. Warny M, Pepin J, Fang A, Killgore G, Thompson A, Brazier J, Frost E, McDonald LC. 2005. Toxin production by an emerging strain of Clostridium difficile associated with outbreaks of severe disease in North America and Europe. Lancet 366:1079-84.
80. Anwar F, Roxas BAP, Shehab KW, Ampel NM, Viswanathan VK, Vedantam G. 2022. Low-toxin Clostridioides difficile RT027 strains exhibit robust virulence. Emerg Microbes Infect 11:1982-1993.
81. Vitucci JC, Pulse M, Tabor-Simecka L, Simecka J. 2020. Epidemic ribotypes of Clostridium (now Clostridioides) difficile are likely to be more virulent than non-epidemic ribotypes in animal models. BMC Microbiol 20:27.
82. Kim DY, Cheknis AK, Serna-Perez F, Lin MY, Hayden MK, Moore NM, Harrington A, Tesic V, Beavis KG, Gerding DN, Johnson S, Skinner AM. 2022. 403. Strain Epidemiology of Clostridioides difficile across Three Geographically Distinct Medical Centers in Chicago. Open Forum Infect Dis 9:ofac492.481. doi: 10.1093/ofid/ofac492.481. eCollection 2022 Dec.
83. McDermott LA, Thorpe CM, Goldstein E, Schuetz A, Johnson S, Gerding DN, Carroll KC, Garey KW, Lancaster C, Walk S, Duperchy E, Snydman DR, Gluck L, Bourdas D. 2022. 1669. A US Based National Surveillance Study for the Susceptibility and Epidemiology of Clostridioides difficile Associated Diarrheal Isolates with Special Reference to Ridinilazole: 2020-2021. Open Forum Infectious Diseases 9.
84. Guh AY, Mu Y, Winston LG, Johnston H, Olson D, Farley MM, Wilson LE, Holzbauer SM, Phipps EC, Dumyati GK, Beldavs ZG, Kainer MA, Karlsson M, Gerding DN, McDonald LC. 2020. Trends in U.S. Burden of Clostridioides difficile Infection and Outcomes. New England Journal of Medicine 382:1320-1330.
85. Liu C, Monaghan T, Yadegar A, Louie T, Kao D. 2023. Insights into the Evolving Epidemiology of Clostridioides difficile Infection and Treatment: A Global Perspective. Antibiotics (Basel) 12.
86. Collins DA, Marcella S, Campbell M, Riley TV. 2022. Linkage study of surveillance and hospital admission data to investigate Clostridium difficile infection in hospital patients in Perth, Western Australia. Anaerobe 74:102528.
87. (ACSQHC). Australian Commission on Safety and Quality in Health Care. 2018. *Clostridium difficile*. Infection 2018 Data Snapshot. Available online: https://www.safetyandquality.gov.au/publications-and-resources/resource-library/clostridium-difficile-infection-2018-data-snapshot#:~:text=to-person%20contact.-,C.,around%206%2C000%20cases%20of%20CDI (accessed 10 December 2023).
88. Collins DA, Riley TV. 2016. Routine detection of Clostridium difficile in Western Australia. Anaerobe 37:34-37.
89. Hong S, Putsathit P, George N, Hemphill C, Huntington PG, Korman TM, Kotsanas D, Lahra M, McDougall R, Moore CV, Nimmo GR, Prendergast L, Robson J, Waring L, Wehrhahn MC, Weldhagen GF, Wilson RM, Riley TV, Knight DR. 2020. Laboratory-Based Surveillance of Clostridium difficile Infection in Australian Health Care and Community Settings, 2013 to 2018. J Clin Microbiol 58.
90. Knight DR, Imwattana K, Collins DA, Lim S-C, Hong S, Putsathit P, Riley TV. 2023. Genomic epidemiology and transmission dynamics of recurrent Clostridioides difficile infection in Western Australia. European Journal of Clinical Microbiology & Infectious Diseases 42:607-619.
91. Songer JG, Hien TT, George EK, Angela DT, McDonald LC, Brandi ML. 2009. *Clostridium difficile* in Retail Meat Products, USA, 2007. Emerging Infectious Disease journal 15:819.

92. Knight DR, Squire MM, Collins DA, Riley TV. 2016. Genome Analysis of *Clostridium difficile* PCR Ribotype 014 Lineage in Australian Pigs and Humans Reveals a Diverse Genetic Repertoire and Signatures of Long-Range Interspecies Transmission. Front Microbiol 7:2138.
93. Brajerova M, Zikova J, Krutova M. 2022. Clostridioides difficile epidemiology in the Middle and the Far East. Anaerobe 74:102542.
94. Imwattana K, Knight DR, Kullin B, Collins DA, Putsathit P, Kiratisin P, Riley TV. 2019. Clostridium difficile ribotype 017 - characterization, evolution and epidemiology of the dominant strain in Asia. Emerg Microbes Infect 8:796-807.
95. Collins DA, Sohn KM, Wu Y, Ouchi K, Ishii Y, Elliott B, Riley TV, Tateda K. 2020. Clostridioides difficile infection in the Asia-Pacific region. Emerg Microbes Infect 9:42-52.
96. King AM, Mackin KE, Lyras D. 2015. Emergence of toxin A-negative, toxin B-positive Clostridium difficile strains: epidemiological and clinical considerations. Future Microbiol 10:1-4.
97. Kullin B, Brock T, Rajabally N, Anwar F, Vedantam G, Reid S, Abratt V. 2016. Characterisation of Clostridium difficile strains isolated from Groote Schuur Hospital, Cape Town, South Africa. Eur J Clin Microbiol Infect Dis 35:1709-18.
98. Kullin B, Wojno J, Abratt V, Reid SJ. 2017. Toxin A-negative toxin B-positive ribotype 017 Clostridium difficile is the dominant strain type in patients with diarrhoea attending tuberculosis hospitals in Cape Town, South Africa. Eur J Clin Microbiol Infect Dis 36:163-175.
99. Hoda H, Gehan H. 2006. Prevalent PCR ribotypes and antibiotic sensitivity of clinical isolates of clostridium difficile.
100. Navaneethan U, Giannella RA. 2009. Thinking beyond the colon-small bowel involvement in clostridium difficile infection. Gut Pathog 1:7.
101. Dabard J, Dubos F, Martinet L, Ducluzeau R. 1979. Experimental reproduction of neonatal diarrhea in young gnotobiotic hares simultaneously associated with Clostridium difficile and other Clostridium strains. Infect Immun 24:7-11.
102. al Saif N, Brazier JS. 1996. The distribution of Clostridium difficile in the environment of South Wales. J Med Microbiol 45:133-7.
103. Borriello SP, Honour P, Turner T, Barclay F. 1983. Household pets as a potential reservoir for Clostridium difficile infection. J Clin Pathol 36:84-7.
104. Bojesen AM, Olsen KE, Bertelsen MF. 2006. Fatal enterocolitis in Asian elephants (Elephas maximus) caused by Clostridium difficile. Vet Microbiol 116:329-35.
105. Andrés-Lasheras S, Martín-Burriel I, Mainar-Jaime RC, Morales M, Kuijper E, Blanco JL, Chirino-Trejo M, Bolea R. 2018. Preliminary studies on isolates of Clostridium difficile from dogs and exotic pets. BMC Vet Res 14:77.
106. Weese JS. 2020. Clostridium (Clostridioides) difficile in animals. J Vet Diagn Invest 32:213-221.
107. Perez J, Springthorpe VS, Sattar SA. 2005. Activity of selected oxidizing microbicides against the spores of Clostridium difficile: relevance to environmental control. Am J Infect Control 33:320-5.
108. Buggy BP, Wilson KH, Fekety R. 1983. Comparison of methods for recovery of Clostridium difficile from an environmental surface. J Clin Microbiol 18:348-52.
109. Hota B. 2004. Contamination, disinfection, and cross-colonization: are hospital surfaces reservoirs for nosocomial infection? Clin Infect Dis 39:1182-9.
110. Verity P, Wilcox MH, Fawley W, Parnell P. 2001. Prospective evaluation of environmental contamination by Clostridium difficile in isolation side rooms. J Hosp Infect 49:204-9.
111. Poutanen SM, Simor AE. 2004. Clostridium difficile-associated diarrhea in adults. Cmaj 171:51-8.
112. Health Do. 2009. *Clostridioides difficile* infection: How to Deal with the Problem. Department of Health.

113. Boyce JM, Ligi C, Kohan C, Dumigan D, Havill NL. 2006. Lack of association between the increased incidence of Clostridium difficile-associated disease and the increasing use of alcohol-based hand rubs. Infect Control Hosp Epidemiol 27:479-83.
114. Song JH, Kim YS. 2019. Recurrent Clostridium difficile Infection: Risk Factors, Treatment, and Prevention. Gut and liver 13:16.
115. Stevens VW, Nelson RE, Schwab-Daugherty EM, Khader K, Jones MM, Brown KA, Greene T, Croft LD, Neuhauser M, Glassman P, Goetz MB, Samore MH, Rubin MA. 2017. Comparative Effectiveness of Vancomycin and Metronidazole for the Prevention of Recurrence and Death in Patients With Clostridium difficile Infection. JAMA Intern Med 177:546-553.
116. Zar FA, Bakkanagari SR, Moorthi KM, Davis MB. 2007. A comparison of vancomycin and metronidazole for the treatment of Clostridium difficile-associated diarrhea, stratified by disease severity. Clin Infect Dis 45:302-7.
117. Johnson S, Louie TJ, Gerding DN, Cornely OA, Chasan-Taber S, Fitts D, Gelone SP, Broom C, Davidson DM. 2014. Vancomycin, metronidazole, or tolevamer for Clostridium difficile infection: results from two multinational, randomized, controlled trials. Clin Infect Dis 59:345-54.
118. Venugopal AA, Johnson S. 2012. Fidaxomicin: a novel macrocyclic antibiotic approved for treatment of Clostridium difficile infection. Clin Infect Dis 54:568-74.
119. Al-Jashaami LS, DuPont HL. 2016. Management of Clostridium difficile Infection. Gastroenterol Hepatol (N Y) 12:609-616.
120. Lawley TD, Walker AW. 2013. Intestinal colonization resistance. Immunology 138:1-11.
121. FAO/WHO. 2002. Guidelines for the Evaluation of Probiotics in Food.
122. Lawrence SJ, Korzenik JR, Mundy LM. 2005. Probiotics for recurrent Clostridium difficile disease. J Med Microbiol 54:905-6.
123. Gorbach SL, Chang TW, Goldin B. 1987. Successful treatment of relapsing Clostridium difficile colitis with Lactobacillus GG. Lancet 2:1519.
124. Biller JA, Katz AJ, Flores AF, Buie TM, Gorbach SL. 1995. Treatment of recurrent Clostridium difficile colitis with Lactobacillus GG. J Pediatr Gastroenterol Nutr 21:224-6.
125. Hickson M, D'Souza AL, Muthu N, Rogers TR, Want S, Rajkumar C, Bulpitt CJ. 2007. Use of probiotic Lactobacillus preparation to prevent diarrhoea associated with antibiotics: randomised double blind placebo controlled trial. BMJ 335:80.
126. Zuo T, Wong SH, Lam K, Lui R, Cheung K, Tang W, Ching JYL, Chan PKS, Chan MCW, Wu JCY, Chan FKL, Yu J, Sung JJY, Ng SC. 2018. Bacteriophage transfer during faecal microbiota transplantation in Clostridium difficile infection is associated with treatment outcome. Gut 67:634-643.
127. Lee CH, Belanger JE, Kassam Z, Smieja M, Higgins D, Broukhanski G, Kim PT. 2014. The outcome and long-term follow-up of 94 patients with recurrent and refractory Clostridium difficile infection using single to multiple fecal microbiota transplantation via retention enema. Eur J Clin Microbiol Infect Dis 33:1425-8.
128. Lee CH, Steiner T, Petrof EO, Smieja M, Roscoe D, Nematallah A, Weese JS, Collins S, Moayyedi P, Crowther M, Ropeleski MJ, Jayaratne P, Higgins D, Li Y, Rau NV, Kim PT. 2016. Frozen vs Fresh Fecal Microbiota Transplantation and Clinical Resolution of Diarrhea in Patients With Recurrent Clostridium difficile Infection: A Randomized Clinical Trial. Jama 315:142-9.
129. Drekonja D, Reich J, Gezahegn S, Greer N, Shaukat A, MacDonald R, Rutks I, Wilt TJ. 2015. Fecal Microbiota Transplantation for Clostridium difficile Infection: A Systematic Review. Ann Intern Med 162:630-8.
130. van Nood E, Vrieze A, Nieuwdorp M, Fuentes S, Zoetendal EG, de Vos WM, Visser CE, Kuijper EJ, Bartelsman JF, Tijssen JG, Speelman P, Dijkgraaf MG, Keller JJ. 2013. Duodenal infusion of donor feces for recurrent Clostridium difficile. N Engl J Med 368:407-15.
131. Shogbesan O, Poudel DR, Victor S, Jehangir A, Fadahunsi O, Shogbesan G, Donato A. 2018. A Systematic Review of the Efficacy and Safety of Fecal Microbiota Transplant for Clostridium

difficile Infection in Immunocompromised Patients. Can J Gastroenterol Hepatol 2018:1394379.
132. DeFilipp Z, Bloom PP, Torres Soto M, Mansour MK, Sater MRA, Huntley MH, Turbett S, Chung RT, Chen YB, Hohmann EL. 2019. Drug-Resistant E. coli Bacteremia Transmitted by Fecal Microbiota Transplant. N Engl J Med 381:2043-2050.
133. Shin JH, Hays RA, Warren CA. 2021. Hospitalized Older Patients with Clostridioides difficile Infection Refractory to Conventional Antibiotic Therapy Benefit from Fecal Microbiota Transplant. Adv Geriatr Med Res 3.
134. Gambino M, Nørgaard Sørensen A, Ahern S, Smyrlis G, Gencay YE, Hendrix H, Neve H, Noben JP, Lavigne R, Brøndsted L. 2020. Phage S144, A New Polyvalent Phage Infecting Salmonella spp. and Cronobacter sakazakii. Int J Mol Sci 21.
135. Duc HM, Son HM, Yi HPS, Sato J, Ngan PH, Masuda Y, Honjoh K-i, Miyamoto T. 2020. Isolation, characterization and application of a polyvalent phage capable of controlling Salmonella and Escherichia coli O157:H7 in different food matrices. Food Research International 131:108977.
136. Chanishvili N. 2012. Phage therapy--history from Twort and d'Herelle through Soviet experience to current approaches. Adv Virus Res 83:3-40.
137. Hodyra-Stefaniak K, Miernikiewicz P, Drapała J, Drab M, Jończyk-Matysiak E, Lecion D, Kaźmierczak Z, Beta W, Majewska J, Harhala M, Bubak B, Klopot A, Górski A, Dabrowska K. 2015. Mammalian Host-Versus-Phage immune response determines phage fate in vivo. Scientific reports 5:14802.
138. Summers WC. 2012. The strange history of phage therapy. Bacteriophage 2:130-133.
139. Lin DM, Koskella B, Lin HC. 2017. Phage therapy: An alternative to antibiotics in the age of multi-drug resistance. World J Gastrointest Pharmacol Ther 8:162-173.
140. Law N, Logan C, Yung G, Furr CL, Lehman SM, Morales S, Rosas F, Gaidamaka A, Bilinsky I, Grint P, Schooley RT, Aslam S. 2019. Successful adjunctive use of bacteriophage therapy for treatment of multidrug-resistant Pseudomonas aeruginosa infection in a cystic fibrosis patient. Infection 47:665-668.
141. Fish R, Kutter E, Bryan D, Wheat G, Kuhl S. 2018. Resolving Digital Staphylococcal Osteomyelitis Using Bacteriophage-A Case Report. Antibiotics (Basel) 7.
142. Khawaldeh A, Morales S, Dillon B, Alavidze Z, Ginn AN, Thomas L, Chapman SJ, Dublanchet A, Smithyman A, Iredell JR. 2011. Bacteriophage therapy for refractory Pseudomonas aeruginosa urinary tract infection. J Med Microbiol 60:1697-1700.
143. Doub JB, Ng VY, Johnson AJ, Slomka M, Fackler J, Horne B, Brownstein MJ, Henry M, Malagon F, Biswas B. 2020. Salvage Bacteriophage Therapy for a Chronic MRSA Prosthetic Joint Infection. Antibiotics (Basel) 9.
144. Gilbey T, Ho J, Cooley LA, Petrovic Fabijan A, Iredell JR. 2019. Adjunctive bacteriophage therapy for prosthetic valve endocarditis due to Staphylococcus aureus. Med J Aust 211:142-143.e1.
145. Hoyle N, Zhvaniya P, Balarjishvili N, Bolkvadze D, Nadareishvili L, Nizharadze D, Wittmann J, Rohde C, Kutateladze M. 2018. Phage therapy against Achromobacter xylosoxidans lung infection in a patient with cystic fibrosis: a case report. Res Microbiol 169:540-542.
146. Lwoff A. 1953. Lysogeny. Bacteriol Rev 17:269-337.
147. Hobbs Z, Abedon ST. 2016. Diversity of phage infection types and associated terminology: the problem with 'Lytic or lysogenic'. FEMS Microbiol Lett 363.
148. Ptashne M. 2004. A genetic switch: phage lambda revisited, vol 3. Cold Spring Harbor Laboratory Press Cold Spring Harbor, NY.
149. Clokie MR, Millard AD, Letarov AV, Heaphy S. 2011. Phages in nature. Bacteriophage 1:31-45.
150. Little JW. 2005. Lysogeny, Prophage Induction, and Lysogenic Conversion, p 37-54, Phages doi:https://doi.org/10.1128/9781555816506.ch3.
151. Little JW, Shepley DP, Wert DW. 1999. Robustness of a gene regulatory circuit. Embo j 18:4299-307.

152. Shearwin KE, Truong JQ. 2021. Lysogeny, p 77-87. *In* Bamford DH, Zuckerman M (ed), Encyclopedia of Virology (Fourth Edition) doi:https://doi.org/10.1016/B978-0-12-809633-8.20963-1. Academic Press, Oxford.
153. Oppenheim AB, Kobiler O, Stavans J, Court DL, Adhya S. 2005. Switches in Bacteriophage Lambda Development. Annual Review of Genetics 39:409-429.
154. Little JW, Mount DW. 1982. The SOS regulatory system of Escherichia coli. Cell 29:11-22.
155. Babić AC, Little JW. 2007. Cooperative DNA binding by CI repressor is dispensable in a phage λ variant. Proceedings of the National Academy of Sciences 104:17741-17746.
156. Schubert RA, Dodd IB, Egan JB, Shearwin KE. 2007. Cro's role in the CI Cro bistable switch is critical for {lambda}'s transition from lysogeny to lytic development. Genes Dev 21:2461-72.
157. Dobbins JJ. 2010. Prescott's Microbiology, Eighth Edition. J Microbiol Biol Educ 11:64-5. doi: 10.1128/jmbe.v11.i1.154. eCollection 2010.
158. Wang IN, Smith DL, Young R. 2000. Holins: the protein clocks of bacteriophage infections. Annu Rev Microbiol 54:799-825.
159. Echols H, Green L. 1971. Establishment and maintenance of repression by bacteriophage lambda: the role of the cI, cII, and c3 proteins. Proc Natl Acad Sci U S A 68:2190-4.
160. Barylski J, Enault F, Dutilh BE, Schuller MB, Edwards RA, Gillis A, Klumpp J, Knezevic P, Krupovic M, Kuhn JH, Lavigne R, Oksanen HM, Sullivan MB, Jang HB, Simmonds P, Aiewsakun P, Wittmann J, Tolstoy I, Brister JR, Kropinski AM, Adriaenssens EM. 2019. Analysis of Spounaviruses as a Case Study for the Overdue Reclassification of Tailed Phages. Systematic Biology 69:110-123.
161. Low SJ, Džunková M, Chaumeil P-A, Parks DH, Hugenholtz P. 2019. Evaluation of a concatenated protein phylogeny for classification of tailed double-stranded DNA viruses belonging to the order Caudovirales. Nature Microbiology 4:1306-1315.
162. Lawrence JG, Hatfull GF, Hendrix RW. 2002. Imbroglios of Viral Taxonomy: Genetic Exchange and Failings of Phenetic Approaches. Journal of Bacteriology 184:4891-4905.
163. Turner D, Shkoporov AN, Lood C, Millard AD, Dutilh BE, Alfenas-Zerbini P, van Zyl LJ, Aziz RK, Oksanen HM, Poranen MM, Kropinski AM, Barylski J, Brister JR, Chanisvili N, Edwards RA, Enault F, Gillis A, Knezevic P, Krupovic M, Kurtböke I, Kushkina A, Lavigne R, Lehman S, Lobocka M, Moraru C, Moreno Switt A, Morozova V, Nakavuma J, Reyes Muñoz A, Rūmnieks J, Sarkar BL, Sullivan MB, Uchiyama J, Wittmann J, Yigang T, Adriaenssens EM. 2023. Abolishment of morphology-based taxa and change to binomial species names: 2022 taxonomy update of the ICTV bacterial viruses subcommittee. Archives of Virology 168:74.
164. Sell TL, Schaberg DR, Fekety FR. 1983. Bacteriophage and bacteriocin typing scheme for Clostridium difficile. J Clin Microbiol 17:1148-52.
165. Horgan M, O'Sullivan O, Coffey A, Fitzgerald GF, van Sinderen D, McAuliffe O, Ross RP. 2010. Genome analysis of the Clostridium difficile phage PhiCD6356, a temperate phage of the Siphoviridae family. Gene 462:34-43.
166. Fortier LC, Moineau S. 2007. Morphological and genetic diversity of temperate phages in Clostridium difficile. Appl Environ Microbiol 73:7358-66.
167. Nale JY, Shan J, Hickenbotham PT, Fawley WN, Wilcox MH, Clokie MR. 2012. Diverse temperate bacteriophage carriage in Clostridium difficile 027 strains. PLoS One 7:e37263.
168. Ackermann HW. 1998. Tailed bacteriophages: the order caudovirales. Adv Virus Res 51:135-201.
169. Shan J, Patel KV, Hickenbotham PT, Nale JY, Hargreaves KR, Clokie MR. 2012. Prophage carriage and diversity within clinically relevant strains of Clostridium difficile. Appl Environ Microbiol 78:6027-34.
170. Phothichaisri W, Ounjai P, Phetruen T, Janvilisri T, Khunrae P, Singhakaew S, Wangroongsarb P, Chankhamhaengdecha S. 2018. Characterization of Bacteriophages Infecting Clinical Isolates of Clostridium difficile. Front Microbiol 9:1701.

171. Hyman P, Abedon ST. 2010. Bacteriophage host range and bacterial resistance. Adv Appl Microbiol 70:217-48.
172. Sekulovic O, Ospina Bedoya M, Fivian-Hughes AS, Fairweather NF, Fortier LC. 2015. The Clostridium difficile cell wall protein CwpV confers phase-variable phage resistance. Mol Microbiol 98:329-42.
173. Boudry P, Semenova E, Monot M, Datsenko KA, Lopatina A, Sekulovic O, Ospina-Bedoya M, Fortier LC, Severinov K, Dupuy B, Soutourina O. 2015. Function of the CRISPR-Cas System of the Human Pathogen Clostridium difficile. mBio 6:e01112-15.
174. Hargreaves KR, Flores CO, Lawley TD, Clokie MR. 2014. Abundant and diverse clustered regularly interspaced short palindromic repeat spacers in Clostridium difficile strains and prophages target multiple phage types within this pathogen. mBio 5:e01045-13.
175. Fortier LC. 2018. Bacteriophages Contribute to Shaping Clostridioides (Clostridium) difficile Species. Front Microbiol 9:2033.
176. Koskella B. 2014. Bacteria-phage interactions across time and space: merging local adaptation and time-shift experiments to understand phage evolution. Am Nat 184 Suppl 1:S9-21.
177. Bourdin G, Navarro A, Sarker SA, Pittet AC, Qadri F, Sultana S, Cravioto A, Talukder KA, Reuteler G, Brussow H. 2014. Coverage of diarrhoea-associated Escherichia coli isolates from different origins with two types of phage cocktails. Microb Biotechnol 7:165-76.
178. Sarker SA, 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, El Aidy S, Berger B, Brussow H. 2016. Oral Phage Therapy of Acute Bacterial Diarrhea With Two Coliphage Preparations: A Randomized Trial in Children From Bangladesh. EBioMedicine 4:124-37.
179. Wittmann J, Riedel T, Bunk B, Spröer C, Gronow S, Overmann J. 2015. Complete Genome Sequence of the Novel Temperate Clostridium difficile Phage phiCDIF1296T. Genome Announc 3.
180. Garneau JR, Sekulovic O, Dupuy B, Soutourina O, Monot M, Fortier LC. 2018. High Prevalence and Genetic Diversity of Large phiCD211 (phiCDIF1296T)-Like Prophages in Clostridioides difficile. Appl Environ Microbiol 84.
181. Goh S, Ong PF, Song KP, Riley TV, Chang BJ. 2007. The complete genome sequence of Clostridium difficile phage phiC2 and comparisons to phiCD119 and inducible prophages of CD630. Microbiology (Reading) 153:676-685.
182. Mayer MJ, Narbad A, Gasson MJ. 2008. Molecular characterization of a Clostridium difficile bacteriophage and its cloned biologically active endolysin. J Bacteriol 190:6734-40.
183. Nale JY, Spencer J, Hargreaves KR, Buckley AM, Trzepiński P, Douce GR, Clokie MRJ. 2016. Bacteriophage Combinations Significantly Reduce Clostridium difficile growth In Vitro and Proliferation In Vivo. Antimicrobial Agents and Chemotherapy 60:968-981.
184. Whittle MJ, Bilverstone TW, van Esveld RJ, Lücke AC, Lister MM, Kuehne SA, Minton NP. 2022. A Novel Bacteriophage with Broad Host Range against Clostridioides difficile Ribotype 078 Supports SlpA as the Likely Phage Receptor. Microbiol Spectr 10:e0229521.
185. Govind R, Fralick JA, Rolfe RD. 2006. Genomic organization and molecular characterization of Clostridium difficile bacteriophage PhiCD119. J Bacteriol 188:2568-77.
186. Goh S, Riley TV, Chang BJ. 2005. Isolation and Characterization of Temperate Bacteriophages of Clostridium difficile. Applied and Environmental Microbiology 71:1079-1083.
187. Sekulovic O, Garneau JR, Neron A, Fortier LC. 2014. Characterization of temperate phages infecting Clostridium difficile isolates of human and animal origins. Appl Environ Microbiol 80:2555-63.
188. Sekulovic O, Meessen-Pinard M, Fortier LC. 2011. Prophage-stimulated toxin production in Clostridium difficile NAP1/027 lysogens. J Bacteriol 193:2726-34.
189. Meessen-Pinard M, Sekulovic O, Fortier L-C. 2012. Evidence of in vivo Prophage Induction during Clostridium difficile Infection. Applied and Environmental Microbiology 78:7662.

190. Rashid SJ, Barylski J, Hargreaves KR, Millard AA, Vinner GK, Clokie MR. 2016. Two Novel Myoviruses from the North of Iraq Reveal Insights into Clostridium difficile Phage Diversity and Biology. Viruses 8.
191. Ramírez-Vargas G, Goh S, Rodríguez C. 2018. The Novel Phages phiCD5763 and phiCD2955 Represent Two Groups of Big Plasmidial Siphoviridae Phages of Clostridium difficile. Frontiers in Microbiology 9.
192. Li T, Zhang Y, Dong K, Kuo CJ, Li C, Zhu YQ, Qin J, Li QT, Chang YF, Guo X, Zhu Y. 2020. Isolation and Characterization of the Novel Phage JD032 and Global Transcriptomic Response during JD032 Infection of Clostridioides difficile Ribotype 078. mSystems 5.
193. Hinc K, Kabała M, Iwanicki A, Martirosian G, Negri A, Obuchowski M. 2021. Complete genome sequence of the newly discovered temperate Clostridioides difficile bacteriophage phiCDKH01 of the family Siphoviridae. Archives of Virology 166:2305-2310.
194. Chan BK, Abedon ST, Loc-Carrillo C. 2013. Phage cocktails and the future of phage therapy. Future Microbiol 8:769-83.
195. Chan BK, Abedon ST. 2012. Phage therapy pharmacology phage cocktails. Adv Appl Microbiol 78:1-23.
196. Nale JY, Redgwell TA, Millard A, Clokie MRJ. 2018. Efficacy of an Optimised Bacteriophage Cocktail to Clear Clostridium difficile in a Batch Fermentation Model. Antibiotics (Basel) 7.
197. Vasu K, Nagaraja V. 2013. Diverse functions of restriction-modification systems in addition to cellular defense. Microbiol Mol Biol Rev 77:53-72.
198. Cheng K, Wilkinson M, Chaban Y, Wigley DB. 2020. A conformational switch in response to Chi converts RecBCD from phage destruction to DNA repair. Nat Struct Mol Biol 27:71-77.
199. Rostøl JT, Marraffini L. 2019. (Ph)ighting Phages: How Bacteria Resist Their Parasites. Cell Host Microbe 25:184-194.
200. Arias CF, Acosta FJ, Bertocchini F, Herrero MA, Fernández-Arias C. 2022. The coordination of anti-phage immunity mechanisms in bacterial cells. Nature Communications 13:7412.
201. Lopatina A, Tal N, Sorek R. 2020. Abortive Infection: Bacterial Suicide as an Antiviral Immune Strategy. Annual Review of Virology 7:371-384.
202. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science 315:1709-12.
203. Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology 151:2551-2561.
204. Rath D, Amlinger L, Rath A, Lundgren M. 2015. The CRISPR-Cas immune system: biology, mechanisms and applications. Biochimie 117:119-28.
205. Kim JG, Garrett S, Wei Y, Graveley BR, Terns MP. 2019. CRISPR DNA elements controlling site-specific spacer integration and proper repeat length by a Type II CRISPR–Cas system. Nucleic Acids Research 47:8632-8648.
206. Noren T, Akerlund T, Back E, Sjoberg L, Persson I, Alriksson I, Burman LG. 2004. Molecular epidemiology of hospital-associated and community-acquired Clostridium difficile infection in a Swedish county. J Clin Microbiol 42:3635-43.
207. Sun CL, Barrangou R, Thomas BC, Horvath P, Fremaux C, Banfield JF. 2013. Phage mutations in response to CRISPR diversification in a bacterial population. Environ Microbiol 15:463-70.
208. Deveau H, Barrangou R, Garneau JE, Labonte J, Fremaux C, Boyaval P, Romero DA, Horvath P, Moineau S. 2008. Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. J Bacteriol 190:1390-400.
209. Semenova E, Jore MM, Datsenko KA, Semenova A, Westra ER, Wanner B, van der Oost J, Brouns SJ, Severinov K. 2011. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. Proc Natl Acad Sci U S A 108:10098-103.

210. Seed KD, Lazinski DW, Calderwood SB, Camilli A. 2013. A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. Nature 494:489-91.
211. Goh S, Chang BJ, Riley TV. 2005. Effect of phage infection on toxin production by Clostridium difficile. J Med Microbiol 54:129-35.
212. Tan KS, Wee BY, Song KP. 2001. Evidence for holin function of tcdE gene in the pathogenicity of Clostridium difficile. J Med Microbiol 50:613-619.
213. Govind R, Vediyappan G, Rolfe RD, Dupuy B, Fralick JA. 2009. Bacteriophage-mediated toxin gene regulation in Clostridium difficile. J Virol 83:12037-45.
214. Hargreaves KR, Clokie MRJ. 2014. Clostridium difficile phages: still difficult? Frontiers in Microbiology 5.
215. Meader E, Mayer MJ, Gasson MJ, Steverding D, Carding SR, Narbad A. 2010. Bacteriophage treatment significantly reduces viable Clostridium difficile and prevents toxin production in an in vitro model system. Anaerobe 16:549-554.
216. Meader E, Mayer MJ, Steverding D, Carding SR, Narbad A. 2013. Evaluation of bacteriophage therapy to control Clostridium difficile and toxin production in an in vitro human colon model system. Anaerobe 22:25-30.
217. Nale JY, Chutia M, Carr P, Hickenbotham PT, Clokie MR. 2016. 'Get in Early'; Biofilm and Wax Moth (Galleria mellonella) Models Reveal New Insights into the Therapeutic Potential of Clostridium difficile Bacteriophages. Front Microbiol 7:1383.
218. Ramesh V, Fralick JA, Rolfe RD. 1999. Prevention of Clostridium difficile -induced ileocecitis with Bacteriophage. Anaerobe 5:69-78.
219. Ott SJ, Waetzig GH, Rehman A, Moltzau-Anderson J, Bharti R, Grasis JA, Cassidy L, Tholey A, Fickenscher H, Seegert D, Rosenstiel P, Schreiber S. 2017. Efficacy of Sterile Fecal Filtrate Transfer for Treating Patients With Clostridium difficile Infection. Gastroenterology 152:799-811.e7.
220. Basdew IH, Laing MD. 2015. Investigation of the lytic ability of South African bacteriophages specific for Staphylococcus aureus, associated with bovine mastitis. Biocontrol Science and Technology 25:429-443.
221. Milase RN. 2020. *Development of a novel bacteriophage cocktail for treatment of American foulbrood disease affecting honeybees of the Western Cape (South Africa)*. https://imbm.co.za/researchers/phds/ridwaan-nazeer-milase/. Accessed 28 July 2020.
222. Bragg R, van der Westhuizen W, Lee JY, Coetsee E, Boucher C. 2014. Bacteriophages as potential treatment option for antibiotic resistant bacteria. Adv Exp Med Biol 807:97-110.
223. van Zyl LJ, Abrahams Y, Stander EA, Kirby-McCollough B, Jourdain R, Clavaud C, Breton L, Trindade M. 2018. Novel phages of healthy skin metaviromes from South Africa. Sci Rep 8:12265.
224. Abdool Karim SS, Churchyard GJ, Karim QA, Lawn SD. 2009. HIV infection and tuberculosis in South Africa: an urgent need to escalate the public health response. Lancet 374:921-33.
225. Pheiffer C, Pillay-van Wyk V, Joubert JD, Levitt N, Nglazi MD, Bradshaw D. 2018. The prevalence of type 2 diabetes in South Africa: a systematic review protocol. BMJ Open 8:e021029.
226. Kullin BR, Reid S, Abratt V. 2018. Clostridium difficile in patients attending tuberculosis hospitals in Cape Town, South Africa, 2014-2015. Afr J Lab Med 7:846.
227. Chibani-Chennoufi S, Bruttin A, Dillmann ML, Brüssow H. 2004. Phage-host interaction: an ecological perspective. J Bacteriol 186:3677-86.
228. Ackermann HW. 2007. 5500 Phages examined in the electron microscope. Arch Virol 152:227-43.
229. Batinovic S, Wassef F, Knowler SA, Rice DTF, Stanton CR, Rose J, Tucci J, Nittami T, Vinh A, Drummond GR, Sobey CG, Chan HT, Seviour RJ, Petrovski S, Franks AE. 2019. Bacteriophages in Natural and Artificial Environments. Pathogens 8:100.

230. Chaturongakul S, Ounjai P. 2014. Phage–host interplay: examples from tailed phages and Gram-negative bacterial pathogens. Frontiers in Microbiology 5.
231. Willner D, Furlan M, Haynes M, Schmieder R, Angly FE, Silva J, Tammadoni S, Nosrat B, Conrad D, Rohwer F. 2009. Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. PLoS One 4:e7370.
232. Pride DT, Salzman J, Haynes M, Rohwer F, Davis-Long C, White RA, 3rd, Loomer P, Armitage GC, Relman DA. 2012. Evidence of a robust resident bacteriophage population revealed through analysis of the human salivary virome. Isme j 6:915-26.
233. Foulongne V, Sauvage V, Hebert C, Dereure O, Cheval J, Gouilh MA, Pariente K, Segondy M, Burguière A, Manuguerra JC, Caro V, Eloit M. 2012. Human skin microbiota: high diversity of DNA viruses identified on the human skin by high throughput sequencing. PLoS One 7:e38499.
234. Manrique P, Bolduc B, Walk ST, van der Oost J, de Vos WM, Young MJ. 2016. Healthy human gut phageome. Proc Natl Acad Sci U S A 113:10400-5.
235. Hargreaves KR, Colvin HV, Patel KV, Clokie JJ, Clokie MR. 2013. Genetically diverse Clostridium difficile strains harboring abundant prophages in an estuarine environment. Appl Environ Microbiol 79:6236-43.
236. Bonilla N, Rojas MI, Netto Flores Cruz G, Hung SH, Rohwer F, Barr JJ. 2016. Phage on tap-a quick and efficient protocol for the preparation of bacteriophage laboratory stocks. PeerJ 4:e2261.
237. Yang Y-W, Chen M-K, Yang B-Y, Huang X-J, Zhang X-R, He L-Q, Zhang J, Hua Z-C. 2015. Use of 16S rRNA Gene-Targeted Group-Specific Primers for Real-Time PCR Analysis of Predominant Bacteria in Mouse Feces. Applied and environmental microbiology 81:6749-6756.
238. Nale JY, Thanki AM, Rashid SJ, Shan J, Vinner GK, Dowah ASA, Cheng JKJ, Sicheritz-Pontén T, Clokie MRJ. 2022. Diversity, Dynamics and Therapeutic Application of Clostridioides difficile Bacteriophages. Viruses 14:2772.
239. Sekulović O, Fortier L-C. 2016. Characterization of Functional Prophages in Clostridium difficile, p 143-165. *In* Roberts AP, Mullany P (ed), Clostridium difficile: Methods and Protocols doi:10.1007/978-1-4939-6361-4_11. Springer New York, New York, NY.
240. Sangster W, Hegarty JP, Stewart DB, Sr. 2015. Phage tail-like particles kill *Clostridium difficile* and represent an alternative to conventional antibiotics. Surgery 157:96-103.
241. Hegarty JP, Sangster W, Ashley RE, Myers R, Hafenstein S, Stewart DB. 2016. Induction and Purification of C. difficile Phage Tail-Like Particles. Clostridium difficile 1476:167-175.
242. Gebhart D, Williams SR, Bishop-Lilly KA, Govoni GR, Willner KM, Butani A, Sozhamannan S, Martin D, Fortier LC, Scholl D. 2012. Novel high-molecular-weight, R-type bacteriocins of clostridium difficile. Journal of Bacteriology 194:6240-6247.
243. Aframian N, Omer Bendori S, Kabel S, Guler P, Stokar-Avihail A, Manor E, Msaeed K, Lipsman V, Grinberg I, Mahagna A, Eldar A. 2022. Dormant phages communicate via arbitrium to control exit from lysogeny. Nature Microbiology 7:145-153.
244. Ballesté E, Blanch AR, Muniesa M, García-Aljaro C, Rodríguez-Rubio L, Martín-Díaz J, Pascual-Benito M, Jofre J. 2022. Bacteriophages in sewage: abundance, roles, and applications. FEMS Microbes 3.
245. Heuler J, Fortier L-C, Sun X. 2021. Clostridioides difficile phage biology and application. FEMS Microbiology Reviews 45.
246. Royer ALM, Umansky AA, Allen M-M, Garneau JR, Ospina-Bedoya M, Kirk JA, Govoni G, Fagan RP, Soutourina O, Fortier L-C. Clostridioides difficile S-Layer Protein A (SlpA) Serves as a General Phage Receptor. Microbiology Spectrum 0:e03894-22.
247. Letellier L, Boulanger P, Plançon L, Jacquot P, Santamaria M. 2004. Main features on tailed phage, host recognition and DNA uptake. Front Biosci 9:1228-339.
248. Kanda R, Griffin P, James HA, Fothergill J. 2003. Pharmaceutical and personal care products in sewage treatment works. J Environ Monit 5:823-30.

249. Yang Y, Ok YS, Kim KH, Kwon EE, Tsang YF. 2017. Occurrences and removal of pharmaceuticals and personal care products (PPCPs) in drinking water and water/sewage treatment plants: A review. Sci Total Environ 596-597:303-320.
250. Dinh Q, Moreau-Guigon E, Labadie P, Alliot F, Teil MJ, Blanchard M, Eurin J, Chevreuil M. 2017. Fate of antibiotics from hospital and domestic sources in a sewage network. Sci Total Environ 575:758-766.
251. Seeley ND, Primrose SB. 1980. The Effect of Temperature on the Ecology of Aquatic Bacteriophages. Journal of General Virology 46:87-95.
252. Jończyk E, Kłak M, Międzybrodzki R, Górski A. 2011. The influence of external factors on bacteriophages--review. Folia Microbiol (Praha) 56:191-200.
253. Tran HN, Le GT, Nguyen DT, Juang RS, Rinklebe J, Bhatnagar A, Lima EC, Iqbal HMN, Sarmah AK, Chao HP. 2021. SARS-CoV-2 coronavirus in water and wastewater: A critical review about presence and concern. Environ Res 193:110265.
254. Thakkar JR, Sabara PH, Koringa PG. 2017. Exploring Metagenomes Using Next-Generation Sequencing, p 29-40. *In* Singh RP, Kothari R, Koringa PG, Singh SP (ed), Understanding Host-Microbiome Interactions - An Omics Approach: Omics of Host-Microbiome Association doi:10.1007/978-981-10-5050-3_3. Springer Singapore, Singapore.
255. Benler S, Koonin EV. 2021. Fishing for phages in metagenomes: what do we catch, what do we miss? Curr Opin Virol 49:142-150.
256. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res 44:W16-21.
257. Roux S, Enault F, Hurwitz BL, Sullivan MB. 2015. VirSorter: mining viral signal from microbial genomic data. PeerJ 3:e985.
258. Guo J, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, Delmont TO, Pratama AA, Gazitúa MC, Vik D, Sullivan MB, Roux S. 2021. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. Microbiome 9:37.
259. Akhter S, Aziz RK, Edwards RA. 2012. PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. Nucleic Acids Res 40:e126.
260. Ren J, Ahlgren NA, Lu YY, Fuhrman JA, Sun F. 2017. VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. Microbiome 5:69.
261. Reis-Cunha JL, Bartholomeu DC, Manson AL, Earl AM, Cerqueira GC. 2019. ProphET, prophage estimation tool: A stand-alone prophage sequence prediction tool with self-updating reference database. PLoS One 14:e0223364.
262. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: a fast phage search tool. Nucleic Acids Res 39:W347-52.
263. Srividhya KV, Rao GV, Raghavenderan L, Mehta P, Prilusky J, Manicka S, Sussman JL, Krishnaswamy S. 2006. Database and Comparative Identification of Prophages, p 863-868. *In* Huang D-S, Li K, Irwin GW (ed), Intelligent Control and Automation: International Conference on Intelligent Computing, ICIC 2006 Kunming, China, August 16–19, 2006 doi:10.1007/978-3-540-37256-1_110. Springer Berlin Heidelberg, Berlin, Heidelberg.
264. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955-64.
265. Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 32:11-6.
266. Hargreaves KR. 2016. Isolation and Characterisation of Bacteriophages Infecting Environmental Strains of Clostridium DifficileUniversity of Leicester.
267. Cairns MD, Preston MD, Hall CL, Gerding DN, Hawkey PM, Kato H, Kim H, Kuijper EJ, Lawley TD, Pituch H, Reid S, Kullin B, Riley TV, Solomon K, Tsai PJ, Weese JS, Stabler RA, Wren BW. 2017. Comparative Genome Analysis and Global Phylogeny of the Toxin Variant Clostridium

difficile PCR Ribotype 017 Reveals the Evolution of Two Independent Sublineages. J Clin Microbiol 55:865-876.
268. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072-1075.
269. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043-55.
270. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.
271. Shaffer M, Borton MA, McGivern BB, Zayed AA, La Rosa Sabina L, Solden LM, Liu P, Narrowe AB, Rodríguez-Ramos J, Bolduc B, Gazitúa MC, Daly RA, Smith GJ, Vik DR, Pope PB, Sullivan MB, Roux S, Wrighton Kelly C. 2020. DRAM for distilling microbial metabolism to automate the curation of microbiome function. Nucleic Acids Research 48:8883-8900.
272. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. 2021. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nature Biotechnology 39:578-585.
273. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM. 2016. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biology 17:132.
274. Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Research 49:W293-W296.
275. Gilchrist CLM, Chooi YH. 2021. Clinker & clustermap.js: Automatic generation of gene cluster comparison figures. Bioinformatics doi:10.1093/bioinformatics/btab007.
276. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, Gladstone RA, Lo S, Beaudoin C, Floto RA, Frost SDW, Corander J, Bentley SD, Parkhill J. 2020. Producing polished prokaryotic pangenomes with the Panaroo pipeline. Genome Biology 21:180.
277. Katoh K, Rozewicki J, Yamada KD. 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform 20:1160-1166.
278. Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. 2019. ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. Molecular Biology and Evolution 37:291-294.
279. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Molecular Biology and Evolution 37:1530-1534.
280. Tesson F, Hervé A, Mordret E, Touchon M, d'Humières C, Cury J, Bernheim A. 2022. Systematic and quantitative view of the antiviral arsenal of prokaryotes. Nature Communications 13:2561.
281. Abby SS, Néron B, Ménager H, Touchon M, Rocha EPC. 2014. MacSyFinder: A Program to Mine Genomes for Molecular Systems with an Application to CRISPR-Cas Systems. PLOS ONE 9:e110726.
282. Hargreaves KR, Clokie MR. 2015. A Taxonomic Review of Clostridium difficile Phages and Proposal of a Novel Genus, "Phimmp04likevirus". Viruses 7:2534-41.
283. Campbell A. 1977. Defective Bacteriophages and Incomplete Prophages, p 259-328. *In* Fraenkel-Conrat H, Wagner RR (ed), Regulation and Genetics: Bacterial DNA Viruses doi:10.1007/978-1-4684-2715-8_3. Springer US, Boston, MA.
284. Garro AJ, Marmur J. 1970. Defective bacteriophages. Journal of Cellular Physiology 76:253-263.
285. Bobay LM, Rocha EP, Touchon M. 2013. The adaptation of temperate bacteriophages to their host genomes. Mol Biol Evol 30:737-51.
286. Ester M, Kriegel H-P, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise, p 226-231. *In* (ed),

287. Roux S, Adriaenssens EM, Dutilh BE, Koonin EV, Kropinski AM, Krupovic M, Kuhn JH, Lavigne R, Brister JR, Varsani A, Amid C, Aziz RK, Bordenstein SR, Bork P, Breitbart M, Cochrane GR, Daly RA, Desnues C, Duhaime MB, Emerson JB, Enault F, Fuhrman JA, Hingamp P, Hugenholtz P, Hurwitz BL, Ivanova NN, Labonté JM, Lee K-B, Malmstrom RR, Martinez-Garcia M, Mizrachi IK, Ogata H, Páez-Espino D, Petit M-A, Putonti C, Rattei T, Reyes A, Rodriguez-Valera F, Rosario K, Schriml L, Schulz F, Steward GF, Sullivan MB, Sunagawa S, Suttle CA, Temperton B, Tringe SG, Thurber RV, Webster NS, Whiteson KL, et al. 2019. Minimum Information about an Uncultivated Virus Genome (MIUViG). Nature Biotechnology 37:29-37.
288. Makarova KS, Wolf YI, Koonin EV. 2018. Classification and Nomenclature of CRISPR-Cas Systems: Where from Here? Crispr j 1:325-336.
289. Chaudhuri A, Halder K, Datta A. 2022. Classification of CRISPR/Cas system and its application in tomato breeding. Theoretical and Applied Genetics 135:367-387.
290. Maikova A, Boudry P, Shiriaeva A, Vasileva A, Boutserin A, Medvedeva S, Semenova E, Severinov K, Soutourina O. 2021. Protospacer-Adjacent Motif Specificity during Clostridioides difficile Type I-B CRISPR-Cas Interference and Adaptation. mBio 12:e0213621.
291. Andersen JM, Shoup M, Robinson C, Britton R, Olsen KEP, Barrangou R. 2016. CRISPR Diversity and Microevolution in Clostridium difficile. Genome Biology and Evolution 8:2841-2855.
292. Koonin EV, Makarova KS, Zhang F. 2017. Diversity, classification and evolution of CRISPR-Cas systems. Current Opinion in Microbiology 37:67-78.
293. Marraffini LA. 2015. CRISPR-Cas immunity in prokaryotes. Nature 526:55-61.
294. Shmakov S, Abudayyeh OO, Makarova KS, Wolf YI, Gootenberg JS, Semenova E, Minakhin L, Joung J, Konermann S, Severinov K, Zhang F, Koonin EV. 2015. Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. Mol Cell 60:385-97.
295. Abudayyeh OO, Gootenberg JS, Konermann S, Joung J, Slaymaker IM, Cox DBT, Shmakov S, Makarova KS, Semenova E, Minakhin L, Severinov K, Regev A, Lander ES, Koonin EV, Zhang F. 2016. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. Science 353:aaf5573.
296. Parma DH, Snyder M, Sobolevski S, Nawroz M, Brody E, Gold L. 1992. The Rex system of bacteriophage lambda: tolerance and altruistic cell death. Genes Dev 6:497-510.
297. Stabler RA, He M, Dawson L, Martin M, Valiente E, Corton C, Lawley TD, Sebaihia M, Quail MA, Rose G, Gerding DN, Gibert M, Popoff MR, Parkhill J, Dougan G, Wren BW. 2009. Comparative genome and phenotypic analysis of Clostridium difficile 027 strains provides insight into the evolution of a hypervirulent bacterium. Genome Biol 10:R102.
298. Ershova AS, Rusinov IS, Spirin SA, Karyagina AS, Alexeevski AV. 2015. Role of Restriction-Modification Systems in Prokaryotic Evolution and Ecology. Biochemistry (Mosc) 80:1373-86.
299. Purdy D, O'Keeffe TA, Elmore M, Herbert M, McLeod A, Bokori-Brown M, Ostrowski A, Minton NP. 2002. Conjugative transfer of clostridial shuttle vectors from Escherichia coli to Clostridium difficile through circumvention of the restriction barrier. Mol Microbiol 46:439-52.
300. Lowey B, Whiteley AT, Keszei AFA, Morehouse BR, Mathews IT, Antine SP, Cabrera VJ, Kashin D, Niemann P, Jain M, Schwede F, Mekalanos JJ, Shao S, Lee ASY, Kranzusch PJ. 2020. CBASS Immunity Uses CARF-Related Effectors to Sense 3'-5'- and 2'-5'-Linked Cyclic Oligonucleotide Signals and Protect Bacteria from Phage Infection. Cell 182:38-49.e17.
301. Millman A, Melamed S, Leavitt A, Doron S, Bernheim A, Hör J, Garb J, Bechon N, Brandis A, Lopatina A, Ofir G, Hochhauser D, Stokar-Avihail A, Tal N, Sharir S, Voichek M, Erez Z, Ferrer JLM, Dar D, Kacen A, Amitai G, Sorek R. 2022. An expanded arsenal of immune systems that protect bacteria from phages. Cell Host Microbe 30:1556-1569.e5.
302. Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, Amitai G, Sorek R. 2018. Systematic discovery of antiphage defense systems in the microbial pangenome. Science 359.
303. Cheng R, Huang F, Wu H, Lu X, Yan Y, Yu B, Wang X, Zhu B. 2021. A nucleotide-sensing endonuclease from the Gabija bacterial defense system. Nucleic Acids Res 49:5216-5229.

304. Vassallo CN, Doering CR, Littlehale ML, Teodoro GIC, Laub MT. 2022. A functional selection reveals previously undetected anti-phage defence systems in the E. coli pangenome. Nature Microbiology 7:1568-1579.
305. Howard-Varona C, Hargreaves KR, Abedon ST, Sullivan MB. 2017. Lysogeny in nature: mechanisms, impact and ecology of temperate phages. Isme j 11:1511-1520.
306. Barnhart BJ, Cox SH, Jett JH. 1976. Prophage induction and inactivation by UV light. J Virol 18:950-5.
307. Raya RR, H'Bert E M. 2009. Isolation of Phage via Induction of Lysogens. Methods Mol Biol 501:23-32.
308. Sekulovic O, Fortier LC. 2016. Characterization of Functional Prophages in Clostridium difficile. Methods Mol Biol 1476:143-65.
309. Selle K, Fletcher JR, Tuson H, Schmitt DS, McMillan L, Vridhambal GS, Rivera AJ, Montgomery SA, Fortier L-C, Barrangou R, Theriot CM, Ousterout DG. 2020. *In Vivo* Targeting of *Clostridioides difficile* Using Phage-Delivered CRISPR-Cas3 Antimicrobials. mBio 11:e00019-20.
310. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2013. GenBank. Nucleic Acids Res 41:D36-42.
311. Mahony DE, Bell PD, Easterbrook KB. 1985. Two bacteriophages of Clostridium difficile. J Clin Microbiol 21:251-4.
312. Nagy E, Foldes J. 1991. Electron microscopic investigation of lysogeny of Clostridium difficile strains isolated from antibiotic-associated diarrhea cases and from healthy carriers. Apmis 99:321-6.
313. Jurczak-Kurek A, Gąsior T, Nejman-Faleńczyk B, Bloch S, Dydecka A, Topka G, Necel A, Jakubowska-Deredas M, Narajczyk M, Richert M, Mieszkowska A, Wróbel B, Węgrzyn G, Węgrzyn A. 2016. Biodiversity of bacteriophages: morphological and biological properties of a large group of phages isolated from urban sewage. Scientific Reports 6:34338.
314. Scholl D. 2017. Phage Tail–Like Bacteriocins. Annual Review of Virology 4:453-467.
315. Kullin B, Knight DR, Reid S, Abratt V, Riley TV. 2019. Whole Genome Sequence Analysis of *Clostridium difficile* Ribotype 017 Strains in Hospital Patients in Cape Town, South Africa. Australian Society for Microbiology Annual Scientific Meeting, Adelaide, Australia.
316. Maciejewska B, Olszak T, Drulis-Kawa Z. 2018. Applications of bacteriophages versus phage enzymes to combat and cure bacterial infections: an ambitious and also a realistic application? Appl Microbiol Biotechnol 102:2563-2581.
317. Fujimoto K, Uematsu S. 2022. Phage therapy for Clostridioides difficile infection. Front Immunol 13:1057892.
318. Dedrick RM, Guerrero-Bustamante CA, Garlena RA, Russell DA, Ford K, Harris K, Gilmour KC, Soothill J, Jacobs-Sera D, Schooley RT, Hatfull GF, Spencer H. 2019. Engineered bacteriophages for treatment of a patient with a disseminated drug-resistant Mycobacterium abscessus. Nat Med 25:730-733.
319. Stewart FM, Levin BR. 1984. The population biology of bacterial viruses: why be temperate. Theor Popul Biol 26:93-117.
320. Hargreaves KR, Kropinski AM, Clokie MR. 2014. What does the talking?: quorum sensing signalling genes discovered in a bacteriophage genome. PLoS One 9:e85131.
321. Meeske AJ, Jia N, Cassel AK, Kozlova A, Liao J, Wiedmann M, Patel DJ, Marraffini LA. 2020. A phage-encoded anti-CRISPR enables complete evasion of type VI-A CRISPR-Cas immunity. Science 369:54-59.
322. Labrie SJ, Tremblay DM, Moisan M, Villion M, Magadï¿½n AH, Campanacci Vr, Cambillau C, Moineau S. 2012. Involvement of the Major Capsid Protein and Two Early-Expressed Phage Genes in the Activity of the Lactococcal Abortive Infection Mechanism AbiT. Applied and Environmental Microbiology 78:6890-6899.
323. Krüger DH, Bickle TA. 1983. Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. Microbiological Reviews 47:345-360.

324. Warren RAJ. 1980. Modified Bases in Bacteriophage DNAs. Annual Review of Microbiology 34:137-158.
325. Sebaihia M, Wren BW, Mullany P, Fairweather NF, Minton N, Stabler R, Thomson NR, Roberts AP, Cerdeño-Tárraga AM, Wang H, Holden MT, Wright A, Churcher C, Quail MA, Baker S, Bason N, Brooks K, Chillingworth T, Cronin A, Davis P, Dowd L, Fraser A, Feltwell T, Hance Z, Holroyd S, Jagels K, Moule S, Mungall K, Price C, Rabbinowitsch E, Sharp S, Simmonds M, Stevens K, Unwin L, Whithead S, Dupuy B, Dougan G, Barrell B, Parkhill J. 2006. The multidrug-resistant human pathogen Clostridium difficile has a highly mobile, mosaic genome. Nat Genet 38:779-86.