UNDERSTANDING *Clostridium difficile* AND THE BACTERIOPHAGES FROM THE ENVIRONMENT

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

By

Saroa Jamal Rashid BSc, MSc,

Department of Infection, Immunity and Inflammation

University of Leicester

November 2016

Abstract

UNDERSTANDING *Clostridium difficile* AND THE BACTERIOPHAGES FROM THE ENVIRONMENT

By

Saroa Jamal Rashid

Clostridium difficile is an important nosocomial pathogen that causes significant human morbidity and mortality. The environment is a natural reservoir for *C. difficile* where it is found in soils, rivers, and other natural settings. It is not clear whether these strains are active, or present as a result of human contamination. To address this, the physiological characterisation of isogenic pairs was tested to determine if strains from the environment had different properties to those derived from a clinical setting. Clinical strains were less motile but produced more toxins and spores than environmental strains.

Little is known about environmental strains of *C. difficile* outside Europe, the USA, and Australia. To explore new areas, *C. difficile* and phages were isolated from environmental samples collected in Kurdistan in northern Iraq. Sixty-one strains were isolated and classified into 9 ribotypes including a new ribotype, R691. The two strains examined from dominant ribotype R604, were non-toxigenic and related to each other but distinct from most described clades of *C. difficile*. Evidence for phages playing important roles in environmental *C. difficile* dynamics can be seen from the large diverse prophage carriage within strains, and the extensive CRISPR system.

Seventeen new phages were isolated and shown to infect ribotype 078 for which few phages exist. Host-range analysis showed that these phages can infect most clinically relevant ribotypes including the novel observation that two myoviruses can infect R027. Genome analysis of three phages revealed that they are distinct but related to known *C. difficile* phages. Interestingly CDKM15 is the first phage isolated to have an active CRISPR system. This study suggests that the natural environment is a potential reservoir for genetically diverse *C. difficile* strains and phages that are regionally structured. This could play a role in the emergence of new strains in hospitals.

ACKNOWLEDGEMENTS

To Almighty God, the most merciful and most benevolent. The completion of this study would not have been possible without the kind help, support and encouragement from many individuals. I would like to give a tremendous thank you to my supervisor Professor Martha Clokie for her assistance and support throughout the project. I was very fortunate to have the opportunity to conduct research under her supervision.

I would like to thank my annual review panel members, Dr. Shaun Heaphy and Dr. Kumar Rajakumar for their advice and guidance throughout my research. My thanks also go to Dr Andrew Milliard, Dr Jakub Barylski and Dr Katherine R. Hargreaves for their assistance in bioinformatic analysis and advice throughout preparation papers for publication.

I would like to gratefully acknowledge Dr Janet Nale who guided me on a daily basis and answered my many questions, as well as for reading over my chapters and giving me authentic feedback. My thanks also to Dr Maria Viskaduraki for her assistance in statistical analysis. Also I would like to acknowledge Dr Jinyu Shan, Peter Hickenbotham and Natalie Allcock for their technical support. My thanks also goes to /Prof M Wilcox, Dr Warren Fawley and Dr Jane Freeman from the *C. difficile* ribotyping network, for their assistance in ribotyping *C. difficile* isolates.

Special thanks also go to all past and present members of lab 105 and 107 especially Ahmed, Guillermo, Jazz, Chris, Anisha, Tolis, Neda, Rin, Ali, Depesh and Sophie for their help in the lab during my PhD project.

I would like to gratefully acknowledge my sister Sana Rashid and Govind Hedayet from the research centre University of Koya in Kurdistan for their assistance and expertise with sampling of sites in Kurdistan. I would like to thank also my husband Dr Fuad Hassen for his assistance in bringing the samples in to the UK.

Last but not least, I could not have completed my PhD journey without love, patience, and understanding from my amazing children Mohamed, Zheer and Dania. I am also indebted to my mum, dad, my sisters, brothers and auntie Kolyzar for their consistent support, inspiration and motivation I needed to succeed.

List of publications

RASHID, S. R. J., CLOKIE, M. R. J. & MILLARD, A. D. 2016. Draft Genome Sequences of Three Novel Clostridium Isolates from Northern Iraq. Genome Announcements, 4, e00033-16.

RASHID, **S**., BARYLSKI, J., HARGREAVES, K., MILLARD, A., VINNER, G. & CLOKIE, M. 2016. Two Novel Myoviruses from the North of Iraq Reveal Insights into *Clostridium difficile* Phage Diversity and Biology. Viruses, 8, 310.

Rashid, S. J., Janet Y. Nale, Millard, A.D and Clokie, M.R.J. Exploring Kurdish *Clostridium difficile* strains: Isolation and characterisation of eleven novel environmental isolates from North of Iraq. (under the preparation).

List of Content

List of Figures

List of Tables

List of Abbreviations

General Abbreviations

AAD: antibiotic associated diarrhoea

ADP: Adenosine diphosphate

AIDS: *acquired immunodeficiency syndrome*.

AP-PCR: arbitrarily-primed polymerase chain reaction

ANOVA: analysis of variance

B1: *C. difficile* group B1 by restriction-endonuclease analysis (REA)

BA: bacteriological agar

BHI: brain heart infusion

BHIS: brain heart infusion supplemented agar

$CaCl_2$: calcium chloride

CCEY: Cycloserine cefoxitin and egg yolk agar (Brazier's selective medium)

CDAD: *C. difficile* associated diarrhoea

CDI: *Clostridium difficile* infection

CFU/ml: Colony-forming unit per millilitre

CI: Ciprofloxacin

CM: Clindamycin

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats

DHB: defribinated horse blood

dNTPs: deoxyribonucleotide triphosphate

EDTA: Ethylene diaminetetra acetic acid

ELISA: Enzyme-linked immunosorbent assay

EBI: European Bioinformatics Institute

FAB: Fastidious anaerobe broth

HCl: hydrogen chloride

HPA: Health Protection Agency

ICTV: International Committee of Viral Taxonomy

IMS: industrial methylated spirits

LTM: long tailed myovirus

M: mitomycin C

MIC: minimum inhibitory concentration

MM: medium myovirus

$MgCl_2$: Magnesium chloride

$MgSO_4$ Magnesium sulphate

MLST: Multilocus sequence type

MLVA: Multilocus variable-number tandem-repeat analysis

MZ: metronidazole

NaCl: Sodium chloride

NCBI: National centre for biotechnology information

NFX: Norfloxacin

NHS: National Health Service

OD: Optical density

PCR: Polymerase chain reaction

PEG: Polyethylene glycol

PFGE: Pulsed-field gel electrophoresis

PFU/ml: Plaque-forming unit per millilitre

PMC: Pseudomembranous colitis

PTLP: phage tail-like particles

SDS: sodium dodecyl sulphate

SM : SM buffer

SMV: small myovirus

SSC: saline-sodium citrate

TAE: Tris-acetate-EDTA

TBE: Tris-Borate-EDTA

TE: Tris-EDTA

Tris HCl: tris (hydroxymethyl) aminomethane hydrochloride

TEM: Transmission electron microscopy

ToxA: *C. difficile* toxin A

ToxB: *C. difficile* toxin B

UPH_20: Ultra-pure water

VA: vacncomycin

Bioinformatic abbreviations

aa: amino acid

bp: basepairs

BLAST: Basic local alignment search tool

Blastp: protein BLAST

Blastn: nucleotide BLAST

BRIG: Blast Ring Image Generator

CDs: coding DNA sequences

DNA: Deoxyribonucleic acid

dsDNA: double stranded DNA

dsRNA: double stranded RNA

GC: Guanine-cytosine

Kbp: kilo basepairs

MEGA: Molecular Evolutionary Genetics Analaysis

NCBI: National Centre for Biotechnological Information

NGS: next generation sequencing

NJ: neighbour joining

ORF: open reading frame

RNA: Ribonucleic acid

rRNA: ribosomal RNA

ssDNA: single stranded DNA

ssRNA: single stranded RNA

List of units

cm: centimetre

ml: Millilitre

mg: Milligram

M: Molar

Mm: Milli molar

µl: Microliter

µg: Microgram

U: Units

°C: degree centigrade

Chapter 1 General Introduction

1.1 *Clostridium difficile* and *C. difficile* infection

Clostridium difficile is one of the most important healthcare-associated pathogens, it is responsible for antibiotic- associated diarrhoea (AAD), which in severe cases can cause pseudomembranous colitis (PMC) (Barnett, 2012, De Sordi et al., 2015). In the last decade of the twentieth century, *C.difficile* Infection (CDI) prevalence increased, becoming a well-publicised source of hospital-acquired infection in developing countries (Martin et al., 2016). Although numbers have decreased in the UK, *C. difficile* infection or CDI is still considered to be a worldwide health threat. It usually occurs following the disruption of the normal gut microbiota, triggered by patients being given broad-spectrum antibiotics (Buffie et al., 2012, Smits et al., 2016, Wieczorkiewicz et al., 2016). Once present within a hospital environment it is difficult to remove and several epidemics and outbreaks have occurred across Europe and North of America (Blossom and McDonald, 2007, Wiegand et al., 2012). There are ~ 500,000 new cases of CDI every year in the US with an estimated cost of nearly $1 billion (Jose and Madan, 2016). In the UK, 14,165 cases of *C. difficile* infection were reported in 2014/2015, which represents an increase of 6.0% over the year 2013/14, where 13,361 cases were reported. This is the first upsurge in CDI since the enhanced mandatory surveillance of CDI was instigated in 2007, and these perceived increases in CDI are currently under investigation (Health Protection Agency, 2015)

Henceforth the incidence of CDI increased, consequently the cost burden of CDI increased too (Avila et al., 2016). CDI is a major health and economic burden that causes a considerable healthcare burden worldwide (Worth et al., 2016). In 2008, CDI resulted in as much as $4.8 billion in excess costs to US acute-care facilities alone, and the annual costs for management of CDI amount to approximately $800 million in the USA and €3000 million in Europe (Dubberke and Olsen, 2012, Wiegand et al., Bouza, 2012). In 2010, the best estimate of quarterly cost to the NHS associated with each CDI was around £10,000 (Health Protection Agency (2012).

A recent systematic review showed the direct economic burden of CDI in USA between 1988 and June 2014, ranged from $8,911 to $30,049 for hospitalized patients (Nanwa et al., 2015). These costs are due to antibiotic treatments, extended bed stay (due to recurrence which is common), and the costs associated with the decontamination

process that is required to maintain hygienic environment within wards (Poli et al., 2015). There is clearly a pressing need to find alternative approaches to efficiently treat and prevent CDI. One promising approach is the use of bacteriophage-based therapies, so using phages to treat infected patients. To exploit a phage for such purpose, it is necessary to be familiar with the biology of *C. difficile* and its associated phages (Hargreaves and Clokie, 2015). This introductory chapter provides an overview of *C. difficile* and the bacteriophage, focusing primarily on virulence factors that are behind the strain variation and reviewing *C. difficile* bacteriophages with its impact on biology of the host.

1.1.1 **Taxonomy of *Clostridium difficile***

Clostridium difficile belongs to the bacterial class Clostridia in the phylum Firmicutes, order Clostridiales, family Clostridiaceae and genus of *Clostridium* (He, 2012). The genus of *Clostridium* phylogenetically is highly diverse group of bacteria (Kalia et al., 2011, Collins et al., 1994). According to estimates based on chemical compounds (2-methylhopanoids) found in cyanobacterial membranes, the *Clostridium* group is thought to have evolved about 2.34 billion years ago (Sheridan et al., 2003, Hegarty et al., 2016), with *C. difficile* having evolved within the last 1.1–85 million years, and recognized as a pathogen only ~ 40 years ago (Bartlett et al., 1978, He et al., 2010). Figure 1-1 shows a phylogenetic tree that illustrates the evolutionary relationships of *C. difficile* to other *Clostridium* species (He, 2012). Recently, Yutin and Galperin proposed reordering *Clostridium* species into six new genera, change name *C. difficile* to '*PeptoClostridium difficile*' based on 16S rRNA and ribosomal protein sequences (Yutin and Galperin, 2013).

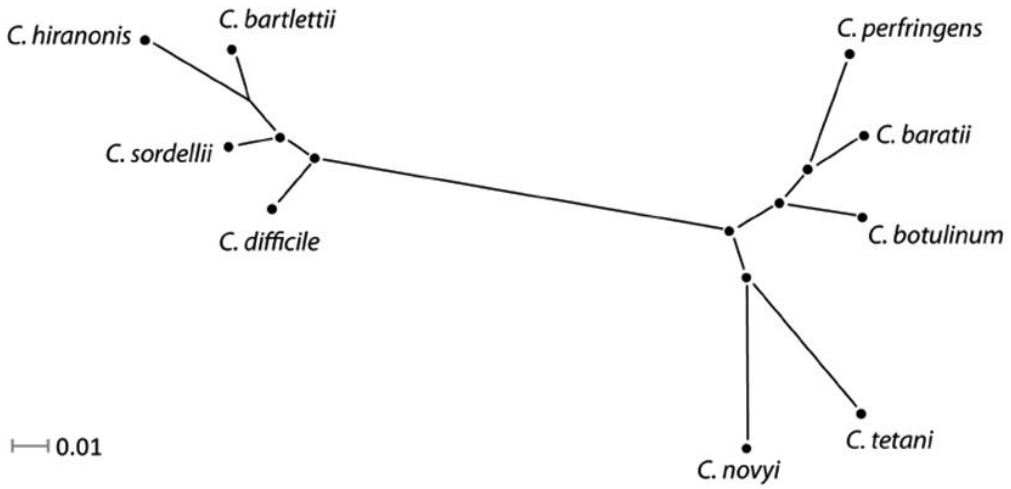


Figure 1-1. Phylogenetic tree of *C. difficile* and other *Clostridium* species.

The tree constructed on 16S rRNA repossessed from The Ribosomal Database Project (RDP) Project, reproduced with permission from (He, 2012).

1.1.2 Morphology

C. difficile is an anaerobic and endospore-forming bacterium. The vegetative cells of *C. difficile* are ~ 3-5 µm in length and ~ 1.5 to 2 µm wide; under the microscope the rod appears singly, in pairs or in short chains. In general, the organism is Gram-positive, but more mature colonies may exhibit Gram stain variability (Brazier and Borriello, 2000, Pechine et al., 2005, Aktories and Wilkins, 2000). This variability in Gram staining may be due the aged and dying cells becoming significantly thinner and more diffuse, and more friable to the Gram stain (Beveridge, 1990). The bacterium produces irregular, flat colonies ranging from 2 to 9 mm in diameter, with a lobate margin and a smooth, glossy, greyish-white appearance on BHI 7% blood medium (Hall and O'Toole, 1935, Hafiz and Oakley, 1976). The majority of *C. difficile* strains are motile with peritrichous flagella and multiple fimbriae (Hafiz and Oakley, 1976, Borriello, 1998). Under adverse growth conditions, *C. difficile* produces sub-terminal elongated spores that are slightly wider than the body of the cell (Hafiz and Oakley, 1976).

1.1.3 Potential reservoirs of *C. difficile*

1.1.3.1 Ecological (non-clinical) reservoir of *C. difficile*

Health-care facilities form the traditional reservoirs of *C. difficile* and consequently, the hospital environment has been extensively studied with respect to the diversity of *C. difficile* genotypes, and to their prevalence and persistence (Johnson et al., 1999,

Svenungsson et al., 2003, Barbut et al., 2007). However, *C. difficile* is also found in the environment outside hospitals, and has been isolated in a variety of locations, including drinking water, untreated water, river, lake, marine environments, sediment, soil samples, as well as air and air vents (al Saif and Brazier, 1996, Simango, 2006, Pirs et al., 2008, Zidaric et al., 2010, Higazi et al., 2011, Best et al., 2010, Pasquale et al., 2011, Hargreaves et al., 2013, Kotila et al., 2013, Diab et al., 2016). These environments offer additional reservoirs for the growth of this bacterium. Furthermore, the capacity for sporulation provides the bacteria with an outstanding means of protection to persist hostile environments, which enhances the potential for the spread of CDI. Researchers have shown that *C. difficile* spores remain stable in the environment on the surfaces, in food, soil and drinking water for several months (Martin et al., 2016).

Furthermore, *C. difficile* presence has been reported in numerous different food types, including ready-to-eat salad (Yamoudy et al., 2015, Bakri et al., 2009), cooked and uncooked meat (Songer et al., 2009, Harvey et al., 2011), fish, shellfish, edible bivalve molluscs and other sea food (Metcalf et al., 2011, Pasquale et al., 2012, Pasquale et al., 2011) and vegetables (Metcalf et al., 2010). Strain types of *C. difficile* found in food products were similar to those in clinical settings. This suggests entry of the bacterium into the food chain and as a consequence has raised questions about food-borne transmission of the pathogen (Gould and Limbago, 2010, Lund and Peck, 2015). In many CDI outbreaks, a contaminated environment such as (soil or irrigation water) is a route for delivery of the bacteria to the to humans via the ingestion of contaminated food (Lynch et al., 2009). Thus, humans and animals are frequently exposed to *C. difficile* spores from multiple environmental sources. Questions of unparalleled importance of the improved control of CDI are how, when, and how regularly this exposure leads to CDI (Gould and Limbago, 2010).

1.1.3.2 **Zoonotic reservoirs of *C. difficile***

In addition to humans, *C. difficile* also infects animals. The bacterium has been isolated from a range of domestic and wild animals from widely geographically dispersed sites. The majority of studies have attempted to confirm the hypothesis that humans acquire *C. difficile* from animals (O'Neill et al., 1993, Keel et al., 2007). Early studies generally emphasised the role of pets as a significant reservoir of *C. difficile* and associated with CDI (Borriello et al., 1983, O'Neill et al., 1993, Silva et al., 2013, Hensgens et al., 2012, Koene et al., 2012). The pathogen causes CDI in a range of

animal species including piglets, calves (Keel et al., 2007, Hensgens et al., 2012, Koene et al., 2012), beef cattle (Costa et al., 2012), chickens (Simango and Mwakurudza, 2008), ostriches (Frazier et al., 1993), rabbits (Perkins et al., 1995), and also has been reported in foals (Baverud et al., 2003, Hargreaves et al., 2013). The incidence of identical strains in various species suggests the possibility of interspecies transmission of *C. difficile* (Arroyo et al., 2005). A comprehensive study revealed a significant diversity of *C. difficile* strains in animals (Janezic et al., 2014). Strains were isolated from 13 animal species between 1998 and 2012. The four most prevalent PCR ribotypes were R078 (14.3% of isolates; 4 hosts), R014/020 (11.6%; 8 hosts); R002 (5.4%; 4 hosts) and R012 (5.4%; 5 hosts) (Figure 1-2). The last three ribotypes are also among the most common human PCR ribotypes, suggesting the extensive distribution of toxigenic *C. difficile* and the great overlap in strain dispersal between species raise concerns about interspecies, including zoonotic transmission (Rupnik and Songer, 2010b, Janezic et al., 2014). Uncertainties still remain with respect to the relationship of any such animal isolates to those connected to with *C. difficile*-associated disease in humans (CDAD) (Keel et al., 2007).

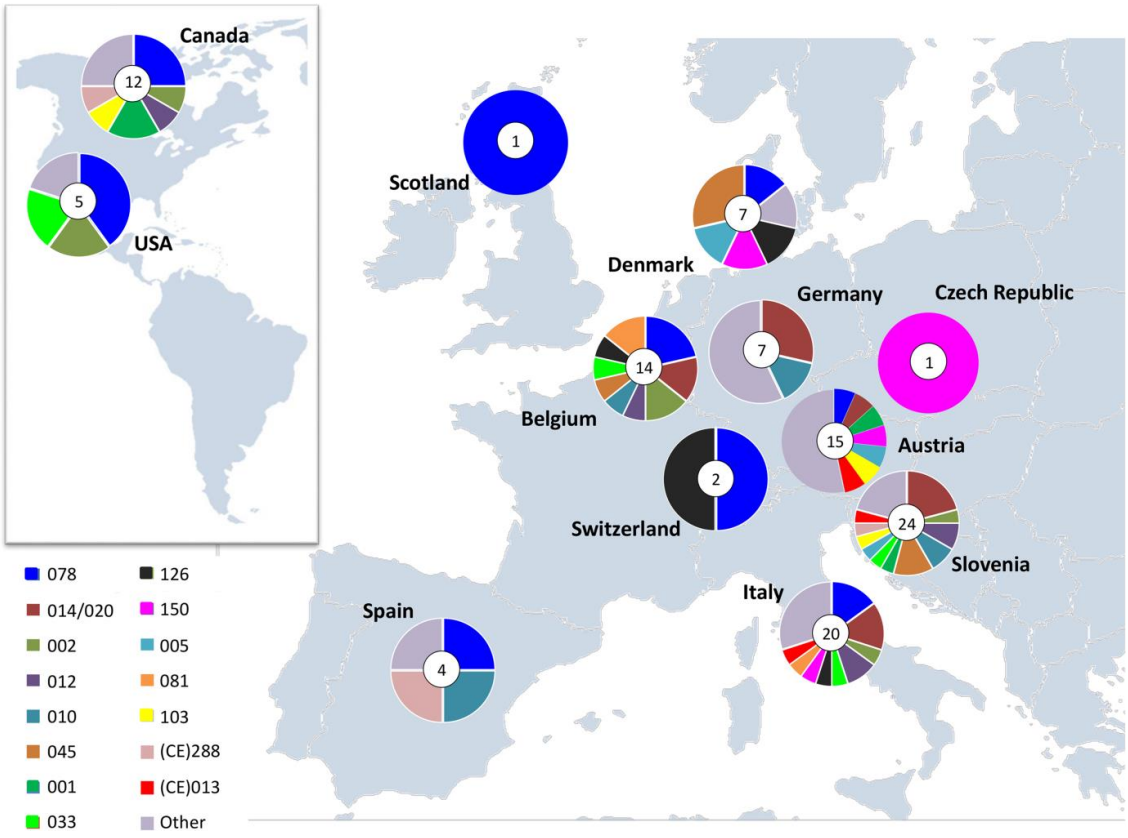


Figure 1-2. Geographical distribution of animal associated *C. difficile* PCR ribotypes from Europe, Canada and the USA.

Pie charts show the percentage of the most common prevalent PCR ribotypes in the collection for each country studied. The number in the middle of pie chart indicates the number of isolates from that country. In each country the diversity of *C. difficile* strains increase with the total number of isolates included in the collection and by strains from particular hosts as well (For example, rabbits in Italy and poultry in Slovenia) reproduced with permission from (Janezic et al., 2014).

1.1.4 **Pathogenesis**

C. difficile infection can be either endogenous or exogenous. Endogenous infection occurs from strains carried by the patient; exogenous is by far the most common and arises through the ingestion of *C. difficile* spores from infected individuals, contaminated health care-workers, nosocomial sources and contaminated environments (Vaishnavi, 2010, Diab et al., 2016). Patients receiving care within hospitals, nursing homes and out-patient surgical centres are often at risk of CDI due to their low immunity and use of antibiotic therapies (McMullen et al., Brown et al., 2015). Most hospitalised patients infected by *C. difficile* are asymptomatic carriers, who assist as

silent reservoirs for continued *C. difficile* contamination of the hospital environment (Vaishnavi, 2010). Approximately 25% of *C difficile* infections are first seen as symptoms among patients in hospitals, and 75% are first seen as symptoms among patients in nursing homes or in patients recently cared for in clinics. Fifty percent of hospital-diagnosed CDI is seen in patients who were transferred or recently discharged from other facilities (Voelker, 2012). Modern genome based techniques such as whole-genome sequencing (WGS) and multiple-locus variable-number tandem-repeat (MLVA) analysis are increasingly used to identify the sources from which *C. difficile* is acquired, thus probing undiagnosed symptomatic patients, symptomatic carriers (including infants), animal and foods (Martin et al., 2016).

CDI develops from the oral ingestion of the spores, via the faecal-oral route. These spores germinate into the vegetative form, and multiply within the small intestine. Flagella and fimbriae enable the vegetative cells of *C. difficile* to adhere to the mucus layer which surrounds the enterocytes, and then to penetrate the mucus layer. This forms the first stage of the pathogenic process, which is known as colonisation (Figure 1-3) (Denève et al., Kelly et al., 1994). This colonisation of the pathogen occurs after the diversity of the gut flora is disturbed due to the use of antibiotics (Tedesco, 1976, Owens et al., 2008, Goodarzi et al., 2012).

C. difficile can adhere to numerous cell lines which have been used as a model to understand this process, so Caco-2 cells and Vero cells are commonly used tools, and *C. difficile* can also adhere *in vivo* to caecal mucus of mice (Cerquetti et al., 2002, Gomez-Trevino et al., 1996, Karjalainen et al., 1994). The factors that are assumed to act as accessories to virulence and are involved in adherence and gut colonisation are proteolytic enzymes such as the cysteine protease Cwp84 (Poilane et al., 1998, Janoir et al., 2007), and adhesins implicated in mucus and cell association. Adhesins and other factors that contribute to the adhesion of *C. difficile* to the gut are discussed in section 1.1.6.1.

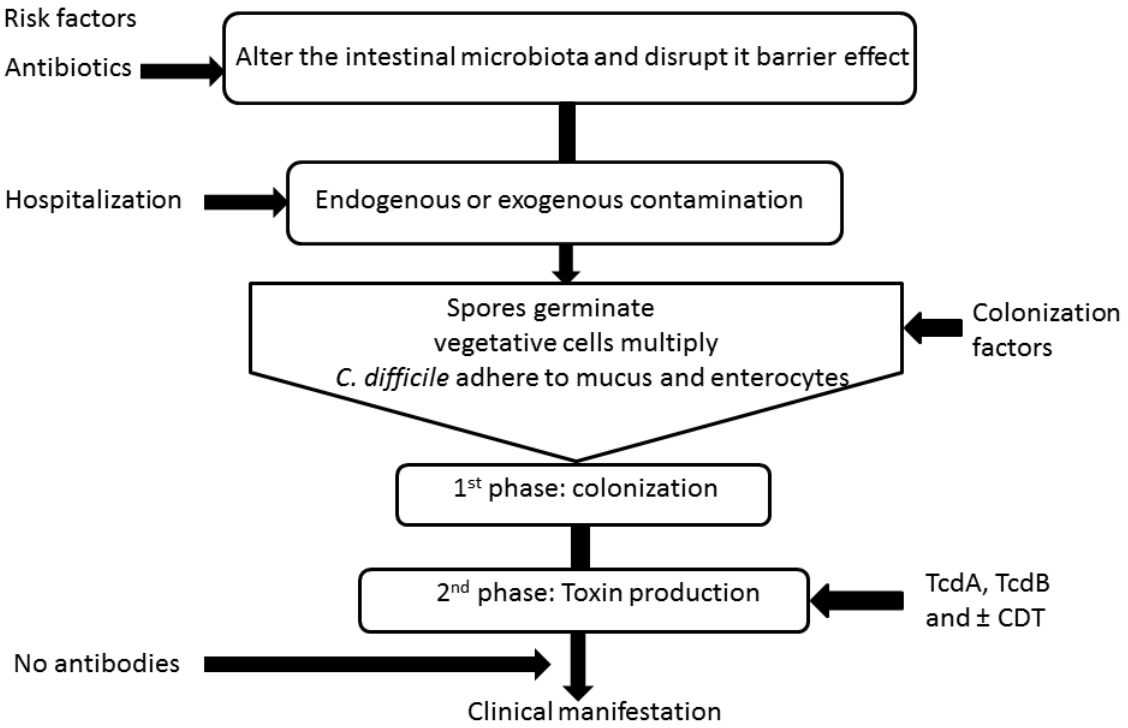


Figure 1-3. Pathogenesis of *Clostridium difficile* (Denève et al., 2009).

The second stage of the pathogenic process is toxin production. The combination of *C. difficile* toxins A and B is responsible for pseudomembranous colitis. ToxA is a potent cytotoxin which results in early damage to the epithelial lining of the intestine (Sutton et al., 2008, Kreimeyer et al., 2011), firstly by destroying the villus tips of the epithelium, and then by damaging the brush border membrane (Aktories and Wilkins, 2000). The resistance of the epithelial cells decreases, which leads to the opening of the tight junctions between epithelial cells. An increase in bacterial migration follows, and the morphological structures of the tight junctions and associated cells undergo changes (Rupnik et al., 2009, Pruitt and Lacy, 2012) (Figure 1-4). Furthermore, the production of neutrophils, monocytes, and necrotic enterocytes plays a significant role in the pathogenesis of CDI (Jose and Madan, 2016, Kelly et al., 1994, Lyerly et al., 1988, Pothoulakis et al., 1993, Barth et al., 2004). The gut mucosa becomes denuded, and penetration simultaneously occurs within protein-rich exudates such as mucin and fibrin.

Early experiments suggested that toxins A and B act synergistically, although later studies have demonstrated the significant role of ToxB in the pathogenesis of *C. difficile* colitis, and suggest that this may be more efficient than ToxA (Lyras et al., 2009). ToxB

is approximately one thousand times as potent as ToxA, and also shows higher enzymatic activity (Riegler et al., 1995, Poxton et al., 2001, Lyras et al., 2009). More recently, animal models have been used to show that ToxB is the main agent inducing the in vivo host-innate immune and inflammatory responses (Carter et al., 2015). Additionally, strains carrying $TcdA^{-}$ $TcdB^{+}$ profiles are periodically identified from clinical isolates; these have been beneficial in illuminating the role of TcdB in *C. difficile*-associated disease. Some of the virulence factors which differentiate various strains of *C. difficile* are discussed 1.1.6 (Kato et al., 1998, Kuijper et al., 2001, Sambol et al., 2000)

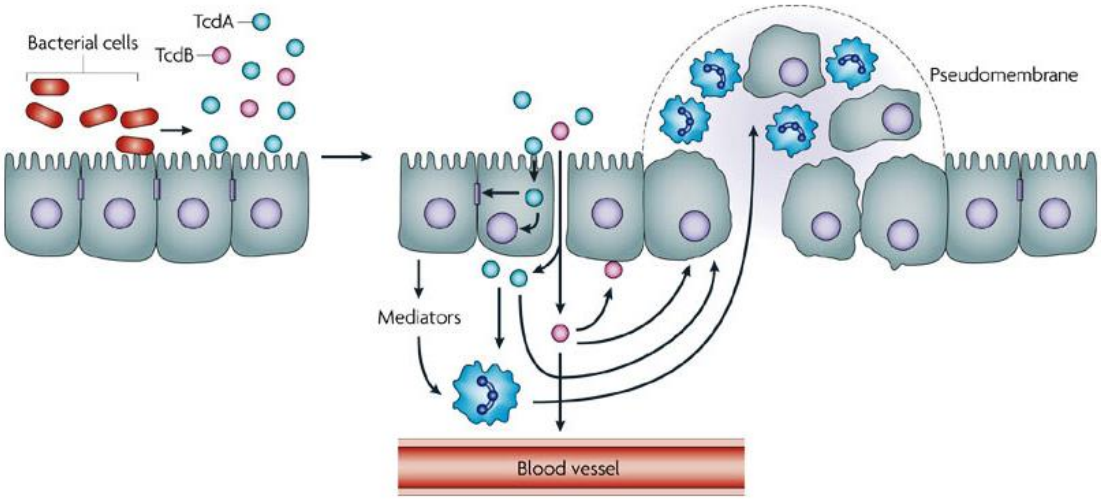


Figure 1-4. Pathogenesis of *Clostridium difficile*.

C. difficile colonizes the intestine (colon) after distortion of stable gut flora. The diagram shows bacterial cells as free cells and attached to host cells. Toxigenic strains produce toxin A and toxin B (TcdA and TcdB). TcdA attach to the apical side of the cell followed change in the cytoskeletal due to internalization of the toxin that result in disturbance of tight junctions and loosening of the epithelial barrier, in cell decease or in the production of inflammatory mediators that attract neutrophils. This damage in the tight junctions allows both TcdA and TcdB to cross the epithelium. TcdB binds specially to the basolateral cell membrane. Both toxins are cytotoxic and prompt the release of numerous immunomodulatory mediators from epithelial cells, phagocytes and mast cells, causing in inflammation and the accumulation of neutrophils. In an animal model, TcdB was shown to have a tropism for cardiac tissue, which would essential that TcdB enter the bloodstream reproduced with permission from (Rupnik et al., 2009).

1.1.5 **Virulence factors of *C.difficile***

1.1.5.1 ***C. difficile* toxin A and B**

The main virulence factors of *C. difficile* are two toxins with high molecular weight, toxin A (TcdA) and toxin B (TcdB). These are among the strongest bacterial toxins identified to date, and they are joined by *Clostridium sordellii* lethal toxin and hemorrhagic toxin and *Clostridium novyi* alpha toxin in making up the group of large Clostridial toxins (Voth and Ballard, 2005, Just and Gerhard, 2004). The exotoxins TcdA and TcdB have molecular mass 308 kDa and 270 kDa respectively (von Eichel-Streiber and Sauerborn, 1990, Barroso et al., 1990). Usually, TcdA is termed an enterotoxin, and TcdB is termed a cytotoxin (Taylor et al., 1981); however, TcdA does have cytotoxic activity, although this is 100 to 1000 times less potent than TcdB as described in section 1.1.4 (Donta et al., 1982, Sullivan et al., 1982). Both toxins are encoded and controlled by genes of the 19.6kb pathogenicity locus (*Paloc*) of *C. difficile* (Figure 1-5). The *tcd*A gene which encodes toxin A is found as a single open reading frame of the *PaLoc*-encoded 8,133-nucleotide. The *tcd*B gene which encodes toxin B is 7,098 nucleotides in length (Lyerly et al., 1988, Voth and Ballard, 2005). The expression of *tcdA* and *tcdB* is high in the late log phase in response to environmental conditions (Dupuy et al., 2008, Hammond and Johnson, 1995) (Figure 1-5).

The *tcdC* gene is present next to *tcdA* gene on the *PaLoc* and is transcribed differently from the toxin genes; the transcription is high during rapid exponential growth, and its expression then declines as cells reach a stationary phase (Figure 1-5) (Braun et al., 1996). The decline in *tcdC* expression is consistent with the increase in expression of *tcdA* and *tcdB*, indicating a negative regulatory function of *tcdC* for both toxin genes (Dupuy et al., 2008, Hundsberger et al., 1997).

tcdD is situated upstream of *tcd*B and also impact the expression of *tcd*A and *tcd*B (Figure 1-5). *tcdD* is comparable to DNA-binding proteins that have been shown to assist the expression of promoter/reporter fusions containing the promoter-binding regions of *tcdA* and *tcdB* (Karlsson et al., 2003, Moncrief et al., 1997). In addition, expression of *tcdD* is responsive to environmental factors such as the cellular growth phase, the growth temperature and the constituents of the medium. As a result, environmental conditions may initiate or inhibit the toxin genes through the regulation

of *tcdD* expression or activity (Matamouros et al., 2007, Mani et al., 2002, Karlsson et al., 2003).

The gene encoding TcdE is located between *tcdB* and *tcdA* in the pathogenic locus, and has homology with phage holin proteins; it is accordingly assumed to assist the release of *tcdA* and *tcdB* through permeabilization of the *C. difficile* cell wall (Figure 1-5) (Tan et al., 2001).Toxin expression therefore seems to be dependent on the declines in *tcdC*, *tcdD*-increased expression, and *tcdE*-mediated release from the cell (Voth and Ballard, 2005).

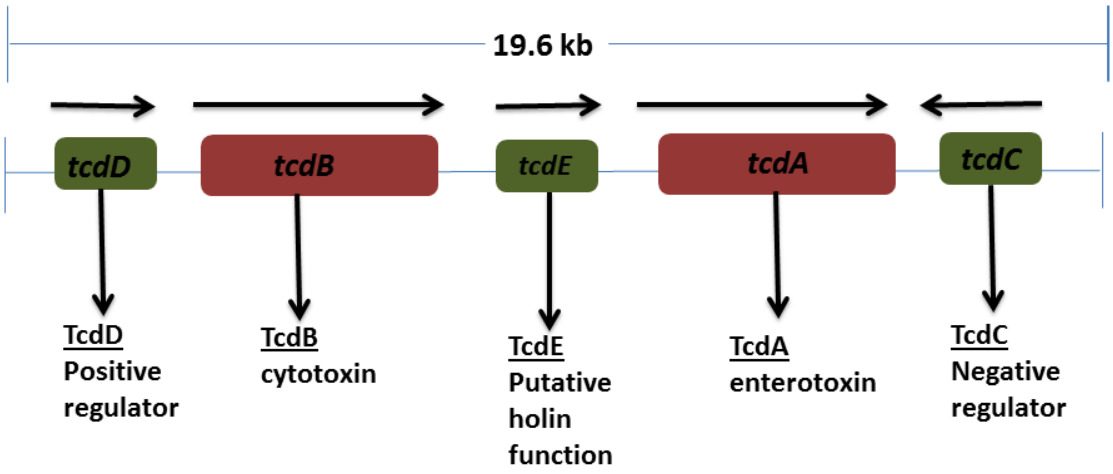


Figure 1-5. The pathogenic locus (Paloc) of *Clostridium difficile*.

The schematic of the 19.6 kb *C. difficile* PaLoc, showing *tcdA* and *tcdB* genes encoding the two major toxins (TcdA and TcdB respectively), the negative and positive regulatory genes (*tcdC* and *tcdD* respectively) and a putative holin-like protein is *tcdE* (Voth and Ballard, 2005).

1.1.5.2 **Binary toxin**

Some strains of *C.difficile* produce a binary toxin in addition to TcdA and TcdB, which is known as *Clostridium difficile* transferase (CDT) (Vedantam et al., 2012). It is a member of the family of Clostridia binary toxins, and has an actin-specific ADP-ribosyltransferase activity; other members of this toxin family are *Clostridium perfringens iota*-toxin, *Clostridium spiroforme* toxin and the *Clostridium botulinum* C2 toxins C & D (Barth et al., 2004, Gerding et al., 2014). It has a cytotoxic effect on

epithelial cells by causing damage to the structure of the actin cytoskeleton (Just et al., 1995). Studies have reported that, in addition to its effect on the cytoskeleton, binary toxin also causes a change in the surface of epithelial cells, allowing the increased adherence of bacteria (Schwan et al., 2009).

The *C. difficile* binary toxin is encoded by two genes, *cdtA* and *cdtB,* and these are located on the *C. difficile* chromosome at a locus 4.3kb outside the *Paloc* called *cdt* locus (Carter et al., 2007), with a positive transcriptional regulator named *cdtR* (Gerding et al., 2008). Both *cdtA* and *cdtB* genes are linked together and responsible for the production of the CDT subunits CDTa and CDTb (Perelle et al., 1997), which form the enzymatic component and binding component of CDT respectively. The other gene *cdtR* encodes the regulatory gene responsible for the positive regulation of the binary toxin, a function similar to *tcdD* in the *PaLoc* (Bouvet and Popoff, 2008, Perelle et al., 1997). Research has confirmed that this region is present in binary toxin positive strains such as (QCD-32 g58) and is absent in non-toxigenic strains such as (CD37), although it is present at about 4.2kb in binary toxin negative strains such as (CD630) (Stare et al., 2007).

The role of the binary toxin in virulence is still under debate, a group of researchers have shown that strains with only CDT gene carriage were not virulent in animal models (Geric et al., 2006). However, to date, it is hypothesized that presence of binary toxin is related to the severe outcome of CDI, thus binary toxin more likely enhance the toxicity of TcdA and TcdB and lead to more severe disease (Sabrina et al., 2011, Gerding et al., 2014). In addition the encoding gene of CDT is always found in some epidemic strains such as ribotype 027 and 078 (Gerding et al., 2014). Furthermore, several case of CDI have been reported in France that are most likely linked to $TcdA^{-}$ $TcdB^{-}$ CDT^{+} strains (Eckert et al., 2015), although it was found that the clinical manifestation of CDI due to those strains expressing only CDT were not typical of CDI, as discovered using the small intestine in animal model (Kuehne et al., 2014).

1.1.6 **Non- toxin virulence factors**

Since the increased rate of CDI patients infected with the new epidemic strains, roles of non-toxin virulence factors in *C. difficile* colonization, proliferation and maintenance in the host gut have been emphasised (Viswanathan et al., 2010). This can be attributed to the improved mechanism of persistence in the intestinal tract regardless

of taking antimicrobial agents in the newly emerged strains. This long term of setting is probably facilitated through the newly-developed colonization tactics (Vedantam et al., 2012). Therefore, it is important to outline these non-toxin virulence factors in *C. difficile* strains, and I outline key facts of these factors below.

1.1.6.1 **Adhesion**

C. difficile has the facility to adhere to the gut surface in the host, which is necessary for the full manifestation of virulence; so far numerous assumed adhesins have been recognized that have microbial surface components recognizing adhesive matrix molecules role (MSCRAMM) (Janoir, 2016). The first sign of *C. difficile* adherence to human gut was reported in 1979 by Borriello (Borriello, 1979), who detected the presence of *C. difficile* cells under the microscope in a washed biopsy sample from patients with PMC. In another study conducted on a model of the disease in hamsters, he showed that the highly virulent toxigenic strains adhered more strongly than strains with less virulence, and both strains adhered more strongly than non-toxigenic strains (Borriello, 1998). Marked adherence is shown in the terminal ileum and caecum, where the disease is more noticeable (Borriello, 1998). This research also revealed that the co-administration of non-toxigenic strains alongside *C. difficile* toxin A elevated the adhesin of the strain to the same rate as that of a toxigenic one. This suggests that adhesion is mediated by toxin A, and that this is the cause of destruction, or that toxin A is directly implicated in binding *C. difficile* to the gut, as set out in section 1.1.4. Various factors influence the ability of *C. difficile* to adhere to the gut, such as:

a) **Fimbriae:** Fimbriae are proteinaceous appendages present in bacteria, and were first identified in 1988. Many strains of *C. difficile* are motile and have fimbriae that are shown to be polar or peritrichous, 4–9 nm in diameter and 6 μm in length (Borriello et al., 1988). The flagella consist of the flagellin FliC and the flagellar cap protein FliD, and both of these have been shown to participate in the process of attachment to the intestinal mucus (Tasteyre et al., 2001). The vegetative cells of *C. difficile* attach to human intestinal epithelial cells (Caco-2 and HT-29MTXC) and to certain elements of the extracellular matrix such as fibrinogen, laminin, fibronectin, collagen I, III and IV (Karjalainen et al., 1994, Cerquetti et al., 2002, Eveillard et al., 1993). *C. difficile* strains also express peritrichous flagella, although flagellar expression and its accompanying

motility vary between *C. difficile* strains, and are regulated by environmental stimuli (Pituch et al., 2002, Kim and McCarter, 2000).

b) **Surface layer proteins**: The S- layer proteins (SLP) within the vegetative cells of *C.difficile* are composed of two main cell surface proteins, high- MW and low-MW proteins which display interstrain variability (Cerquetti et al., 2000, Calabi et al., 2001). Both are encoded by slpA and are formed from the posttranslational cleavage of a common precursor (Calabi et al., 2001). These proteins may have a role in colonization, potentially acting as adhesins assisting the interaction between *C. difficile* cells and gut cells (Cerquetti et al., 2000, Gebhart et al., 2012). The first S-layer protein to be recognised in *C. difficile* and to be identified as having role in adhesion was the Cwp66 protein of 66 kDa (Waligora et al., 2001, Janoir, 2016). The Cwp66 contains two domains, each carrying three imperfect repeats and one showing a correlation with the autolysin CwlB in *Bacillus subtilis* (Waligora et al., 2001). S-layer proteins of *C. difficile* are thus extremely variable, and may contribute to the virulence of the organism by assisting adherence and colonisation or by prompting inflammatory and immunological response from the host (McCoubrey and Poxton, 2001, Poxton et al., 2001, Fagan et al., 2011, Ryan et al., 2011).

c) **Physicochemical properties:** The physicochemical properties of a bacterium can enhance its capability of adhering to surfaces. *C. difficile* cells are hydrophilic and carry a net positive charge which is evenly dispersed over the cell wall (Krishna et al., 1996). Charged interactions with the negatively charged human gut can therefore lead to colonisation (Waligora et al., 1999).

1.1.6.2 *C. difficile* spores

C. difficile produces dormant spores as described in section1.1.3, and these can remain stable in the environment or the gut (Akerlund et al., 2008). Due to the bacterium's strict anaerobic requirement, spores play an important role in the dissemination, perseverance and pathogenesis of *C. difficile* (Jump et al., 2007). It has been shown that *C. difficile* spores can tolerate very adverse environments, heating and disinfection (Akerlund et al., 2008, Lawley et al., 2009). Spores initiate germination in reaction to particular bile acids in the small intestine after being ingested by susceptible hosts, and its vegetative forms then resume and cause infection after secretion of the

toxins. The capability of some *C. difficile* strains to produce very great numbers of spores in the host might therefore be responsible for the variation in the ability of *C. difficile* to spread-easily with in the environment (Viswanathan et al., 2010). Several studies have demonstrated that strains of particular ribotypes of *C. difficile* have the ability to yield larger amount of spore than other types; this characteristic can account for the greater success of dissemination, severity of infection and recurrences of some types in comparison with others (Akerlund et al., 2008, Wilcox and Fawley, 2000, Merrigan et al., 2010, Vohra and Poxton, 2011). However, others research has shown that although there is significant variation in sporulation between the strains of *C. difficile*, it is not type-associated (Burns et al., 2010, Burns and Minton, 2011). In considering the roles of spores in the transmission of CDI, it is necessary to understand the differences in sporulation between strains (Viswanathan et al., 2010). In this project we are interested to know the variation in sporulation rate between strains from environment and hospital to signify role of each strains in the pathogenicity of CDI.

1.1.6.3 **Hydrolytic enzymes**

Several species of anaerobic bacteria produce a comprehensive hydrolytic enzyme, and the role of these enzymes within disease processes is well-established (Steffen and Hentges, 1981). These enzymes break down carbohydrate, protein, and fat particles into their basic forms. The hydrolysis of polymers by hydrolytic enzymes outcome free monomers. A comprehensive study was conducted by Seddon and colleagues to examine whether hydrolytic enzymes are contribute in breakdown of connective tissue; they obtained results showing that all of the *C. difficile* strains had hyaluronidase, chondroitin-4-sulphatase and heparinase action, although the heparinase activity was usually weak (Seddon et al., 1990). Strains with high virulence were more energetic than less virulent strains. Collagenase activity was also existent, but this was usually weak and limited to the more highly virulent strains. It is likely that some of these tissue-degrading enzymes contribute to the noticed pathology and help in further compromising gut integrity and causing consequent fluid accumulation. It is also possible that *C. difficile* develops nutritional advantage from such activity (Seddon et al., 1990)

1.1.7 **Methods for determining the strain relatedness**

C.difficile strains are genetically diverse (He et al., 2010, Knight et al., 2015) and some *C. difficile* strains are more problematic in causing disease or infection than others. Several genome typing methods have been developed to examine the epidemiology, genetic diversity, and evolution of *C. difficile*, these include:

1.1.7.1 **Ribotyping**

Ribotyping is a method that distinguishes between strains according to the size and the number of copies of the *C. difficile* 16S - 23S rRNA inter spacer region in different alleles (O'Neill et al., 1996). It was first applied to *C. difficile* by Gürtler in 1993 (Gurtler, 1993), and this research identified various banding patterns in 24 strains of *C. difficile* by amplifying these regions of DNA and running them on long denaturing polyacrylamide gels, the study was then able to sub-type the 24 strains into 14 ribotypes. Ribotyping is a straightforward and reproducible method which enables the study of a large number of isolates (Cartwright et al., 1995). The method was simplified by Cartwright and colleagues in 1995, and these researchers carried out the process on 102 strains gained from symptomatic patients. Through the use of Gürtler primer, he was able to distinguish between closely related strains using agraose gel electrophoresis instead of denaturing polyacrylamide gel electrophoresis (PAGE) (Cartwright et al., 1995). This technique was modified for routine use by O'Neill et al. (1996), who developed it further into a rapid and simple typing method. Traditionally, the amplified DNA with re-designed primers gives bands between 260 and 585 bp, and high resolution agarose was required for greater discrimination. By 1996, 116 ribotypes had been defined (Stubbs et al., 1999). Bidet and colleagues have improved the ribotyping primers and have to obtained bands that are more stable and readable (Bidet et al., 1999, Bidet et al., 2000). Ribotyping is now the most-widely used typing method in Europe and Australia and its use has now increased in North America (Arroyo et al., 2005, Janezic and Rupnik, 2010, Kuijper et al., 2009, Sadeghifard et al., 2006, Walk et al., 2012, Martin et al., 2016, Brazier, 2001). More recently, capillary gel electrophoresis-based PCR ribotyping has been introduced, as an alternative approach to solving the problem of inter-laboratory variations (Indra et al., 2008).

1.1.7.2 **Toxinotyping**

Toxinotyping is restriction fragment length polymorphism (RFLP)-PCR-based method for categorising *C. difficile* strains into groups according to variation within the

toxin genes, *tcdA* and *tcdB* (Rupnik et al., 1997, Rupnik, 2010). These variants are manifest in the first 3 kb of *tcdB* (PCR fragment B1) and 3 kb of the repetitive region of *tcdA* (PCR fragment A3). The principle of toxinotyping is based on restricting these DNA regions with enzymes like *Eco*RI, *Acc*I and *Hin*cII (Rupnik et al., 1998). The method corresponds to PFGE, serogrouping and ribotyping, even though several ribotypes may grouped within a single toxinotype (Rupnik, 2001). So far 34 toxinotypes have been identified (Rupnik and Janezic, 2016).

1.1.7.3 **Pulsed-field gel electrophoresis**

Pulsed-field gel electrophoresis (PFGE) includes the resolution of large fragments of DNA after restriction using infrequent cutting enzymes such as *Sma*I and *Sac*II (Brazier, 2001). PFGE typing of *C. difficile* corresponds well with sero-grouping (Kato et al., 1994), and can be even more discriminatory than Ribotyping (Bidet et al., 1999) this is considered to be the gold standard for the typing of pathogenic bacteria. By use of the PFGE optimized by Gal and his colleagues (Gal et al., 2005), seven sub-types of PCR ribotype 001 isolates were have been identified. The method is widely used in North America and the strains are designated North American pulsed-field (NAP) types (Tenover et al., 2011, Janezic and Rupnik, 2010).

1.1.7.4 **Arbitrarily-primed polymerase chain reaction**

Arbitrarily-primed polymerase chain reaction (AP-PCR) is based on the amplification of non-specific fragments of DNA through using short primers under low-stringency conditions (Killgore and Kato, 1994, Wilks and Tabaqchali, 1994, Wullt et al., 2003). It is not as reproducible as Ribotyping, but is more discriminatory than immune-blotting.

1.1.7.5 **Multi-locus sequence typing and Multi-locus variable number tandem-repeat analysis**

Multi-locus sequence typing (MLST) was developed to overcome difficulties in intra-laboratory comparisons of typing results, and is used to outline epidemiological associations between *C. difficile* isolates (Lemee et al., 2004, Maiden et al., 1998). The MLST method is discussed more in Chapter 3 of this thesis. Multi-locus variable number tandem-repeat analysis (MLVA) has also been shown to be a convenient method for *C. difficile* typing (Marsh et al., 2006); this technique uses the comparison of seven variable regions in the genome which categorise the strains into groups

correlating well to Restriction Endonuclease Analysis (REA) types. MLST and MLVA used in combination were shown to be positive in regard to the recognition of genetic lineages and the genetic relationships between them (Marsh et al., 2010).

1.1.7.6 **Surface-layer protein A-encoding gene (slpA) typing**

The diversity in the surface-layer protein A gene region between *C. difficile* strains has been found to be effectively discriminatory for use in typing schemes (Karjalainen et al., 2002, Huber et al., 2013). SlpA sequence typing uses PCR amplification of the variable region of *slpA* gene, followed by sequencing or restriction fragment length polymorphism (RFLP). It is a valuable genotyping tool, and has been found to be applicable within epidemiology studies and outbreak detection (Keim et al., 2000, Noller et al., 2003).

1.1.7.7 **Whole genome sequencing**

Whole genome sequencing (WGS) is the definitive pathogen typing tool, and can differentiate between strains at the single nucleotide level by comparing genomes in terms of single nucleotide polymorphisms (SNP). It therefore offers significant improvements in discriminatory power over conventional genetic typing methods (Didelot et al., 2012, Koser et al., 2012, Knetsch et al., 2013). WGS has shown to be advantageous in surveillance studies, outbreak identification and the study of strain transmission from the environment to patients (Knetsch et al., 2013) . It can produce practical, clinically-relevant data within a time frame which allows the improvement of hospital infection control and patient outcomes in routine clinical practice (Eyre et al., 2012). Through the use of WGS, our understanding of the genetic diversity, evolution, pathogenicity and epidemiology of this enigmatic pathogen has been considerably deepened (Knight et al., 2015).

1.1.7.8 **The *C. difficile* genome**

The first complete genome sequence of *C. difficile* strain 630 was performed by Sebaihia and colleagues in 2006 (Sebaihia et al., 2006). This multiple drug-resistant and highly transmissible strain of *C. difficile* was isolated in a patient with severe PMC in Zurich, Switzerland. The genome analysis of strain 630 displayed a large circular chromosome of 4,290,252 bp (4.3 Mb), 3,776 putative protein-coding sequences (CDSs), with a GC content of 29.06% (Knight et al., 2015). Subsequently, the genome of several *C. difficile* strains ranging in size from 4.1 to 4.3 Mbp were fully sequenced

and annotated: CD37 (R009; isolated in the United States in 1980) (Brouwer et al., 2012), M68 (R017; isolated in Ireland in 2006), CF5 (R017; isolated in Belgium in 1995), M120 (RT078; isolated in the United Kingdom in 2007), G46 (R027; isolated in the United Kingdom in 2006) (Gaulton et al., 2015), R20291 (R027; isolated in the United Kingdom in 2006), CD196 (R027; isolated in France in 1985), 2007855 (R027; isolated in the United States in 2007), and BI1 (R027; isolated in the United States in 1988) (He et al., 2010, Monot et al., 2011, Stabler et al., 2006, Knight et al., 2015).

C. difficile reference genomes, such as those of strains 630 and M120, play a significant role in the NGS data analysis pipeline. They comprise a definite and contiguous sequence of well-known nucleotides which cover the entire chromosome and plasmids (if present), therefore providing an extremely high-quality reference for the mapping of draft genomes (Knight et al., 2015). Whole genome analysis of the *C. difficile* strains listed above has revealed a considerable amount of information regarding the architecture of the *C. difficile* genome, including that it has a highly dynamic and mosaic genome which involves a high proportion (11% in strain 630) of mobile genetic elements (MGE). These include transposable elements, prophages, and clustered regularly interspersed short palindromic repeat (CRISPR)-*cas* elements (Brouwer et al., 2012, Cairns et al., 2012, Darling et al., 2014, Hargreaves et al., 2015, He, 2012, Monot et al., 2011, Mullany et al., 2015, Stabler et al., 2009). *C. difficile* possesses an "open" genome with high levels of plasticity, and has access to and frequent transactions with numerous host environments and bacterial gene pools (Knight et al., 2015). Bacteriophage is one of the factors which shape this genome plasticity through gene flow. Reports on prevalence of prophage and CRISPR –cas system in *C. difficile* strains are discussed in chapter 4 of this thesis.

1.1.7.9 **Epidemiology of CDI**

C. difficile was first recognised as a pathogen linked to antibiotic- associated diarrhoea (AAD) in the late 1970s (Martin et al., 2016). During the last decade, the epidemiology of CDI has changed markedly. An increase in the prevalence and severity of CDI has been reported in Canada, USA and Europe (Davies et al., 2015, Dubberke and Wertheimer, 2009, McDonald et al., 2006, Pepin et al., 2005, Johnson, 2014, Freeman et al., 2010, Lessa et al., 2012, Magill et al., 2014), not only within health-care settings but also within communities (Martin et al., 2016). This rapid transformation is due in part to the emergence of the high-virulence strain PCR ribotype 027, which is

sometimes referred to as North American Pulsed-Field type 1 or group B1 (Lessa et al., 2012, Marra and Ng, 2015, Freeman et al., 2010). The first outbreak in the UK of R027 occurred between 2003 and 2004 at the Stoke Mandeville Hospital (Buckinghamshire Hospitals NHS Trust), and resulted in 174 infections and 19 deaths, a mortality rate of 11% (Smith, 2005). Subsequently, CDI due to ribotype 027 was reported in numerous European countries (Kuijper et al., 2006, Barbut et al., 2007, Kuijper et al., 2008). It is now an endemic strain throughout the US, Canada and Europe (Clements et al., 2010, He et al., 2013, Freeman et al., 2010). CDI cases due to infection with R027 strains are recognised as involving more severe diarrhoea and higher mortality, with increased risk of recurrence (Goorhuis et al., 2007, Loo et al., 2005, Hubert et al., 2007).

The increasing rates of CDI are not caused solely by the spread of R027 strains. Increased CDI is also reported due to strains of *C. difficile* ribotypes 001, 106 and 078 (Rupnik et al., 2009). Recently, an increase in the prevalence of *C. difficile* R078 was observed in several countries in Europe, from 3% in 2005 to 13% in 2008 (Goorhuis et al., 2008b, Rupnik et al., 2008). The severity of CDI due to types 027 and 078 is similar, although the latter affects a younger population and is more usually found in community associated disease (Goorhuis et al., 2008b). Today, R078 is acknowledged to be the third most common ribotype in Europe, with the first and second most common being 014 and 001(Bauer et al., 2011). Several clarifications for this 'hypervirulence' were suggested, and it seems possibly pathogenic factors such as germination, sporulation, epithelial adherence and toxin production could be the factors behind the success of some strains (Martin et al., 2016).

1.1.8 **Risk Factors**

To reduce *C. difficile* infection rates requires an understanding and reduction of CDI risk factors (Bignardi, 1998, Goudarzi et al., 2013). The most significant primary risk factors are an age of over 65 years, an age of less than 1 year with comorbidity or underlying conditions, hospitalisation and antimicrobial therapy (Goudarzi et al., 2014, McFarland et al., 1999, Smits et al., 2016). The major secondary risk factors include underlying conditions; existing problems or intestinal disease such as inflammatory bowel disease or colon cancer; a weak immune system arising from chemotherapy, other drugs which suppress the immune system or AIDS; malnutrition; low serum albumin level (<2.5 g/dL); neoplastic diseases; cystic fibrosis; diabetes; and malignancy

(Vecchio and Zacur, 2012, Bignardi, 1998, Nicholson et al., 2015, Goudarzi et al., 2014, Klingler et al., 2000, Gupta and Khanna, 2014, Khanna and Pardi, 2012). The administration of broad-spectrum antibiotics, which are destructive to the balance of the gut flora and promote the propagation of toxigenic *C. difficile*, is the most extensively recognized risk factor (Owens et al., 2008, Goudarzi et al., 2014, Goudarzi et al., 2013, Vecchio and Zacur, 2012, Smits et al., 2016). One current study has established that inpatients taking antibiotics are, on average, 60% more likely to acquire the infection (Brown et al., 2015). In addition, proton pump inhibitors may also disturb the microbiota of the gut and the large intestine, thereby possibly increasing the risk to the patient of CDI (Williams, 2001, Gupta and Khanna, 2014). This is largely under control now and their use has been minimised (McDonald et al., 2015)

1.1.8.1 **Antibiotic treatment of CDI**

Traditionally, CDI has been connected with antibiotic-associated pseudomembranous colitis. Clindamycin, third-generation cephalosporins and penicillin have been identified as posing the highest risk; more recently fluoroquinolones have also begun to be considered high-risk agents (Owens et al., 2008, McDonald et al., 2005). Frequently, the cessation of the antimicrobial agent causing CDI leads to termination of the infection; however if this fails, standard antibiotic treatment for over 30 years has involved the administration of either vancomycin or oral metronidazole (Norén, 2010, Pepin, 2008, Aslam et al., 2005). Vancomycin is the treatment of choice for severe or problematical CDI, with or without adjuvant therapies and metronidazole is applicable in mild infection (Bagdasarian et al., 2015). However, failure rates of treatments with metronidazole and vancomycin of 22.4% and 14.2% respectively were discovered in a recent review by Vardakas and colleagues (Vardakas et al., 2012). Due to the high rates of treatment failure, CDI can relapse; a reoccurrence rate of 27.1% was noted among patients treated with metronidazole, and lower levels at 24% was observed for vancomycin. Further research has also reported the emergence of strains that have lower susceptibility to metronidazole and vancomycin (Tickler et al., 2014, Freeman et al., 2007).

The new antibiotic Fidaxomicin (Optimer Pharmaceuticals, USA) permitted in 2011 by the US Food and Administration (FDA) for treatment of CDI, and this has a similar effect to that of vancomycin. This antibiotic has considerably decreased the level of reoccurrence, which was shown to be 13.3% during clinical trials. However, its

clinical utility is limited, due to the high cost associated with this antibiotic in comparison to metronidazole and vancomycin (Ivarsson et al., 2015). Typical 10-day treatments with metronidazole, vancomycin and fidaxomicin are estimated to cost in about $22, $680 and $2800 respectively (Surawicz et al., 2013). Vancomycin is almost thirty times as costly as metronidazole, and metronidazole is consequently recommended for the treatment of mild infections, with vancomycin being administered only if this fails (Ivarsson, 2014). The reason for this is the awareness of the possibility of emerging strains of vancomycin-resistant *C. difficile* and vancomycin-resistant enterococci; this antibiotic has the potential to kill the majority of intestinal microflora, and this would therefore allow the emergence of a vancomycin-resistant enterococci population (Ananthakrishnan, 2011). Vancomycin is therefore only used when absolutely necessary (Ramesh et al., 1999). In view of the incomplete effectiveness and high rates of relapse developing with metronidazole and vancomycin, it is clear that there is a significant requirement for novel potential therapies for treating CDI. Phage therapy offers one potential approach towards combating CDI.

1.2 Bacteriophages

Bacteriophages (phages) are viruses that infect and kill bacteria, generally with a high degree of specificity. Phages are obligate intracellular parasites that replicate inside bacteria by making use of some or all of the host biosynthetic machinery (Parija, 2014). They are widely distributed in nature, and are closely associated with bacteria. Phages can be readily isolated from sea water, soil, faeces, sewage and from most other places where their hosts are found. Consequently, phages are considered to be the most abundant life form on Earth; it has been estimated that 10^{31} phages exist, although very few have been studied in detail (Hendrix, 2003).

Bacteriophages were discovered independently in 1915 by William Twort in England, and in 1917 by Felix d'Herelle at the Pasteur Institute in France, and it was observed that they had the potential to kill bacteria (Clokie et al., 2011). Research exploring therapeutic application of bacteriophages has been conducted since their discovery at the beginning of the 20^{th} century (Merril et al., 2003). Studies in the development of Phage therapy study have been carried out in Eastern Europe since 1923, particularly at the Eliava Institute of Bacteriophage, Microbiology and Virology, Georgia (Merril et al., 2003).

1.2.1 **Morphology and classification bacteriophages**

The typical features of a bacteriophage, as visualised by Transmission Electron Microscope (TEM) consist of a head or capsid in the shape of an icosahedron, which encloses the viral genome (commonly double stranded DNA) as shown in (Figure 1-6), a tail consisting of a hollow tube through which the genetic material passes during infection, and which has a tendency to contract, and tail fibres which bind to the surface of their target bacteria (Ackermann, 2006, Hanlon, 2007, Ackermann and Krisch, 2014, Merril et al., 2003, Parija, 2014).

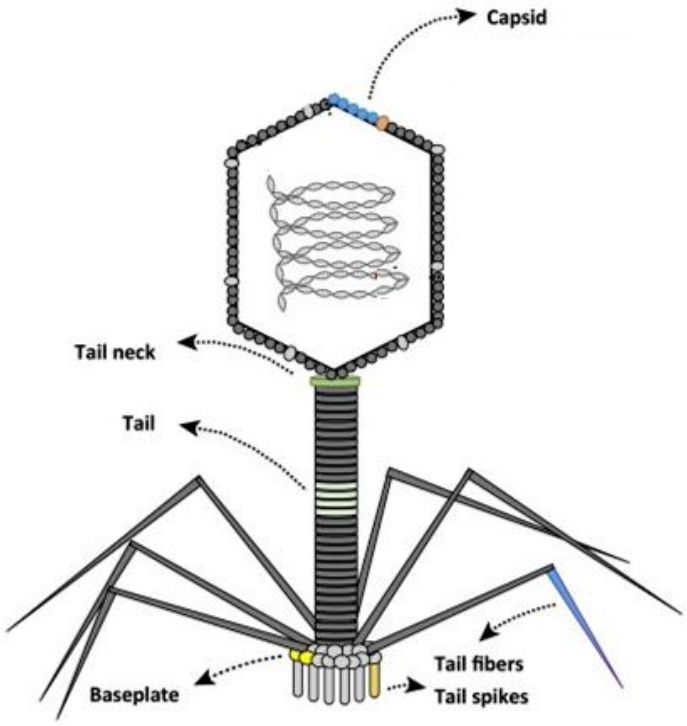


Figure 1-6. A schematic representation of a typical feature of a common phage from the order Caudovirales reproduced with permission from (Nobrega et al., 2015).

Phages are classified into 13 families based on their morphology, the type of genetic material and their mode of replication (Ackermann, 2007) (Figure 1-7). The morphology of phages may occur in nature as tailed or polyhedral, filamentous, or pleomorphic. Most of the phages have double-stranded dsDNA, but some phages also contain single-stranded ssDNA, dsRNA and ssRNA (Ackermann, 2006, Ackermann, 2003).

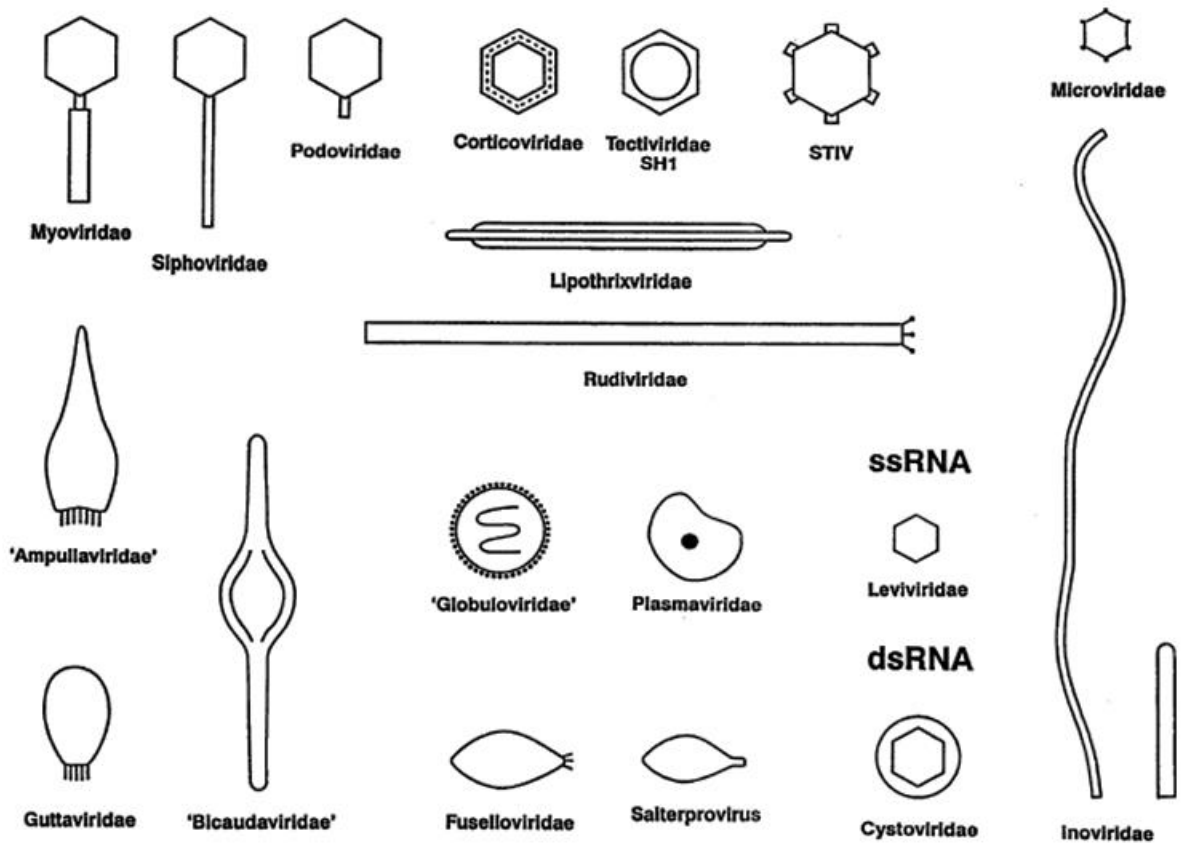


Figure 1-7 . Morphptypes of bacteriophages (Ackermann, 2007).

Of the 5568 phages that have been examined by electron microscope since 1959, 5360 (96.2%) are tailed. A tailed bacteriophage has a tail and an icosahedral head with double-stranded DNA. They are classified into the order Caudovirales, which has three large phylogenetically related families (Ackermann, 2003): Myoviridae with a contractile tail consisting of a sheath and a central tube, which comprise 25% of the tailed phages; Siphoviridae, with a long, flexible (non-contractile) tail, which comprise 61%; and Podoviridae, with a very short and non-contractile tail, which make up 14% (Ackermann, 2006, Ackermann, 2007) (Figure 1-7).

1.2.2 **Bacteriophage life cycles**

In general, phages can be classified into two classes based on their cycles: lytic phages, which are virulent or lysogenic (Sulakvelidze et al., 2001). Lytic phage undergoes a lytic life cycle, with the phage attaching to the host and injecting its genetic material, which causes disruption in the metabolism of the bacterial cell. The injected

nucleic acid exploits all resources of the host towards the production of new phages. Within minutes after injection, the host cells are lysed, releasing many new phage virions (Burrowes et al., 2011). Lysogenic phages are referred to as temperate phages. In the same way as lytic phages, they attach and inject nucleic acid into the host cell; this either initiates cell lysis or integration into the genome of their host to form prophages. Prophages are phages that have become integrated into the genome of bacterial hosts and are replicated with the bacterial genome. The new generation of bacterial cells acquire the phage genome within their chromosome (Merril et al., 2003, Sulakvelidze et al., 2001, Burrowes et al., 2011). Prophages can remain repressed and stable for thousands of generations (Clokie et al., 2011).

In either case, several steps happen before lysis of the host (Campbell, 2003) (Figure 1-8). The first stage in phage replication within a host cell begins with adsorption, in which the phage adheres to the surface of the host by means of its tail fibres (Figure 1-8, stage 1). The adsorption phase is a key stage which is defined by the phage-host specificity and the mechanism behind bacteria's resistance to phage infection (Rakhuba et al., 2010). The adsorption rate and efficiency are significant elements and rely on the physiological state of the host (Kutter and Sulakvelidze, 2004). Adhesion is increased by the ability of the phage's adhesions to identify specific binding sites on the host that are known as receptors and which determine the range of possible hosts (Hyman and Abedon, 2010). Receptors can either be protein, glycoprotein, phospholipids or lipopolysaccharide (Rakhuba et al., 2010).

Following successful attachment, the phage injects its DNA into the host (Figure 1-8, stage 2). This mechanism is called penetration, and varies from one phage to another. In myovirus, which has a contractile tail, the mechanism is complicated. The tail sheath contracts, and the DNA is injected into the bacterium, leaving the empty phage, known as the 'ghost', on the surface of the host cell (Cann, 2005, Parija, 2014). After the DNA injection, the processes of transcription and translation begin immediately (Figure 1-8, Stage 3A). This is commonly seen in lytic phages; in this case, the phages take over the replicatory mechanisms of their host for their own replication. This causes the production of large amounts of both nucleic acids and the structural and scaffolding proteins necessary for assembly, lysis and release. After the production of enough of this material, the phage particles begin to be assembled and the nucleic acid is packaged into the capsid (Figure 1-8, Stage 4A) (Nicklin, 1999). This is followed by

lysis of the bacterial cell (Figure 1-8, stage 5A), and then the release of phage particles (Figure 1-8, stage 6A).

In the case of lysogenic phages, the phage DNA does not necessarily replicate immediately, but instead is integrated into the chromosome of the host and replicates within the host chromosome (Figure 1-8, stage 3B). This integrated prophage DNA then replicates alongside the bacterial genome, and the next generation of bacterial cells inherit the phage DNA (Figure 1-8, stage 4B). However, the prophages can be induced to begin a lytic cycle, and to replicate and release phage progeny (Figure 1-8, stage 5B). This prophage induction may be spontaneous or induced using stress, UV radiation or antibiotics (Lydersen and Pollard, 1975, Giacomoni, 1982, Choi et al., 2010). The release is usually assisted by the phage enzyme holin, which lyses the cell wall of the host to release the phages (Wang et al., 2000, Krupovic and Bamford, 2008).

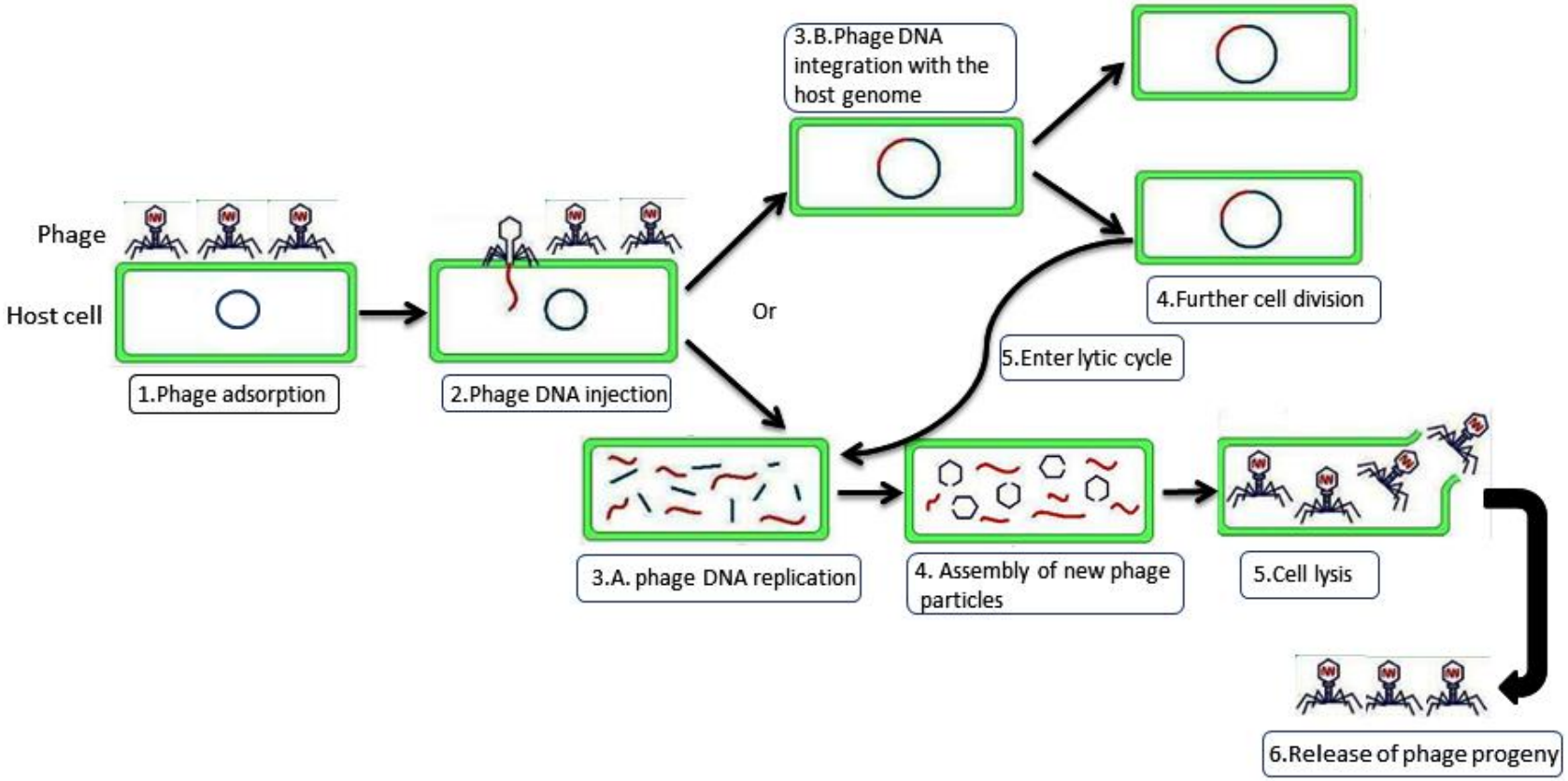


Figure 1-8. Diagram showing bacteriophage life cycle.

Phages can either follow the lytic cycle, which leads to lysis of the bacterial cell or the lysogenic cycle where phage DNA is incorporated within the bacterial genome.

1.2.3 *C.difficile* phages

C. difficile phages were first isolated in 1980s and used as a typing tool, but their efficacy for typing was limited due to their narrow host ranges (Dei, 1989, Sell et al., 1983). The rapid increase in *C. difficile* prevalence and severity over the last few decades has promoted a renewed the interest in the use of phages, this time in their potential use of phage therapy to challenge this pathogen (Hargreaves and Clokie, 2014). Aside from the aforementioned early research on *C. difficile* phages and research on phages that target this bacterium is limited. This is in part due to difficulties in the isolation and propagation of these phages (Hargreaves and Clokie, 2014). Small numbers of temperate phages have been identified following antibiotic induction, and the majority of these phages are not amenable to propagation or had very low frequency of propagation (Mahony et al., 1985, Mayer et al., 2008, Nale et al., 2012, Sell et al., 1983). Attempts to isolate free phages from patients (Nale et al., 2012), animal and environment samples (Goh et al., 2005b) were reported to have failed despite the use of multiple methods to isolate phages, and the use of multiple hosts.

The isolation of lytic phages has become of increasing interest to researchers of late, due to their potential use in therapeutic applications (Adhya and Merril, 2006). To date, no lytic phages have been isolated that actively kill *C. difficile*. All the phages described in the literature as being able to infect this species, appear to have access to the temperate lifecycle and impact of phage infection on *C.difficile* (Bacon et al., 1988, Dei, 1989, Goh et al., 2005b, Govind et al., 2006, Hawkins et al., 1984, Horgan et al., 2010, Mahony et al., 1985, Meessen-Pinard et al., 2012, Nagy and Foldes, 1991, Nale et al., 2012, Ramesh et al., 1999, Sekulovic et al., 2014, Shan et al., 2012).

Phage populations are highly dynamic, and it has been estimated that ~ 10^{23} phage infections occur per second in marine environments (Suttle, 2007). Phages therefore have a vital role in bacterial evolution, from killing bacteria to being a tool for horizontal gene transfer (Kropinski et al., 2007). Genome sequencing of *C. difficile* phages provides insight into how phages impact the biology of *C. difficile*, and studies have therefore focused on the applications of these within therapeutic use and the effects of phage infection on the virulence *C. difficile* (Hargreaves and Clokie, 2014).

No phages have been found that encode genes for toxin. Experiment have demonstrated the presence of the accessory toxin gene, *tcdE* within the genome of temperate phages, but were not able to detect the *tcdA* and *tcdB* genes responsible for *C. difficile* ToxA and ToxB production (Goh et al., 2005a). Another study described the genome characterisation of phiCD38-2, a temperate *C. difficile* siphovirus with a relatively wide host range (Sekulovic et al., 2011), and showed that infection of *C. difficile* strains with phiCD38-2 phage resulted in 6-fold increase in toxin production, although these researchers could not detect any virulence factor or *C. difficile* toxin genes within the genome of phiCD38-2 (Goh et al., 2005a). Another piece of research indicated that purified phage endolysin from a myovirus phiCD27 was active and could lyse a diverse set of *C. difficile* strains, including strains of the major epidemic ribotype 027. This endolysin could therefore be used for therapeutic purposes (Mayer et al., 2008).

The therapeutic application of phages in treating *C. difficile* has been investigated both *in vivo* and *in vitro*. The first *in vivo* model was developed using phage CD140 to treat clindamycin-induced CDI in hamsters. This research showed the efficacy of phages in treating CDI; in this study, 14/18 hamsters treated with phages survived, while the entire infected control group died (Ramesh et al., 1999). An *in vitro* model system was conducted by (Meader et al., 2010) using a batch fermentation model of a system colonised with *C. difficile*. A remedial phage treatment regime caused a transient decrease in the number of viable *C. difficile* cells, and decreased the level of toxin production in comparison with the control, while the gut microbiota remained unaffected by this therapy.

To identify candidate phages for therapeutic purposes, it is necessary to assess their effect on the biology of *C. difficile*, and an update on taxonomic classification of bacteriophages is also required. Rob Lavigne and colleagues have addressed the traditional and genomic classification of bacteriophages using the integration of protein sequence data and physicochemical parameters (Lavigne et al., 2009). Accordingly, the three myoviruses (phiCD119, phiCD27, phiC2) discussed by the International Committee of Viral Taxonomy (ICTV) were assigned to genus phicd119virus, based on the similarity of their genomes and morphological features. Recently, a new species of "phiMMP04virus" has been proposed as forming part of the taxonomic group of *C. difficile* phages; however there remains a less well defined genetic diversity across the sequencing of phages and related prophage (Hargreaves and Clokie, 2015).

1.3 Justification of study

Clostridium difficile is an important nosocomial pathogen that impacts significantly on human morbidity and mortality. The environment is a natural reservoir for *C. difficile*, which can be found in soils, rivers and other natural settings. The prevalence and diversity of *C. difficile* from various environmental sources in Europe has been extensively studied to some extent although most studies predate the use of molecular approaches. However, very little information exists on the occurrence of *C. difficile* in the Middle East (Al-Thani et al., 2014, Rotimi et al., 2003, Esfandiari et al., 2014, Jalali et al., 2012, Nikaeen et al., 2015, Rahimi et al., 2015, Khoshdel et al., 2015), especially in the north of Iraq, where the prevalence of *C. difficile* carriage and CDI are mostly unknown. Furthermore, no previous studies have looked at the biology of this pathogen from natural environments in this geographical region. This can be attributed to the fact that research in anaerobic bacteriology has only recently gathered pace, as beforehand there was a lack of required skills, technology and facilities for culturing anaerobic pathogens (Rotimi et al., 2003).

To address this lack of knowledge regarding *C. difficile* in these areas, we have screened environmental samples for the presence of *C. difficile* and explored the diversity of the strains. Soil and sediment samples were taken from the mountains and alongside rivers in Kurdistan, in northern Iraq. Screening for the presence of *C. difficile* was carried out on the enrichment of these samples.

Since *C. difficile* phages in the area have not been researched either, screening for novel phages was carried out to reveal the diversity of *C. difficile* phages in this poorly studied area and we were particularly interested to find phages that could infect strains of ribotype 078. This ribotype emerged recently as a predominant type in cattle and the CDI which it causes in humans is similar in severity to the CDI caused by R027. Several *C. difficile* phages that were previously isolated from sediment samples in the UK demonstrated a capability to infect many clinically relevant strains. However, only a few phages can infect R078 at present. For therapeutic purposes, it is necessary to have a combination of phages that could infect diverse strains of *C. difficile*, specially the problematic ones.

In addition, we have addressed the phenotypic features of strains isolated from environmental sources under a range of environmental factors in comparison to

genetically related strains isolated from patients with CDI, followed by toxin and sporulation comparison as well as motility with antibiotic sensitivity. Understanding the biology of strains from a marine environment may provide information as to whether strains in this environmental reservoir could potentially contribute to this pathogen's continued evolution through genetic exchange in metabolically active populations.

1.4 Research aim

The aims of this study were to:

1) Isolate *C. difficile* from the natural environment in Kurdistan, northern Iraq, and undertake genotypic and phenotypic characterization of the isolates.
2) Isolate and characterise *C. difficile* phages that could infect *C. difficile* strains of ribotype 078 in a lytic manner;
3) Determine the potential suitability of phages as phage therapeutics and their taxonomic relationship;
4) Determine the phenotype features of environmental strains and compare them to clinical strains of *C. difficile*.

Chapter 2 Materials and methods

This chapter describes the general protocols and reagents used in this study.

2.1 Media, buffers and solutions

All reagents, buffers and solutions used in this study are described in (Appendix 1.and Appendix 2).

2.2 Bacterial isolates and phages

A set of 106 *C. difficile* isolates belonging to 23 ribotypes was used in this study for phage host range analysis and phage plaque efficiency (Appendix3). Of the 106 strains, 38 were isolated from horse faeces, soil and sediment samples collected around Hampshire, UK, by Dr Katherine Hargreaves (Hargreaves et al., 2013), and 31 strains were isolated from clinical samples. Of the 31 clinical strains, 20 strains were isolated by Dr Krusha Patel from toxin A/B +ve faecal samples from patients attending University Hospitals of Leicester (Patel, 2013). The 31 clinical strains included 11 clinical isolate of ribotype 027 typed with multilocus variable-number tandem-repeat analysis (MLVA). Eight strains belong to R001, R012, R017, R017 and R106 obtained from *Clostridium difficile* Ribotyping network (CDRN) were used as reference strains for ribotype profiling. Also another 6 reference strains obtained for Liverpool University were used in this study for the host range analysis. Additionally, 23 strains were isolated in this project from soil and sediment samples from Kurdistan, northern Iraq used for phage host range analysis (Appendix3). A group of 54 *C. difficile* phages available in the databases were used in the study, of these 5 phages were isolated in this study (Appendix 4).

2.3 Routine laboratory culture of *C. difficile* strains

C. difficile was routinely cultured on Brain Heart Infusion (BHI, Oxoid, UK) 1% (w/v) Bacteriological Agar (BA, Oxoid, UK) supplemented with 7% (v/v) defibrinated horse blood (DHB) and incubated at 37°C under anaerobic conditions (10 % H_2, 10 % CO_2 and 80 % N_2) in a MiniMACS chamber (Don Whitley, UK). Colonies were identified by typical characteristics including the distinct yellow/green florescence under long wave UV light (365*nm*), "ground glass" colony morphology about 3-5 mm in diameter and a typical *C. difficile* odour of horse dung/urine (Brazier and Borriello, 2000) (Figure 2-1).

For broth culture, a single colony 48h was removed from the BHI blood plate and transferred to 7ml of pre-reduced Fastidious Anaerobic broth (FA, Oxoid, UK). Pre-reduction was conducted anaerobically for 6 h to pre-warm and eliminated oxygen from the media.

All *C. difficile* isolates were maintained in Protect Bacterial Preservers (Technical Service Consultants Limited, Heywood, UK). To do this, 1 ml of an overnight FA *C. difficile* culture was centrifuged at 13,000 x *g* (Beckmen Coulter, UK) for 5 min at room temperature. The supernatant was decanted and the pellet was re-suspended in 200 µl of glycerol taken from the vials. The solution was transferred back to the vials and stored at -80˚C.

2.3.1 **Isolation of *C. difficile* strains from the environmental samples**

2.3.2 **Sampling sites**

Soil and sediment samples were taken from the mountains and alongside rivers from Kurdistan the North of Iraq. Forty five soil and sediment samples were taken throughout 15 sampling sites. Preference was given to non- habitable areas but samples from areas where there were human activities were also collected. Sampling was performed between August 2012 and January 2014, assistance in performing the sampling was gratefully received from Sana J Rashid. Samples were stored at 4°C until processed.

2.3.3 **Soil and sediment enrichment and recover *C. difficile* isolates**

Approximately 1 g of a sample was suspended in 10 ml of FA broth. This was supplemented with 250 µg ml $^{-1}$ cycloserine and 8 µg ml $^{-1}$ cefoxitin (antibiotics selective for *C. difficile*) and 0.1 % (w/v) sodium taurocholate to enhance spore germination (Wilson et al., 1982). Enrichment was conducted anaerobically for 10 days at 37°C. Afterwards, the cultures were centrifuged at 3398 × g for 10 min. The supernatant was removed, filtered using 0.22 µm filters (Millipore, UK) and stored at 4°C for later use in phage isolation (Hargreaves et al., 2013). The resultant pellets were mixed with an equal volume of industrial methylated spirits (IMS) and vortexed until it was homogenously mixed, then further incubated for 30 min at room temperature. A loop-full of the sample (5 to 15 µl) was sub cultured on Brazier's cycloserine, cefoxitin, and egg yolk (CCEY) agar (Bio Connections, UK) for 24 - 48 h. Colonies were

identified by typical features of grey opaque flat colonies, raised elevation, 2-3mm diameter, ground glass appearance and a rough, fimbriate edge (Figure 2-1). These colonies were purified through three rounds of streaking on 7% (w/v) defibrinated horse blood (DHB) 1% (w/v) agar. Glycerol cryostocks were prepared from the clonal isolates as described in section 2.3.

2.3.4 **Methods for *C. difficile* confirmation**

Suspected *C. difficile* colonies were confirmed by the "ground glass" colony morphology, the characteristic yellow/green florescence under long wave UV light (365nm), and the typical *C. difficile* odour of horse dung/urine (Brazier, 1998). As required, colonies were further confirmed using the *C. difficile* Test Kit (Oxoid, UK) based on latex agglutination was performed according to the manufacturer's guidelines.

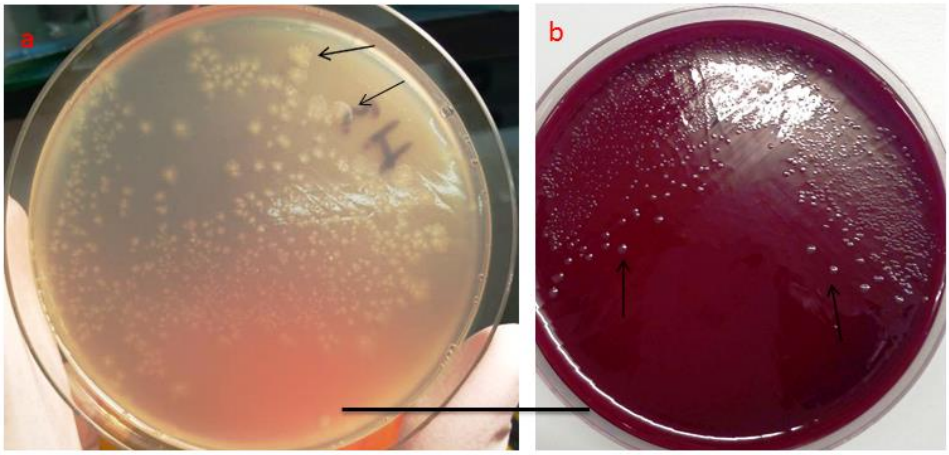


Figure 2-1. Picture showing characteristic colony morphology of *C. difficile* after 48 h.

a: grey opaque flat colonies of *C. difficile* on CCEY agar medium. b: Characteristic ''ground glass'' colony growth on BHI 7 % blood medium. Bar represents ~5 cm.

2.3.5 **Bacterial DNA isolation**

2.3.5.1 **Bacterial DNA isolation using Chelex-100 solution**

Crude DNA template was extracted from identified *C. difficile* colonies using Chelex® 100 Molecular Biology Grade Resin (BioRad Laboratories, California, and USA). To do this, 5% (w/v) Chelex-100 was prepared in 150 µl ultra-pure water, by adding a loop-full of *C. difficile* colonies from a 48 h plate culture. Samples were vortexed and boiled at 100°C for 12-15 min. After cooling for 5 min, the mixture was

centrifuged at 16,000 × g for 10 min. The supernatants containing the DNA template were transferred into a sterile eppendorf tubes and used immediately for PCR ribotyping or stored at -20°C for a maximum of 2 weeks (Bidet et al., 2000, Arroyo et al., 2005, Marsh et al., 2006).

2.3.5.2 **Bacterial DNA isolation using phenol/ chloroform/ isopropanol method**

An inoculum of *C. difficile* starter culture was prepared by transferring a single colony from a 48 h culture on BHI blood plate into 2 ml pre-reduced FA broth. The culture was allowed to grow for 18-24 h. A 5 ml BHI broth was inoculated with 10 % of the starter culture followed by 10-15 h incubation. Bacterial DNA was extracted using phenol/ chloroform/ isopropanol as previously described (Sambrook and Russell, 2001). In brief, the BHI culture was centrifuged at 15,000 × g for 5 min. The resulting pellet was washed once with 500 µl of 0.1 x SSC buffer (Appendix 2) followed by re-suspending the pellet in 300 µl of 10 mM Tris.Cl pH 8.0 containing lysozyme at 2.5 mg/ml. This was incubated at 37°C for 30 min, later the cells were lysed with 500 µl of bacterial lysis buffer containing 0.2 mg/ml of proteinase K. The sample was incubated at 37°C for a further 30 min to allow complete lysis of cells before being centrifuged at 15,000 × g for 5 min. The supernatant transferred into a sterile Eppendorf tube, then an equal volume of phenol/ chloroform/ isopropanol (25: 24:1 v/v/v) was added. The solution incubated and mixed for 2 min at room temperature followed by centrifugation at 15,000 × g for 5 min. The aqueous layer containing the DNA material was cautiously collected and transferred into a sterile Eppendorf tube. This solution was mixed with an equal volume of chloroform/ isopropanol (24:1 v/v) and centrifuged at 15,000 × g for 5 min after 2min incubation at room temperature. After collecting the aqueous layer into a sterile tube, the DNA was precipitated using 2 volumes of isopropanol and 0.4 volume of 7.5 M ammonium acetate followed by 1 h incubation on ice. The DNA material was centrifuged down at 21,000 × g for 30 min at 4°C. The supernatant was discarded and the resulting DNA pellet washed once with 1 ml of 75 % (v/v) ethanol. Finally, the DNA was pelleted by final centrifugation at 21,000 × g for 15 min. The pellet was air dried for 5 min and re-suspended in elution buffer (10 mM Tris. HCl, pH 8). The quantity and quality of the DNA sample samples were checked by NanoDrop ND-1000 spectrophotometer (Thermo Scientific, Wilmington, USA).

2.3.6 **PCR targeting *C. difficile* 16S rRNA gene**

Molecular confirmation for *C. difficile* isolates were achieved on all isolates by amplification of the 16S rRNA gene using *C. difficile* specific primers: cdif-F (5'-TTGAGCGATTTACTTCGGTAAAGA-3') and cdif-R (5' CCATCCTGTACTGGCTCACCT-3') as described previously by Rinttila and, colleagues producing an expected 157 bp product (Rinttila et al., 2004). The amplification reaction was carried out in a total volume 50 µl standard reaction containing DNA template, 4 µM forward and reverse primers, 0.25 mM dNTPs, 4 mM $MgCl_2$, 1 × PCR buffer, and 0.5 U of BioTaq DNA polymerase (Bioline, UK). Amplification conditions were: initial denaturation at 95°C for 5 min; 40 cycles of denaturation at 95°C for 30 s, annealing at 55°C for 30 s, extension at 72°C for 45 s; with a final extension of 5 min at 72°C. The expected 157 bp products were visualised alongside 7.5 µl of GeneRuler 1 kbp DNA ladder (Fermentas, UK) on a 1% 9w/v) Helena Agarose gel (Helena Biosciences, Tyne and Wear, UK). This was prepared in 1x TAE buffer stained with Gel Red and run at 90 Volts for 1 hour in TAE buffer. Gels were imaged using SynGene software in a in a transilluminator.

Further molecular confirmation of the isolates was carried out using universal bacterial 16S rRNA primers 8F (5'-AGAGTT TGATCCTGGCTCAG-3') and 1391R (5'-GACGGGCGGTGTGTRCA-3') producing a 1800 bp product as designed by previously (Lane et al., 1985, Edwards et al., 1989). PCR reactions were performed in a volume of 50 µl with: template DNA; 2 µM forward and reverse primers; 0.25 mM dNTPs; 2 mM $MgCl_2$; 1 × PCR buffer and 0.5 U of BioTaq DNA polymerase. Amplification was carried out at: 95°C for 5 min followed by 30 cycles of 95°C for 30 s, 55°C for 60 s and 72°C for 2 min; with a final extension step at 72°C for 10 min. PCR products were gel-purified by using Qiagen Gel Elute kit. The purified PCR products were sent to GATC biotech, UK for sequencing, and data were analysed using Chromasv1.45. Sequences were blasted against the NCBI non-redundant nucleotide collection using the Blastn algorithm accessed online to assign bacterial species.

2.3.7 **Agarose gel-based and Capillary-based PCR ribotyping**

PCR ribotyping is considered on the isolates by amplification of 16S to 23S intergenic spacer regions in the rRNA operon. Several alleles of different sizes are

present in this operon on the *C. difficile* chromosome (Bidet et al., 2000). To do this one set of primers was used, forward (5'-GTGCGGCTGGATCACCTCCT-3') and reverse (5'-CCCTGCACCCTTAATAACTTGACC-3'). The reaction was carried out in a volume of 50 µl. PCR reactions were performed with SensoQuest LabCycler (SensoQuest GmbH, Germany). The 50 µl PCR reaction mixtures contained 0.5 µM of each of the primers, 0.25 mM of dNTPs, 1 x PCR buffer (Bioline, London, UK), 0.1 U of Taq polymerase (Bioline, London UK) and 2 µl of DNA template except for the negative control where equal volume of ultra-pure water replaced the template. The PCR ribotyping conditions were: denaturation at 95^{o}C for 120 s followed by 30 cycles denaturation at 92^{o}C for 60 s, annealing at 55^{o}C for 60 s, elongation at 72^{o}C for 90 s and a final extension at 72^{o}C for 5 min. The reaction mixture was reduced in volume to ~20 µl on a heating block set at 75^{o}C. PCR products resolved in 1 % w/v RESponse Regular PCR Agarose gel (Bioplastics, The Netherlands) prepared in 1×Tris-acetate-EDTA (TAE, pH 8) containing 20 µl/ml Ethidium Bromide. HyperLadder IV (5 µl) (Bioline, London, UK) was used as a molecular standard. The gel was run in TAE buffer at 90V for 4 h. Gels were analysed using Genesnap software (SynGene) in the UV transilluminator.

Capillary gel electrophoresis an alternative ribotype method was performed on all isolates. PCR reactions of 25 µl contained, 12 µl of HotStarTaq Mix (Qiagen, UK) added to 9µl upH_20, 1µl of the same primers as listed above were used except that the 16S primer was fluorescently labelled with Biosystems carboxyfluorescein (FAM) at the 5' end, and 1µl of the 23S rRNA gene primer (Applied, UK and vhbio, UK) respectively with 1µl DNA template (Indra et al., 2008). The reaction was performed in a PCR thermo cycler running for 1 cycle of 6 min at 94°C as initial step for denaturation, followed by 35 cycles (1 min at 94°C, 1 min at 56°C and 1min at 72°C) followed by final elongation for 7 min at 72°C (Stubbs et al., 1999). Thereafter the PCR fragments were analysed by PNACL at the University of Leicester on an ABI 3730 genetic analyser with a 50 cm capillary, loaded with a POP7 gel (Applied Biosystems, UK). The reaction carried out in a total volume 25 µl, 1µl of DNA template with 10µl of a master mix, containing 9µl Formaldehyde and 1µl DNA ladder (Gene scan 1200 LIZ). The samples were injected with 1.6 kV over 15 seconds with a total running time of 103 min at 8 kV run voltage. A 20-1200 bp fluorescent Genescan ladder (Applied Biosystems, UK) was added to each sample. Data was analysed using Peak Scanner

v1.0 (Applied Biosystems). Peaks which were at least 10% of the highest peak of the analysis were considered, and those between 250 and 650 bp were accepted as genuine fragments. A subset of strains (n=12) those with different banding patterns were sent to the Leeds Reference Laboratory, UK for ribotyping, confirming toxin profiles and their antibiotic sensitivity.

2.3.8 **Toxin gene screening by PCR of *C. difficile* isolates**

The presence of toxin genes was determined by carrying out the PCR assay established by (Kato et al., 1991). A PCR was run using the primer pairs NK2 (5'-CCCAATAGAAGATTCAATATTAAGCTT-3') and NK3 (5'-GGAAGAAAAGAACTTCTGGCTCACTCAGGT-3'), which amplified the partial sequences of tcdA and generating an expected product of 251 bp, and primers NK9 (5'-CCACCAGCTGCAGCCATA-3') and NK11 (5'-TGATGCTAATAATGA ATCTA AAATGGTAAC-3') targeting the essential repeat region within tcdA generating an expected product of 1,265 bp. These primers were used in a multiplex reaction in a 50 µl volume: template DNA; 1 µM forward and reverse primers; 0.25 mM dNTPs; 1.5 mM $MgCl_2$; 1×PCR buffer and 0.5 U of BioTaq DNA polymerase. Toxin B gene was detected using primers NK104 (5'-GTGTAGCAATGAAAGTCCAAGTTTACGC-3') and NK105 (5'-CACTTAGCTCTTTGATTGC TGCACCT-3') to generate a product of 203 bp. (Barroso et al., 1990). The PCR was carried out with the primers in a reaction of 50 µl volumes with: template DNA; 2 µM forward and reverse primers; 0.25 mM dNTPs; 1.5 mM $MgCl_2$; 1× PCR buffer and 0.5 U of BioTaq DNA polymerase. Amplification conditions for all NK primers were: denaturation at 95°C for 5 min followed by 32 cycles of 95°C for 20 s, 62°C for 120 s and 72°C for 2 min; and final extension phase for 5 min at 72°C. (Kato et al., 1998, Barroso et al., 1990). Carriage of the binary toxin genes *cdtA* and *cdtB* was determined using primers designated by Stubbs and colleagues (Stubbs et al., 2000), primers cdtA forward (5'-TGAACCTGGAAAAGGTGATG-3') and reverse (5'-AGGATTATTTACTGGACCATTTG-3') producing an expected product of 375 bp and *cdtB* forward (5'-CTTAATGCAAGTAAATACTGAG-3') and reverse (5'-AACGGATCTCTTGCTTCAGTC-3') with a product size of 510 bp. The PCR run in a 50 µl multiplex reaction: template DNA; 1 µM forward and reverse primers; 0.25 mM dNTPs; mM $MgCl_2$; 1× PCR buffer and 0.5 U of BioTaq DNA polymerase. PCR reaction conditions were: denaturation at 95°C for 5 min followed by 30 cycles of 94°C

for 45 sec, 52°C for 60 s and 72°C for 2 min; and final extension stage for 5 min at 72°C. PCR products were separated on a 1% (w/v) Helena agarose gel prepared in a 1 × Tris-TAE buffer and visualised using Ethidium Bromide and compared to a 1 kb ladder.

2.3.9 **Prophage Induction**

Phage induction is the switch form lysogeny cycle to lytic cycle after antibiotic treatment (Clokie and Kropinski, 2009). Twelve of the strains were examined for prophage release by inducing with mitomycin C and norfloxacin. Approximately 10 ml of overnight BHI broth cultures were induced with mitomycin C (Fisher Scientific, Loughborough, UK) and norfloxacin (Sigma Aldrich, Dorset, UK) at a final concentration of 3 μg ml^{-1} for 24 h (Nale et al., 2012, Shan et al., 2012). Cultures were then centrifuged at 3400 × *g* for 10 min. The resulting supernatants were filtered through a 0.22 μm filter and examined by using a Transmission Electron Microscope (TEM). The Polyethylene glycol (PEG) 8000 was used to concentrate the phage particles (Clokie and Kropinski, 2009, Nale et al., 2012).

2.3.10 **Phage purification using polyethylene glycol**

The induced phage supernatants were purified using a solution of polyethylene glycol (PEG) and NaCl (Yamamoto et al., 1970, Goh et al., 2005b). The solution of PEG and NaCl enable the bacteriophages to be pelleted down during centrifugation. The resultant pellet was resuspended in a small amount of phage buffer, resulting to ~100-fold concentration of the original the induced lysate (Yamamoto et al., 1970). To purify the phage supernatant using PEG, NaCl was added to 1 M, mixed and incubated on ice for 1 h and then centrifuged at 10,000 × g for 5 min. Then 10 % (w/v) of PEG 8000 (Fisher Scientific, New Jersey, USA) was slowly added and continuously stirred until completely dissolved. This was kept at 4°C overnight and afterwards centrifuged at 15,000 × g for 20 min at 4 oC to precipitate the phage. The supernatant was discarded and the pellet was re-suspended in 1 ml of SM buffer (10 mM NaCl, 8 mM $MgSO_4$.7H2O and 50 mM Tris-Cl), washed with equal volume of chloroform and centrifuged for 15 min at 5,000 × g. The top phase was collected and filtered using membrane filter of pore size 0.22 μm. The filtrate was used for TEM analysis

2.3.11 **Transmission electron microscopy (TEM)**

TEM was carried out by Stefan Hyman and Natalie Allcock (The Electron Microscope Laboratory University of Leicester). Briefly, samples were placed on individual glow-discharged pioloform/carbon-coated copper grids (Athene type 3mm, Agar Scientific, UK), prepared by applying a thin film of 0.25% pioloform and allowed to dry for 1 h. The pioloform coated copper grids were then carbon coated for 1 to 3 s before using high-voltage glow discharged for 30 to 60 s. Immediately afterwards immediately 5 µl of phage sample was placed onto glow discharged pioloform / carbon coated grid for 2 min. The excess sample was absorbed using Whatman paper. The grid was washed twice with 5 µl of double distilled water to eliminate buffer salts. Amount of 5 µl of 1% uranyl acetate (w/v) was used for staining the phage samples for 5 to 10 s. The excess was removed immediately by blotting with Whatman paper then air dried for 3-5 min. The samples were examined with JEOL 1400 (Joel, UK) TEM with an accelerating voltage of 80kV. The sample images were taken using SIS Megaview III digital camera with analysis software (Olympus) (Nale et al., 2012, Shan et al., 2012).

2.3.12 **Sensitivity of *C. difficile* strains to bacteriophages.**

Susceptibility of environmental strains to 5 sipho and 7 myoviruses were tested. Five phages (CDKM9, CKM15, CDKM7, CDKM7 and CKM8) were isolated from the same reservoir and 7 phages (phiCDHM1, phiCDHM2, phiCDHM3, phiCDHM4, phiCDHM5, phiCDHM6 and phiCDHS1) were from different bacterial communities (Appendix3). Spot test were used for testing phage sensitivity of the strains and were performed by Gurinder Vinner during her training project in the lab. The whole spot test was performed in triplicate and repeated three times.

2.3.13 **Spot test**

The standard spot test was used to determine the susceptibility of each *C. difficile* strains to phages using phage stock of 10^8 PFU/mL as follows. An amount 250 µl of overnight culture of each of the bacterial strain was mixed with 3 mL of BHI 0.5% (w/v) agar over lays supplemented with salt solution (0.4 M $MgCl_2$ and 0.01 M $CaCl_2$), and poured onto BHI 1% (w/v) Agar plate. After the overlay had solidified, 10 µL of the diluted phage lysate at each dilution to $10^{-}3$ was spotted onto the agar. Following 24 h incubation at 37°C under anaerobic conditions plates were inspected for lysis (a clearing of the bacteria within bacteriophage spot) occurred (Goh et al., 2005a).

2.3.14 **Isolation and characterization of *C. difficile* bacteriophage**

2.3.15 **Isolation of phages from the environmental samples**

In order to isolate free phages, 1-2 g of soil and sediment samples were mixed with an equal volume of SM buffer (0.1 M NaCl, 8 mM $MgSO_4.7H_2O$ 50 ml of 1 M Tris–HCl) for one hour on a shaking platform. This suspension was centrifuged at 3,400 x g for 10 min. The supernatants were stored at 4 °C after filtration through 0.22 µm pore size filters. Enrichment cultures from the isolation of *C. difficile* as mentioned previously in section 2.3.3 were also used to isolate the phages. Screening for virulent phages was done by spot test assay using a double agar overlay lawn method as described previously (Goh et al., 2005b). Different plaque sizes were picked from the agar using sterile Pasture pipettes and placed in 500 µl of BHI broth; this was incubated for minimum one hour at 4°C to allow the phage to diffuse out of the agar. The resuspended plaques were centrifuged at 10000 × g for 5 min and the filtered supernatant was used in the next round of purification. Plaques were subjected to this purification 3 times to ensure each phage clonal (Hargreaves, 2012).

2.3.16 **Bacteriophage propagation**

From the time since discovery of bacteriophages in the twentieth century, there have been two basic methods of propagating virulent phage, both based on the ability of phage to kill the host bacterial cell (Adams, 1959).

2.3.16.1 **Plaque assay**

Plaque assay is the most regularly used today, and the one that is taught in most microbiology teaching laboratories (Lillehaug, 1997). Briefly, BHI 1% (w/v) agar plates were dried in the flow hood and labelled. Soft agar and salts were added in ratio of 1:1 and mixed well. 3 ml of soft agar was added to a bijou containing 250 µl of overnight FA culture and 10-250µl phage solution dependent on the titre of the phage stock. After mixing, the solution was poured onto a BHI plate, left to set for 5 min and incubated overnight in the 37°C anaerobically. The plates were then checked for complete lysis or plaques (Cornax et al., 1990, Mahony et al., 1985, Sekulovic et al., 2011).

To make up phage stocks, confluent lysis was required after incubation of ten plates with plaque assays prepared as above. However, if plaques were not confluent the amount of phage added was increased in the next attempt. If confluent lysis was

obtained, the soft agar layer of the plates was scraped into a centrifuge tube. An equal volume of BHI broth was added to the 15ml tube, mixed well by inversion and flicking, rocked for 30 min to mix and stored overnight at 4˚C to allow phage to dissociate from the agar. Next day, the mixture was centrifuged at 10,000 x g for 20 min and the supernatant filtered through a 0.22 μm filter (Millipore, USA) and transferred to a clean tube and stored at 4˚C. Phage titre of stock solution was determined by spot tests as described in section 2.3.13.

2.3.16.2 **Bacteriophage propagation in liquid**

This method is usually used for the growth of significant amounts of phage. A 50 ml log phase culture of the host bacteria (OD_{550} 0.2) were infected with amount of phage lysate (10^8 to 10^9 PFU/ml) dependent on the multiplicity of infection (MOI) being tested (MOI 10). After a gentle mix by inversion the culture was incubated anaerobically at 37°C for 24 h. The next day the culture centrifuged at 4000 x g for 10 min (Clokie and Kropinski, 2009). This was then filtered through a 0.22μm filter (Millipore) and transferred to a clean 15ml centrifuge tube. The supernatants were assayed by spot test as described in section 2.3.13.

2.3.17 **Host range of *C. difficile* phages**

The host range of the isolated phages was determined by using spot test as described previously in section 2.3.13. A diverse collection of strains were used including environmental and clinical isolates representing different PCR ribotypes (Appendix3). The phages were also tested against isolates obtained from this project (local environment). 10 μl of diluted and undiluted of phage stock (10^8 Pfu/ml) were spotted onto a lawn of host bacteria then plates were examined for cell lysis after incubation anaerobically for 24 h. These spot tests were performed at three independent times using three replicates. Infection was considered positive when clearance was seen in all three tests.

2.3.18 **Phage genomic DNA extraction**

Extraction of phage DNA was performed using phenol: chloroform: isolamyl method from a high titter phage lysate 10^9 PFU/mL following the described protocol by Nale et al. (2015b). Samples were treated with 1.4 μg μL^{-1} DNase, 3 μg μl^{-1} RNase (Sigma, USA) and 12.5 μL 1 M $MgCl_2$ (Acros Organics, New Jersey,USA) then incubated overnight at 37˚C. Proteinase K (Fisher Scientific, UK), EDTA (Sigma,

USA), and SDS were added to final concentration of 0.5 mg mL^{-1}, 20 mM and 0.5% (w/v) respectively, followed by 1 hour incubation at 55°C. An equal volume of phenol: chloroform: isoamyl alcohol (25:24:1) (v/v/v) was added to the samples and mixed by inversion then centrifuged at 13,000 × g for 5 min, the top aqueous layer was taken and transferred into new Eppendorf. This step was repeated 3-4 times until the white interphase ewas not observed. The collected aqueous layer was treated with 3M sodium acetate solution (Fisher Scientific, UK) and 95% (v/v) cold ethanol followed by 30 min incubation on ice. DNA pelleted by centrifugation for 30 min at 4°C, then washed with 500 µL of 70 % (v/v) ethanol before re-suspension of the DNA with 60 µL of 5 mM Tris Cl, pH 8.5.

DNA was quantified using a Nanodrop 2000and Qubit dsDNA HS Assay Kit on a Qubit Fluorometer using manufacturer's recommendations (ThermoScientific, UK) respectively. To inspect genomic DNA contamination, PCR was carried out, using universal bacterial 16S rRNA primers 8F and 1391R that amplify the bacterial 16S rRNA gene as described in section 2.3.6.

2.3.19 **Pulsed-field gel electrophoresis (PFGE) of phage particles.**

To determine the whole genome size of the phages PFGE was performed on phage lysate (Goh et al., 2007). As PFGE needs a high concentration of DNA, a high titre of Phage lysate 10^9 PFU/mL was used for plug formation.

2.3.19.1 **Plug formation**

The formation of PFGE plugs were performed as previously described (Wakita et al., 2002). Concisely, the casting mould of plug was fastened at the bottom by a piece of masking tape before labelling the wells. Seaplaque CTG agarose (2 % w/v) (Cambrex Bio Science Wokingham, Berkshire, UK) in 0.5 × Tris-Borate-EDTA (TBE, pH 8) was prepared and maintained fluid by incubation on a heat block at 100°C. Approximately 40µl of phage samples were transferred into marked tubes, and then 60 µl of warm plug agarose were added, followed by mixing by gentle pipetting up and down, avoiding the formation of bubbles. . The mixture was directly placed into the marked gel moulds and left to set for 30 min at 4°C. The phage samples were digested with 1ml of lysis buffer added 0.5 mg/ml final concentration of proteinase K (proteinase K was added just before use) (Appendix 2).The plugs were added to the lysis buffer by pushing them

from the back of the mould and incubated in a water bath at 55°C for overnight before being washed three times with 1x Tris-EDTA buffer (TE, pH 8) (Appendix 2).

2.3.20 **Pulsed-field gel electrophoresis process**

The PFGE tank and combs were assembled after 1% (w/v) pulsed-field certified agarose (BioRad Laboratories, Carlifornia, USA) was prepared in 0.5 x TBE and boiled in microwave. The warm agarose was poured into the tank and left to cool at room temperature for 1 h. The plugs were washed three times with TE buffer and loaded into wells of the Pulsed field certified agarose gel using 70 % v/v ethanol-sterilised spatulas. About 2 mm of the low range PFG marker (New England Biolabs, Herts, UK) was cut and loaded in one of the empty wells of the gel using sterilised blade. The wells were sealed up with the remaining molten agarose. The products were separated in 0.5 × TBE buffer using a Bio-Rad CHEF-DR-II Pulsed Field Electrophoresis System (Bio-Rad, Richmond, CA) at 6 Volt cm^{-1} for 17 h at 14 °C (ramped pulse times were: initial 5 sec; final 13 sec) followed by staining with GelRed (Cambridge Biosciences, Cambridge, UK) (Nale et al., 2012). Synegene box was used to visualise the gel and images captured using Genesnap software.

2.4 Genome sequencing and annotation of *C. difficile* strains and their phages.

2.4.1 *C. difficile* strains genome sequencing and annotations

Genomes of 14 *C. difficile* strains isolated in the north of Iraq were subjected to whole genome sequencing using Illumina MiSeq with 2 x 250 bp paired-end performed kindly by Dr Andrew Millard, University of Warwick. The genomic DNA was prepared from overnight culture using a QIAGEN Genomic- tips kit (Qiagen, UK) according to manufacturer`s instructions. An amount of 1 ng of bacterial DNA was used with the Nextera XT DNA sample preparation kit (Illumina), following the manufacturer's instructions. Libraries were sequenced using a MiSeq V2 reagent kit (2 x 250 bp). Genomes were assembled using SPAdes 2.0 with the following parameters "-k 21,33,55,77,99,127 --careful". Contigs were ordered against the reference strain *C. difficile* CD630 (NC_009089) using MAUVE (Darling et al., 2010). PROKKA was used to annotate the genomes with the following settings "-- compliant -- genus *Clostridium* use genus" (Seemann, 2014). The genomes of the 14 strains were submitted to the EBI under the Project 193 accession PRJEB8702.

Sequence types (ST) were determined form the assembled genomes by extracting the alleles and querying these against the curated *C. difficile* database (http://pubmlst.org/*C. difficile*/). Core genes were identified using reciprocal best blast hit, utilising orthAgogue with the following settings "-b -e 6 ", followed by clustering using mcl: "–abc – I 1.5" to allow identification of single copy core genes (Darling et al., 2010).

Genome of the strains were subjected to phylogenetic analysis by constructing a phylogenetic tree was based on the core genes; from this single copy genes that were found in all genomes were extracted. To do this, the sequences for each gene set were first aligned using MUSLCE (Edgar, 2004) and further concatenated into a single alignment. Alignments were trimmed with trimal options '- automat (Capella-Gutiérrez et al., 2009). Phylogenetic reconstruction was carried out with RaxML using a GTRGAMMA model with 100 bootstrap replicates (Stamatakis, 2014). Trees were visualized in an NJ plot and edited in Inkscape (Perrière and Gouy, 1996). Whole genome comparison was performed for the 9 strains of clade 1 against the type strain CD630 as a references, and visualised in BRIG v 0.95 using BLASTn (Alikhan et al., 2011).

2.4.2 **Prophage Prediction and induction from *C. difficile* strains**

Genomes of the 11 strains representative from each ribotype were screened for predicted regions using PhiSpy with default parameters (Akhter et al., 2012), then followed by manual checking and examination using Artemis Genome Browser and Annotation Tool (Carver et al., 2012). To further confirm phage presence, mid-log cultures of the 11 strains were treated with mitomycin C and norfloxacin to mediate phage release. Approximately 10 ml of overnight BHI broth cultures were induced with mitomycin C (Fisher Scientific, Loughborough, UK) and norfloxacin (Sigma Aldrich, Dorset, UK) at a final concentration of 3 µg/ mL for 24 h incubation at 37°C in anaerobic chamber (Nale et al., 2012, Shan et al., 2012). Polyethylene glycol (PEG) 8000 was used to concentrate the phage particles (Clokie and Kropinski, 2009, Nale et al., 2012, Hargreaves, 2012). The supernatants were filtered through 0.22 µm, after being centrifuged at 3400 × *g* for 10 min. Phages within the induced lysates were identified using a TEM analysis. TEM was carried out by the Electron Microscope Laboratory at the University of Leicester as described in section 2.3.11

2.4.3 **Identification of Clustered Regularly Interspaced short Palindromic Repeats (CRISPR) in *C.difficile* strains**

The genomes of the 11 strains were examined for the CRISPR arrays, arrays were predicted using PILERC-CR 1.06 with default settings (Edgar, 2007, Ekseth et al., 2014). The CRISPR spacers from the strains were extracted and compared with the *C. difficile* phage genomes from the GenBank (Appendix3) and the default plasmid database. CRISPRtargets were used to identify a possible extrachromosomal origin (match reward +1, mismatch penalty -10, minimum score 30) (Biswas et al., 2013).

2.4.4 ***C. difficile* phage genome sequencing and annotation**

Genomes of three phages (CDKS6, CDKS8, & CDKM15) were submitted for sequencing using the Illumina Hiseq 2000 platform at Beijing Genomics Institute (Shenzhen, China). Paired-end libraries were prepared using 3 µg of DNA with insert size of 170 bp that was sequenced and assembled using SOAP denovo version 2.04 with the following parameters ‘‘*-K*-P*-F*-M*-d*-R*-u*-k*-o’’ (Li et al., 2010, Li et al., 2008). The reads were assembled into one contig for CDKS6, CDKS8 and two contigs for CDKM15. The contigs were joined by PCR using primers designed using Primer 3 v0.4.0 (Rozen and Skaletsky, 1999). To close the contigs F1 5’-TCCAGCAATAACACCACCAA-3’ and R1 5’-TCACCATCTCACTCCATCCA-3’ with a 545 bp product were used. PCR reactions were performed in 50 µl volumes containing: template DNA; 10µM forward and reverse primers; 2.5 mM dNTP; 3mM $MgCl_{2;}$ 1× PCR buffer and 0.5 U of BioTaq DNA polymerase. Amplification conditions were: denaturation at 98°C for 120 s; followed by 30 cycles denaturation at 98°C for 20 s; annealing at 57°C for 20 s; elongation at 72°C for 30 s; with final extension of 60 s at 72°C. All products were visualized on a 1% (w/v) Helena Agarose gel prepared in 1x TAE buffer and stained with Gel Red. This was run at 90 volts for 1 h in TAE buffer and samples were loaded alongside 10µl of GeneRuler 1 kbp DNA ladder. Sequencing was performed on the gel-purified PCR products prepared using the QiAquick Gel Extraction kit (Qiagen, UK) following the manufacturer’s instructions. Sanger sequencing was carried out at GATC biotech Ltd (UK). Data were edited using Chromas v1.45 and sequences aligned with ClustalW2 ((Larkin et al., 2007, Goujon et al., 2010), the genome was re-annotated in Artemis.

Genomes of phages CDKM9 and CDKS7 were sequenced and assembled at Warwick University by Dr Andrew Millard using Illumina Nextera XT DNA sample kit as manufacturer's protocol (Illumina, San Diego, USA) with 3800 coverage. Amount of 1 ng input DNA, as quantified by Qubit (Life Technologies, Carlsbad, USA) was used to prepare a genomic library. Sequencing was performed on an Illumina MiSeq using the paired-end 2 x 250 bp protocol (version 2, 500 cycles) 3800x coverage. Read quality was checked with FastQC version 0.11.3 (Andrews, 2010), trimmed with sickle (Joshi NA, 2011) and assembled into one contig with SPades 3-1 (Bankevich et al., 2012).

The reads of phages were mapped against the assembled genome with bwa-mem 0.7.5a to detect assembly errors. Sam and bam files were manipulated with Samtools V1.4 (Li et al., 2009). Depth of coverage and insert size were calculated using QualimapV2.2 (Okonechnikov et al., 2015).

After re-mapping of the reads to final assemblies their arrangement was analyzed to determine the position of physical ends of the DNA molecule using Pause tool (https://cpt.tamu.edu/computer-resources/pause/). The physical terminus of the CDKM15, CDKM9 genomes were searched for by PCR using specific designed primer but were not detected. Protein coding genes were predicted using GeneMarkS, GeneMark.hmm (Besemer and Borodovsky, 2005), Glimmer 3 (Delcher et al., 2007), RAST (Overbeek et al., 2014), FGENESV (Softberry, Inc.) and Prodigal 1.20 (Hyatt et al., 2010). CDSs with no overlapping BLASTx hits (against the NCBI nr database) predicted by only a single tool were discarded. BLASTx analysis was also used for functional annotation of CDSs. Predicted coding sequences were then translated and their initial annotation was re-assessed using BLASTp, InterProScan 5 and CD-Search (Altschul et al., 1990, Marchler-Bauer and Bryant, 2004, Jones et al., 2014). tRNA genes were predicted by tRNAscan-SE version 1.21, while other non-coding RNAs by Infernal 1.1.1 (Schattner et al., 2005, Nawrocki and Eddy, 2013). CRISPR arrays located by the Infernal were confirmed by PILER-CR 1.06 and CRISPR Finder (Edgar, 2007, Grissa et al., 2007). CRISPRtarget was used to identify matches to the spacers in the array (match reward +1, mismatch penalty -10, minimum score 30) (Biswas et al., 2013). Protospacer adjacent motifs (PAMs) were identified by visual inspection of the protospacer flanking sequence. Genome maps were generated and visualised in Geneious (Kearse et al., 2012). The annotated genomes of the three phages were oriented to start with the small terminase subunit to be consistent with the other

sequenced *C. difficile* phage genomes (Govind et al., 2006, Hargreaves et al., 2014b, Goh et al., 2007).

2.4.5 **Phylogenetic analysis of *C. difficile* phages**

To determine the taxonomic position of CDKS8, CDKM15 and CDKM9 relative to the known *C. difficile* phages (n = 22) genome comparisons were performed using Gegenees 2.2.1(Appendix 4**)**. The program calculates global similarity between pairs of sequences based on BLAST local alignments (we used both BLASTn and BLASTx method with fragment size of 200 sliding window size of 100). The resulting BLASTx similarity matrix was used to construct BioNJ phylograms with SplitsTree 4.13.1(Ågren et al., 2012, Huson and Bryant, 2006). The generated tree was visualized in the Geneious tree viewer (Kearse et al., 2012).

2.4.6 **Phylogeny of the endolysin genes**

To determine whether the genes responsible for host cell lysis followed the same evolutionary history of the overall phage genomes, a neighbour-joining tree based on the amino acid sequences of endolysin gene was constructed. The amino acid sequences of homologues for the endolysin gene in related phages were retrieved from GenBank using BLASTp and confirmed following protein domain identification using InterProScan and NCBI Conserved Domain Search databases (Appendix 5). After aligning the sequences with ClustalX2 (Larkin et al., 2007), evolution models were inferred using the neighbour-Joining method and bootstrap of 1000 replicates, the analysis performed on MEGA 7.0 (Kumar et al., 2016). The tree was visualised with Fig tree v1.4.2.

2.4.7 **Phylogenetic analysis of TerL and the packaging strategy of isolated phage**

The packaging strategy and type of DNA ends that a phage uses can be positively predicted from the amino acid sequence of a phage's large terminase subunit (Casjens and Gilcrease, 2009). The packaging strategy of the phages was predicted by aligning large subunit terminase genes from *C. difficile* phages to corresponding proteins from reference phages (Appendix 6) using ClustalX2. From the aligned sequences, a neighbour-joining tree was constructed as proposed by Casjens and Gilcrease (Casjens and Gilcrease, 2009), and phylogenies were determined by bootstrap analysis of 10000 replicates in MEGA 7.0 (Kumar et al., 2016). The tree was visualised in Fig tree v1.4.2.

2.4.8 **Genome comparison of CDKM9 and CDKM15**

In order to visualise the patterns of sequence similarity across the genomes of phages CDKM9 and CDKM15, each were used as a reference to compare with 22 *C. difficile* phage genomes using BLAST Ring Image Generator (BRIG). tBLASTx as a comparison algorithm was used to generate circular maps of genomic similarity (Alikhan et al., 2011, Katoh et al., 2002). Default parameters and settings were used for all tools unless stated otherwise.

2.4.9 **Growth curve comparison of *C. difficile* isolates**

Growth comparison for representative strains from the environment and the hospital sets of the ribotypes 027, 078 and 010 was performed. The selected ribotypes each had representative strains from environment and clinical sources to allow comparison between the strains within each ribotype. All the growth experiments were carried out under anaerobic conditions using Hungate tubes (SciQuip, UK). 4.5 ml of standard pre-reduced BHI broth was placed in the tubes. 0.5 ml of an overnight FA culture was used as to inoculate 4.5 ml of pre-reduced BHI broth. Growth of *C. difficile* was monitored by measurement of $OD_{600\ nm}$ (WPA, Biowave C08000) over a period of 48 h, with a tube containing uninoculated FA broth was used as a control. Cultures were checked for purity by colony morphology and PCR targeting the *C. difficile* 16 S as described previously in section 2.3.6.

BHI broth at differing pH (8) was prepared by adding drops of 1 N NaOH to a desired pH and checked with a pH meter (JENWAY 3310) prior to autoclaving. BHI broth with salt concentrations of 2 % (w/v) was prepared using the appropriate amount of analytical grade NaCl before autoclaving (Appendix 2).

2.4.10 **Toxin production assay**

As described previously by Vohra and Poxton (Vohra and Poxton, 2011), total toxin A and B production was measured for *C. difficile* environmental and clinical strains using the *C. difficile* TOX A/B II™ kit (TechLab) according to the manufacturer's instructions. Briefly, an equal amount (20 µl) of 10^5/ml titre of the purified spore suspensions was added to 1 ml of pre-reduced BHIS supplemented with of L-cysteine and sodium taurocholate. This was incubated anaerobically for 18 h at 37°C until an OD_{600} of 1.0 (±0.05) was achieved. The appropriate volume 1% (v/v) of the starter culture was then inoculated into pre-reduced BHI to obtain a culture, which

was grown anaerobically at 37°C for 48 h. Culture supernatants were collected at 6, 8, 24, 32 and 48 h by centrifugation at 13,000 g for 1 min, then aliquoted and stored at -80°C until use; samples were thawed only once. The supernatants were diluted 1: 5 in the supplied diluent. The wells provided with the kit were coated with affinity-purified goat antibodies specific for toxin A and toxin B (Vohra and Poxton, 2011). Two wells were prepared for each sample by adding an aliquot (50 µl) of the conjugate containing two horseradish peroxidase coupled antibodies: a mouse monoclonal antibody specific for toxin A, and a goat polyclonal antibody specific for toxin B. 100 µl of diluted sample was added to each well. For the positive controls, one drop of the provided positive control was added, while for the negative controls, one drop of diluent was added. The wells were covered, incubated at 37°C for 50 min, followed by five washes with the supplied wash buffer solution containing phosphate-buffered saline and detergent (Vohra and Poxton, 2011).

Two drops (100 µl) of substrate (tetramethyl benzidine and peroxide) were then added to the wells, followed by incubation for 10 min at room temperature with continuous shaking. To stop the reaction, an aliquot (50 µl) of the stop solution (0.3 M sulphuric acid) was added to each well. Afterwards, the plate was incubated at room temperature for 2 min. The intensity of the colour developed was determined by measuring absorbance at 450 nm.

2.4.11 **Induction, isolation and purification of spores**

C. difficile spores were prepared according to the basic protocol proposed by Sorg and Dineen (2009). An amount of 1 ml of overnight liquid culture of FA broth was transferred into fresh pre-reduced BHI. When OD_{550} ~ 0.7-0.9, 150 µl of culture was transferred into the wells of the 6-well plates containing pre-reduced BHI 1% (w/v) agar. The plates were incubated anaerobically for 7 days to induce sporulation. The bacterial culture from the top of the agar from the six wells was gently scraped into a 15 ml tube containing 1 ml of ice cold water using a sterile inoculation loop. Residual vegetative cells and spores were recovered by washing the plate wells with ice cold water to obtain a total of 10 ml spore suspension in a tube for each sample. The suspension was centrifuged for 1 min at 14,000 × g at room temperature. The supernatant was gently removed; the washing and centrifugation steps were repeated 9 times. The final pellet was then re-suspended with 20% w/v HistoDenz and combined to

make 5 ml. This suspension was gently lowered into 10 ml 50 % HistoDenz in 15 ml tube (care was taken not to disturb the density interphase) (Nale, 2013, Paredes-Sabja et al., 2008). This was then centrifuged for 20 min at 15, 000 x g and 4°C. The supernatant was then removed (debris at the interphase was removed first) and the spores re-suspended in 1 ml ice cold water. The spores were washed once and finally resuspended in 200 ml of ice cold water. The spore suspension was heated at 65°C for 3h in order to completely eliminate vegetative cells. The concentration of spores in the preparations was quantified by determining the viable count. To do this, spore samples were serially diluted in ultra-pure water. BHI agar plates supplemented with 0.1 % (w/v) each of L-cysteine (Sigma-Aldrich, Dorset, UK) and sodium taurocholate (Sigma-Aldrich, Dorset, UK) (Buggy et al., 1985, Buggy et al., 1983). The plates were pre-reduced anaerobically at 37^{o}C for 1 h prior to use. A 20 µl volume of the diluted spore preparations were applied onto plates and spread evenly using a sterile plastic spreader. The plates were then incubated anaerobically for 24 h at 37^{o}C to allow the spores to germinate and grow. Two groups of negative controls were involved. One was the plating of 20 µl of ice cold distilled water on BHI supplemented plates and a 20 µl of the prepared spore suspension on plain BHI with no supplements (Nale et al., 2012). After 24 h incubation, the numbers of colonies per plate were counted and the CFU/ml count was calculated using the formula below:

$$\text{CFU/ML} = \frac{\text{Number of colonies formed}}{\text{volume plated}} \text{x dilution factor}$$

2.4.12 **Motility assay for *C. difficile* strains**

A motility assay was performed for *C. difficile* isolates using a Motility Agar stab (Stabler et al., 2009, Tasteyre et al., 2001). Motility Agar is soft agar in a test tube (0.175% w/v agar) bacteria are stab-inoculated into the agar. This was incubated anaerobically at 37 °C for 48 h. Non-motile bacteria will only grow where they are inoculated, while motile bacteria will grow along the stab and will also ‘’swim’’ out away from the inoculated area. The assay was performed with two technical replicates and three biological repeats (Hargreaves et al., 2013).

2.4.13 **Minimum Inhibitory Concentrations (MICs) of *C. difficile* strains**

Antibiotic susceptibility tests for vancomycin (VA), metronidazole (MZ), clindamycin (CM) and ciprofloxacin (CI) of *C. difficile* strains were carried out by using Etest strips (Biomérieux, Marcy l'Etoile, France) following the manufacturer's guidelines. The Etest based on given gradients of antibiotic densities on a plastic strip that enabled us to determine the precise MIC of anti-bacterial agents. 300 µL of overnight FA culture was used to make a lawn on 1% (w/v) agar BHI plate. Lawns were left to dry for 5 min followed by applying the antibiotic strips onto the lawn. Plates were inverted and incubated overnight. The MIC value was read at the point of intersection between the inhibition ellipse edge and the E-test strip.

Chapter 3 Isolation and characterization of *Clostridium difficile* from the natural environments in the north of Iraq

3.1 Introduction

Clostridium difficile is a Gram positive pathogenic bacterium that causes nosocomial infections in humans (Ghose, 2013). It can also be carried asymptomatically and accordingly promote the dissemination of the disease (Eyre et al., 2013, Hargreaves et al., 2015). This bacterium inhabits a variety of environments and has been isolated from water (Zidaric et al., 2010, Pasquale et al., 2011, Romano et al., 2012), soil (al Saif and Brazier, 1996, Kim et al., 1981, Simango, 2006) and sediment (Hargreaves et al., 2013) as well as in clinical settings (Kuijper et al., 2001, Geric et al., 2004, Martirosian et al., 2005, Kuijper et al., 2006, He et al., 2013, Kalchayanand et al., 2013, Gupta and Khanna, 2014, Zhou et al., 2014). *C. difficile* is associated with food borne infections (Rupnik and Songer, 2010a, Gould and Limbago, 2010, Lund and Peck, 2015) as it has been isolated from a variety of contaminated food sources (Simango and Mwakurudza, 2008, Gould and Limbago, 2010, Metcalf et al., 2010, Kalchayanand et al., 2013, Quesada-Gomez et al., 2013, Yamoudy et al., 2015); a number of *C. difficile* genotypes have been recovered from animals and the environment have been observed also in human (Weese, 2010, Janezic et al., 2012). This suggests that the natural environment is not only a source of *C. difficile* strains but favours evolution of microorganisms due to constant interaction within a population and with members of a different species (Hargreaves et al., 2015).

C. difficile is the most problematic bacterium in clinical settings in Europe and the USA, and a range of studies have focused on these locations (Geric et al., 2004, Wilcox et al., 2012, Lessa et al., 2015, Kurti et al., 2015, Zarb et al., 2012, Bauer et al., 2011, Barbut et al., 2003). However, many other countries have remained unexplored and with constant migration and associated human activities, it is not known how strain variation emanating from these geographical areas is contributing to the overall complexities associated with CDI. It is therefore of interest to explore other geographical locations in order to isolate and characterise the diverse strains in order to understand *C. difficile* biogeography and how this may relate to epidemiology (al Saif and Brazier, 1996, Wilcox and Fawley, 2000, Waslawski et al., 2013, Bauer et al., 2011).

There is great diversity within *C. difficile* population (He et al., 2010, Knight et al., 2015, Stabler et al., 2012), and different approaches have been established to characterise them. Polymerase Chain Reaction (PCR) ribotyping is the standard method worldwide which distinguishes between strains according to the size and the number of copies of the *C. difficile* 16S - 23S rRNA inter spacer region (Arroyo et al., 2005, Martin et al., 2016, Bidet et al., 1999, Brazier, 2001). However, although ribotyping is useful for outbreak studies, more sensitive approaches are required to study long-term epidemiology, and to carry out population genetic analysis. With this in mind, multisequence typing has been developed as a tool to consider the genetic relatedness of many bacterial species and population structure (Maiden et al., 1998, Maiden, 2006, Griffiths et al., 2010).

Multilocus sequence typing is a bacterial genotyping method that enables strain differentiation by utilising sequence fragments from seven housekeeping genes (*adk*, *atpA*, *dxr*, *glyA*, *recA*, *sodA* and *tpi*) (Maiden et al., 1998). Each of these genes has a unique sequence form with a unique allele number. The combination of the seven available allele numbers gives a sequence type (Lemee et al., 2004). Thus, if used together, MLST and ribotyping are dynamic tools for classifying phylogenic relationships within *C. difficile* strains. However, if more precise resolution is needed then whole genome sequencing (WGS), which permits single nucleotide-level resolution for a strain can be applied (Zhou et al., 2014, Martin et al., 2016).

Whole genome sequencing (WGS) is a highly discriminatory tool for outbreak investigations and its use has considerably increased the understanding of both pathogenesis and the dynamic populations of *C. difficile* (Wormser et al., 2009, Grad et al., 2012, Knetsch et al., 2014, Martin et al., 2016). Indeed, phylogenetic analysis of well-described *C. difficile* strains suggests that they are categorised into six unique clades (1, 2, 3, 4, 5 and C-I) (Dingle et al., 2014, Monot et al., 2015, Stabler et al., 2006). Of these groups clade 1 is a highly diverse clade in terms of alleles and ribotypes (Cairns et al., 2012, Dingle et al., 2014, Janezic and Rupnik, 2015, Stabler et al., 2012). Whilst C-I clade has a distinctive lineage, it was hypothesized that it might represent a novel species (Dingle et al., 2014), and recently strains with new toxin profiles (A^+B^-) were found belong to this clade (Monot et al., 2015)

The major contributing factors for the pathogenicity of *C. difficile* are the virulence determinants toxins A and B (Voth and Ballard, 2005, Dingle et al., 2011, Janoir, 2016), that cause the characteristic pathology of CDI (Poxton et al., 2001). In addition, other non-toxin virulence factors are also associated with *C. difficile* pathogenesis, including colonisation and sporulation factors (Vedantam et al., 2012, Janoir, 2016). Flagella are also acknowledged as important virulence factors that mediate motility in several enteropathogenic species (O'Neil and Marquis, 2006, Stabler et al., 2006). Recently, studies have unravelled the role of flagella in *C. difficile* pathogenicity, and shown that they are involved in the regulation of toxin gene expression species (Baban et al., 2013, Stevenson et al., 2015) however, other work has shown that the capability of non-flagellated and flagellated strains to cause CDI is similar (Goorhuis et al., 2008a).

The onset of CDI is linked with broad spectrum antibiotic use, including clindamycin and ciprofloxacin (Gerding, 2004), therefore, resistant strains to these antibiotics are of interest. In addition, vancomycin and metronidazole are the first line of treatment for the disease, although infection relapse has been reported for both drugs (Alonso et al., 2001, Marra and Ng, 2015). Studies have emphasized the large role of environmental strains in the distribution of antibiotic resistance genes (Berglund, 2015). Mobile genetic elements play a role in dissemination of antibiotic resistance genes in *C. difficile* strains, and compromise 11% of the genome (Sebaihia et al., 2006, Mullany et al., 2015).

Mobile genetic elements are any piece of DNA that can move around within the genome (Siefert, 2009). *C. difficile* in particular has an extremely mobile, mosaic genome; thus, mobile elements exert a great influence on the properties of a host and may result in divergence as a strain evolves increasing virulence or antibiotic resistance (Sebaihia et al., 2006). Bacteriophages are the mobile genetic elements that facilitate horizontal gene transfer (HGT), and also impact the evolution of the host genome (Siefert, 2009). Prophages are encoded in the genomes of most sequenced *C. difficile* strains (Shan *et al*., 2012). Prophages with highly diverse morphology and genome size have been detected in environmental and clinical strains belonging to different ribotype groups (Nale et al., 2012, Hargreaves et al., 2013). At least one prophage has been identified in all of the sequenced *C. difficile* isolates, whilst some strains have been found to encode two prophage (Goh et al., 2005b, Hargreaves et al., 2013, Sebaihia et al., 2006, Shan et al., 2012).

One way for *C. difficile* to combat phage infection is by having Clustered Regularly Interspaced short Palindromic Repeats (CRISPR) and an associated Cas proteins system (Bhaya et al., 2011, Hargreaves et al., 2014a, Boudry et al., 2015). CRISPRs are an adaptive bacterial and archaeal "immune system" that directly target infecting DNA. This system is composed of arrays of short (25–50) bp direct repeats (DRs) split by variable sequence spacers of similar length (26 to 72) along with Cas proteins. These Cas directly target and degrade foreign DNA homologous to a spacer. As spacers are inherited, they can also be acquired through the integration of foreign DNA sequences (Barrangou et al., 2007). Because different strains of *C. difficile* encode multiple CRISPR arrays (Sebaihia et al., 2006, Hargreaves et al., 2014a), the dispersal and diversity of these spacers in the bacteria and its prophage could control phage infections in this pathogen and influence both its evolution and pathogenicity (Hargreaves et al., 2014a).

Although there are growing numbers of studies which have investigated *C. difficile* strains in Western countries in both clinical and environmental settings, little has been published on strains of this pathogen from the Middle East (Al-Thani et al., 2014, Jalali et al., 2012, Khoshdel et al., 2015, Rotimi et al., 2003, Sadeghifard et al., 2010a, Sadeghifard et al., 2010b, Al-Tawfiq and Abed, 2010). Recent studies have reported the isolation of *C. difficile* from non-clinical samples such as wastewater treatment plants and ready food salad in Iran (Nikaeen et al., 2015, Rahimi et al., 2015). Despite the extensive studies on the clinical isolates, no reports currently exist on the isolation and characterisation of environmental *C. difficile* strains and their distribution in this geographical area. In order to address this shortcoming, this study was designed to isolate and characterise *C. difficile* strains in Kurdistan in the north of Iraq and I state the aims of this study out below.

3.2 Aims and objectives of the study

1. Isolate strains of *C. difficile* from the natural environment in Kurdistan Northern Iraq (from water and soil collected from rivers and mountains).
2. Characterise *C. difficile* isolates using PCR ribotyping and determine their toxin profiles.
3. Sequence the genome of selected isolates and identify multilocus sequence type (MLST) within the genome of the strains.

4. Analyse genomes for content and determine phylogenetic relationships with known strains.
5. Examine the diversity of the CRISPR/Cas system in selected strains from their genomes and correlate this diversity to the susceptibility of strains to phage infection.
6. Characterise the strains according to their motility and sensitivity to antibiotics commonly associated with CDI (vancomycin, metronidazole, clindamycin and ciprofloxacin).
7. Isolate and characterise inducible prophages from the strains.

3.3 Results

3.3.1 Isolation of *C. difficile* from the environment

C. difficile was isolated from 13 out of 45 subsurface soil and sediment samples collected from 15 sample sites; of these, seven sites yielded 61 strains (Figure 3-1, Table 3-1 and Table 3-2). The bacterium was recovered from both polluted and clean sites. For example strains of *C. difficile* were isolated from Hamamoke resort (site 7) where human activities were highly unlikely and from Dokan resort (site 9) where human habitation is likely. In addition, the type of sample does not appear to make a difference, as *C. difficile* has been isolated from both although high percentages (75%) of isolates were recovered from sediment samples and just 25% from soils Table 3-2.

Table 3-1. Table of sampling sites, sample type and ribotype.
Numbers of sites are corresponding to the map in Figure 3-1.NA= not assigned ribotype.

Sites	Source	Notes	Ribotype
1	Dalal Bridge	Sediment	-ve
2	Duhok Valley	Sediment	-ve
3	Safien Mountain	Soil	-ve
4	Kalak River	Sediment	R014/NA
5	Chnarok Mountain	Soil	R091
6	Hizope River	Sediment	-ve
7	Hamamok Resort	Soil/Sediment	R010/R604
8	Bestana	Soil/Sediment	R604
9	Dokan Resort	Sediment	R001/R091/R011
10	Khalakan	Soil	-ve
11	Kalkasmaq	Soil	-ve
12	Taq Taq River	Soil/Sediment	R691/NA
13	Jalee River	Soil/Sediment	R011/R035/ R001/R081
14	Haibat sultan Mountain	Soil	-ve
15	Sirwan River	Sediment	-ve

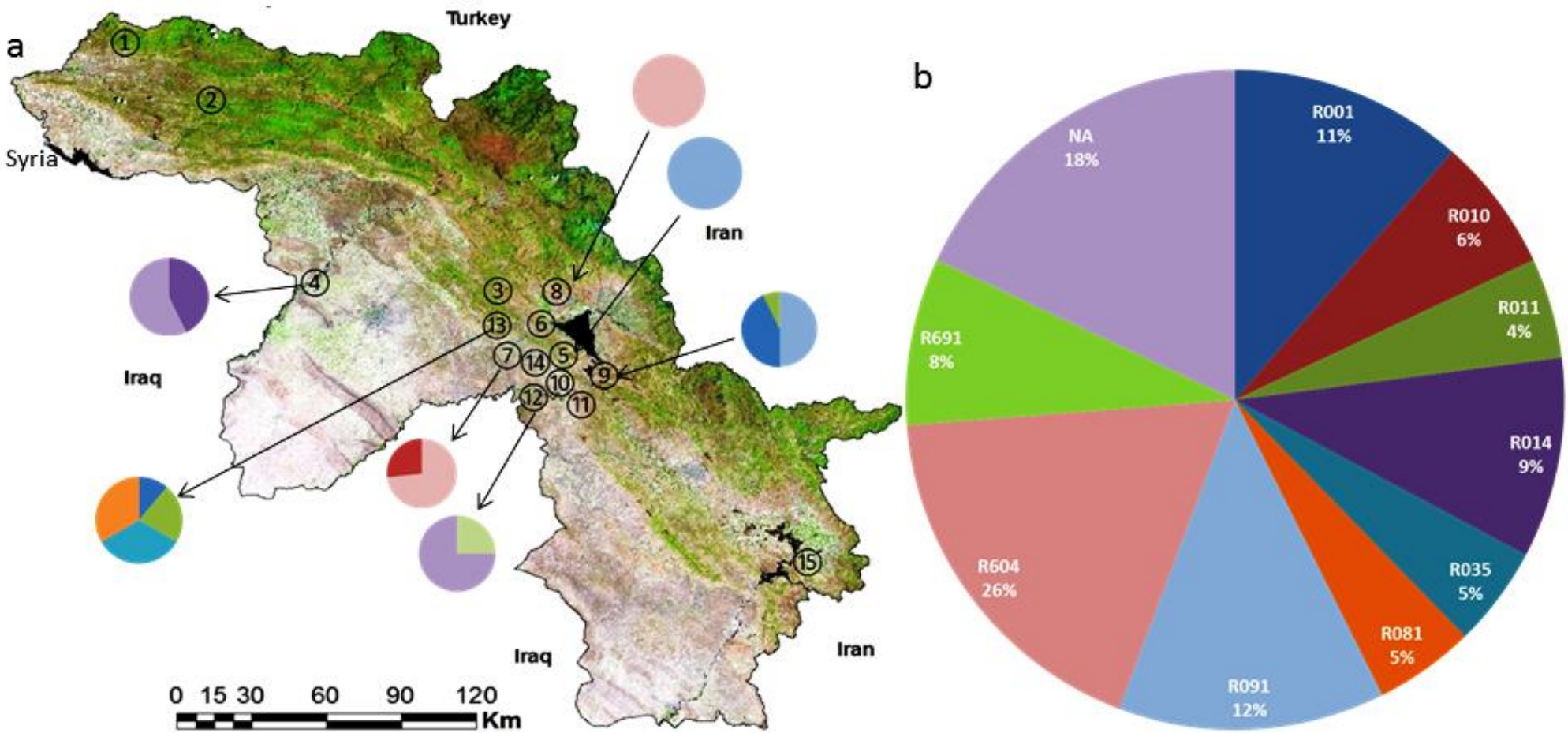


Figure 3-1. Attendance and ribotype diversity of *C. difficile* across Kurdistan region the North of Iraq.
a: Map of sampling sites in Kurdistan the North of Iraq, sampling sites are numbered to correlate to the Table 3-1. **b**: Pie chart showing the proportion of ribotypes recovered from the soil and sediment samples. Abbreviation NA= not assigned to ribotype.

Table 3-2. Sampling site, *C. difficile* isolate ribotypes, origin, toxin profiles and motility.

Abbreviation, R= ribotype, NA= not assigned ribotype, +++; high level of motility, ++; medium level of motility and +: low motility.

Site	Ribotype	Source	Toxin profile	Motility	Site	Ribotype	Source	Toxin profile	Motility
9	091	Sediment	A-B-	++	12	691	Sediment	A-B+	+++
9	091	Sediment	A-B-	+++	4	014	Soil	A-B+	++
9	091	Sediment	A-B-	++ +	4	014	Sediment	A+B+	+++
9	091	Sediment	A-B-	++	4	014	Sediment	A+B+	+++
9	091	Sediment	A-B-	+++	4	014	Sediment	A+B+	+++
9	091	Sediment	A-B-	++	4	014	Sediment	A+B+	+++
9	091	Sediment	A-B-	+++	4	014	Sediment	A+B+	++
9	001	Sediment	A+B+	++	4	NA	Sediment	A-B-	++
9	001	Sediment	A+B+	+++	4	NA	Sediment	A-B-	+
9	001	Sediment	A+B+	++	4	NA	Sediment	A-B-	+++
9	001	Sediment	A+B+	++	4	NA	Sediment	A-B-	+++
9	001	Sediment	A+B+	++	4	NA	Sediment	A-B-	+++
9	011	Sediment	A+B+	++	4	NA	Sediment	A-B-	++
13	001	Sediment	A+B+	+++	4	NA	Sediment	A-B-	+++
13	011	Sediment	A+B+	++	4	NA	Sediment	A-B-	+++
13	011	Sediment	A-B-	++	12	691	Soil	A-B-	++
13	035	Sediment	A-B-	++	12	691	Sediment	A-B-	++
13	035	Sediment	A-B-	++	12	691	Sediment	A-B+	++
13	035	Sediment	A-B-	++	12	691	Sediment	A-B+	++
7	604	Soil	A-B-	++	13	081	Soil	A-B+	++
7	604	Soil	A-B-	+	13	081	Soil	A-B+	+
7	604	Soil	A-B-	+	13	081	Soil	A-B+	++
7	604	Soil	A-B-	+	12	NA	Sediment	A-B+	++
7	604	Soil	A-B-	+	12	NA	Sediment	A-B+	++
7	604	Soil	A-B-	+++	12	NA	Sediment	A-B+	+++
7	604	Soil	A-B-	++	9	001	Sediment	A+B+	++
7	604	Soil	A-B-	++					
8	604	Soil	A-B-	++					
8	604	Sediment	A-B-	++					
8	604	Sediment	A-B-	++					
5	091	Soil	A-B-	++					
7	010	Sediment	A-B-	+					
7	010	Sediment	A-B-	+					
7	010	Sediment	A-B-	+++					
7	010	Sediment	A-B-	+++					

3.3.2 **Diverse and novel strains are found in the natural environment.**

Ribotyping analysis reveals that the *C. difficile* strains in natural environments are mixed and varied with nine known ribotypes identified as well as a novel form termed R691. In addition, 11 isolates from 2 samples sites could not be assigned to known ribotypes, and were thus grouped into two based on their ribotype profiles. Of the nine known ribotypes identified, seven are common in the UK and Europe (R001, R010, R011, R014, R081 & R035) based on the profile/banding pattern on the gel (Figure 3-2). While the other two (R091 and R604) are unusual even in this area alongside the one novel ribotype (R691) which has not identified before (Table 3-1). The distribution of PCR ribotypes also differed at each sample site, with two ribotypes R001 and R011 isolated at two sample sites (site 9 and site 13), and R091 isolated at two others (site 9 and site 5) (Figure 3-1).

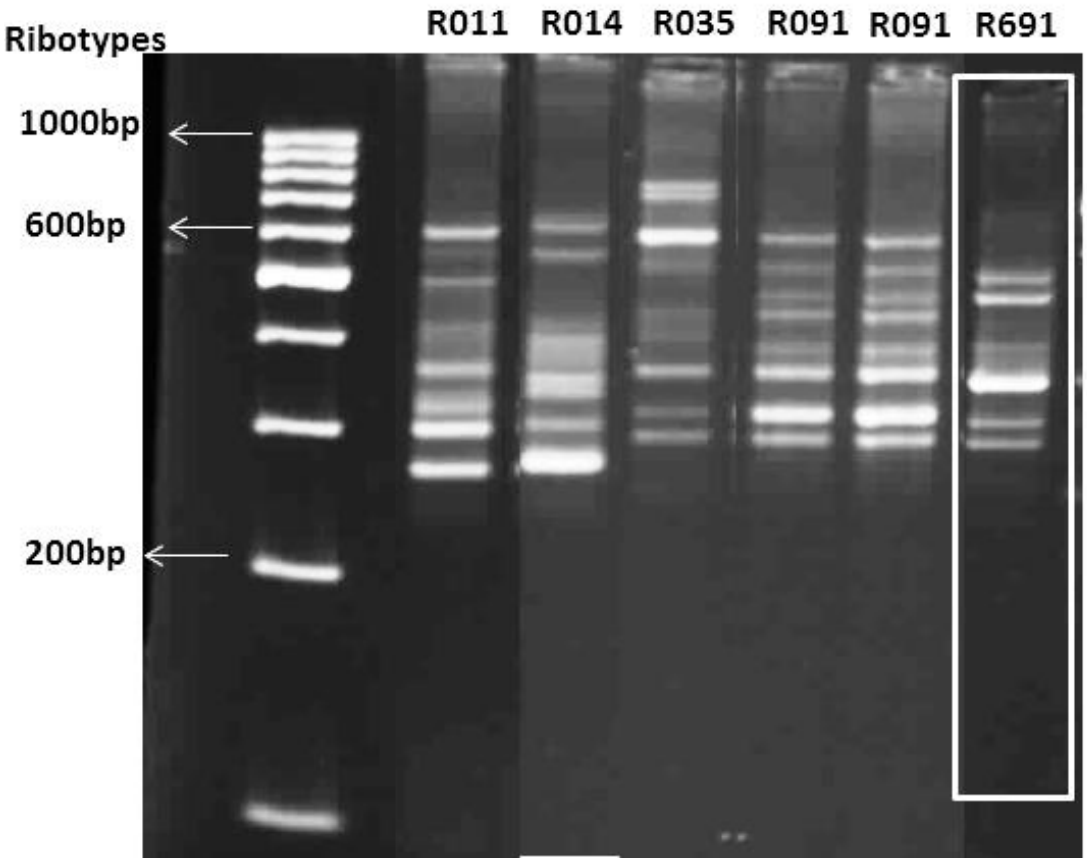


Figure 3-2. Gel electrophoresis images of *C. difficile* PCR ribotype profiles

Representation of five ribotype patterns of *C. difficile* strains isolated from environmental samples. The ribotype profile in the white box shows is the novel ribotype.

3.3.3 **Toxin profile determination of *C. difficile* isolates**

The toxin profiles of isolates were determined by examining toxin-related genes including *tcdA*, *tcdB*, and the binary genes *cdtA* and *cdtB*. Strains with different toxin profiles were found within the area, and isolates were grouped into three categories based on the presence or absence of toxin genes; non-toxigenic strains negative for the targeted regions *tcdA* and *tcdB* (A^- B^-) comprised 55.7% of the sample, toxigenic isolates positive for *tcdA* and *tcdB* (A^+ B^+) comprised 23% of the sample, and toxigenic isolates negative for *tcdA* and positive for *tcdB* (A^- B^+), comprised 21.3% of the sample (Figure 3-3 and Figure 3-4). Notably, the two isolates of R011 came from different sample sites (site 2 and site 13), and had different toxin profiles, one negative for *tcdA* and *tcdB*, and one positive for both genes. In addition, the tested isolates were negative for the binary toxin genes (CDT^-).

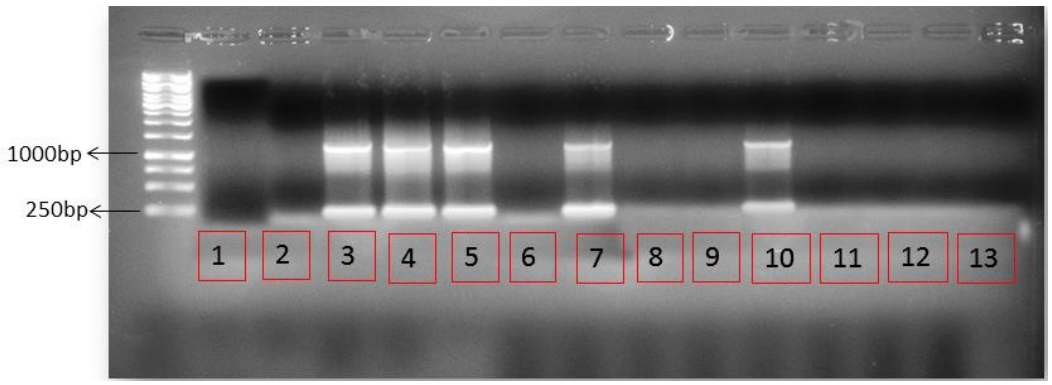


Figure 3-3. Agros gel electrophoresis for Toxin A gene screening in *C. difficile* isolates.

PCR based toxin A gene screening for 13 strains of *C. difficile* isolates. Two primer sets were used to amplify the partial sequences of *tcdA* producing 251 bp and an essential repeat region producing product 1,265bp. The gel shows 5/13 selected strains were positive for both region of toxin A gene.

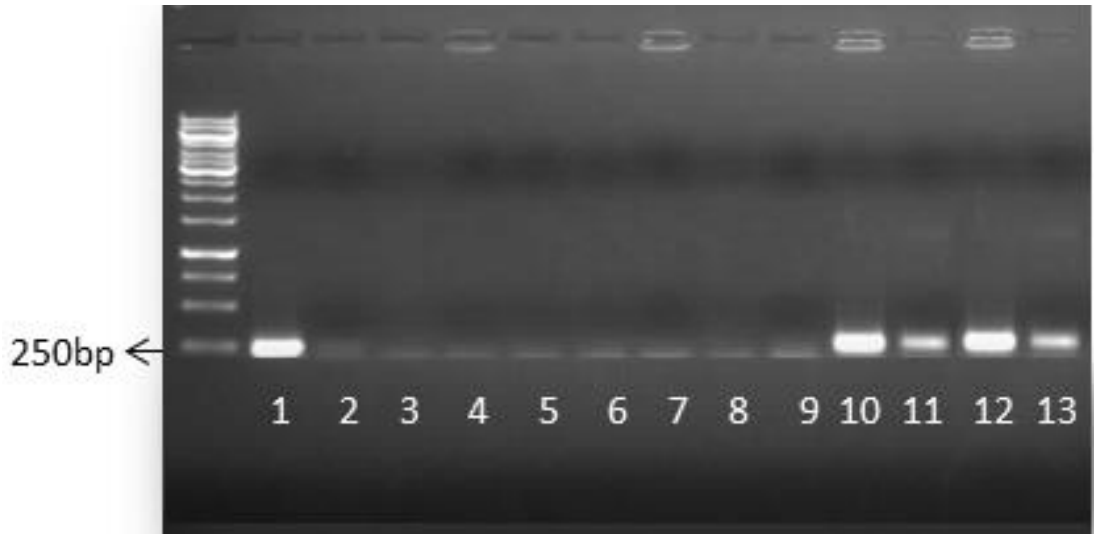


Figure 3-4. Agrose gel electrophoresis for toxin B gene screening in *C. difficile* isolates.
PCR based toxin B gene screening for 13 strains of *C. difficile* isolates. One set of primers was used to target *tcdB* gene producing 203bp. 5/13 of tested strains are encode toxin B gene.

3.3.4 **Motility assays of *C. difficile* isolate**

Most bacteria are motile which is important for survival as it enables movement toward nutrient-rich sources and movement away from noxious environments (Stabler et al., 2009). Thus, to further describe isolated strains, motility traits were considered. The motility assays revealed that all tested isolates were motile but to different degrees. Thus, isolates were described in one of three categories; 34.4% of the strains have a high level of motility (+++), while 54.1% have medium motility (++) and 11.5% have low motility (Figure 3-5 and Table 3-2).

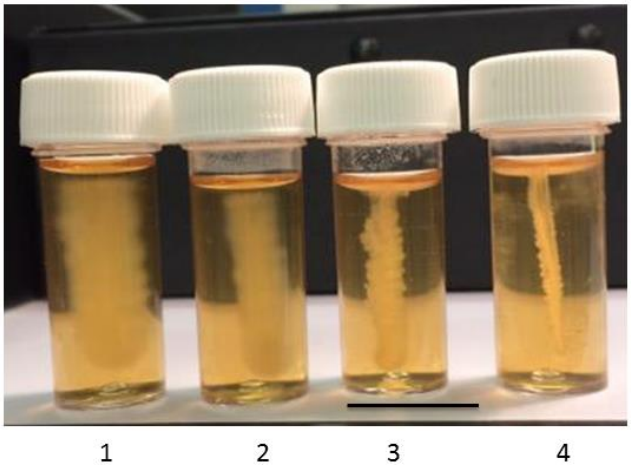


Figure 3-5. Result of motility assay for *C. difficile* isolates.

Strains were stabbed in BHI 0.175% (w/v) agar medium and incubated for 24 h in anaerobic chamber. Motility was clearly visible in agar stabs for all strains except for the negative control strain M120. Lane 1: high level of motility +++, lane 2 medium level of motility ++, lane 3 low motility + and lane 4 –ve control (strain M120). Bar represents 2cm.

3.3.5 Antimicrobial susceptibility of *C. difficile* strains

Susceptibility to the antibiotics vancomycin, metronidazole, clindamycin, and ciprofloxacin were determined for 20 isolates (two strains from each ribotype were tested, except for R091 and R001 for which three strains were tested) (Figure 3-6 and Table 3-3). These displayed antimicrobial susceptibilities ranging from being sensitive, borderline sensitivity, all the way to resistance. All isolates were sensitive to Metronidazole (MIC < 8 μg ml^{-1}) and vancomycin (MIC < 4 μg ml^{-1}), and the greatest range in MIC was found for clindamycin and ciprofloxacin. MIC variability was also observed between different strains within a ribotype; for example, five isolates of R091 include some that are resistant, some borderline resistant, and some sensitive to clindamycin and R001 also contains strains which are both sensitive and borderline resistant to this antibiotic. Both strains of R011 showed resistance to clindamycin, while a wide range of sensitivities were determined against ciprofloxacin that range from sensitive at MICs 0.5 μg/ml to fully resistant at MICs >32 μg/ml. Four isolates were observed to have MICs > 8 μg/ml.

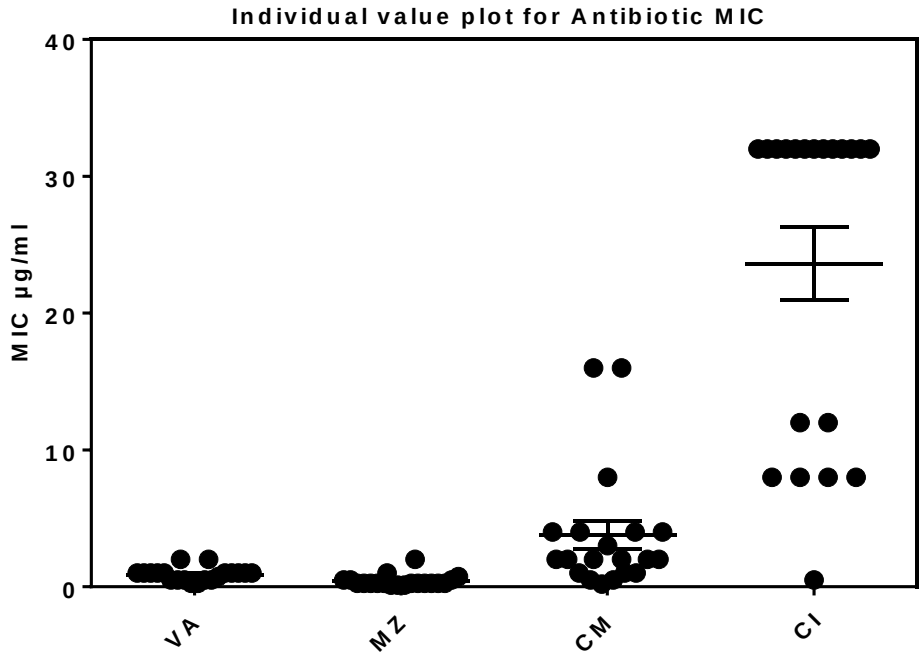


Figure 3-6. Individual value plot of MICs.

MICs (on the y axis) for each antibiotic (on x axis) of 20 isolates plotted in an individual value plot and error bars calculated using GraphPad Prism6. Abbreviations: VA, vancomycin; MZ, metronidazole; CM, clindamycin; CI, ciprofloxacin. The breakpoints according to the SRGA (http://www.srga.org) used were as follows: vancomycin susceptible, ≤ 4.0 mg/litre; vancomycin resistant, > 4.0 mg/litre; metronidazole susceptible, ≤ 8.0 mg/litre; metronidazole resistant, >8 .0 mg/litre; clindamycin susceptible, ≤ 2.0 mg/litre; ciprofloxacin resistant > 2.0 mg/litre.

Table 3-3. E-Test results for isolated strains from the environment.

		MIC in μg^{-1}			
Ribotype	**Isolate**	**Ciprofloxacin**	**Vancomycin**	**Metronidazole**	**Clindamycin**
R091	DF1	0.5	0.25	8	32
	DF9	0.25	0.5	4	32
	DF6	1	0.25	2	32
R001	J3	1	0.25	4	32
	F6	1	1	1	32
	F8	1	0.25	4	32
R011	J5	0.5	0.25	16	32
	F9	0.5	0.50	16	32
R035	J6	0.25	2	2	32
	J10	1	0.25	2	32
R604	K2	2	0.25	1	32
	K5	1	0.13	2	32
R010	4M	1	0.25	4	32
	7M	1	0.125	2	8
R691	C	2	0.25	0.5	12
	12C	0.75	0.5	1	12
R014	K4	0.5	0.25	0.2	8
	T	1	0.75	3	8
R081	11E	1	0.25	0.5	0.5
	11I	0.5	0.125	2	8

3.3.6 **Whole genome sequencing and phylogenetic analysis of *C. difficile* isolates**

To further examine the diversity within our *C. difficile* isolates, genome of 14 strains were selected and sequenced using Illumina MiSeq with 2 x 250 bp paired-ends. A 16S rRNA gene phylogeny tree was constructed that was based on 16S genes sequences to determine their position in the tree. The analysis revealed that three strains C105KSO13, C105KSO14, and C105KSO15 were grouped within Lachnospriaceae, whereas the other grouped with other difficile strains within the Peptostreptococococaceae. The three strains were subjected to further analysis including the 40 single copy-phylogenetic makers using specI (Mende et al., 2013), accordingly the three strains were assigned as a novel species of *Clostridium* (Rashid et al., 2016).

From the sequence data the 11 strains of *C. difficile* were subjected to MLST analysis and 6 sequence types (ST) were determined and correlated to PCR ribotypes (Figure 3-7). This consists of ST-3, ST-15, ST-137, ST-107, ST-177 & 181. Of these two PCR ribotypes (R035 and R091) were associated with single ST (ST-137), and two STs (ST-177 and ST-181) were occurred in a single PCR ribotype 604.

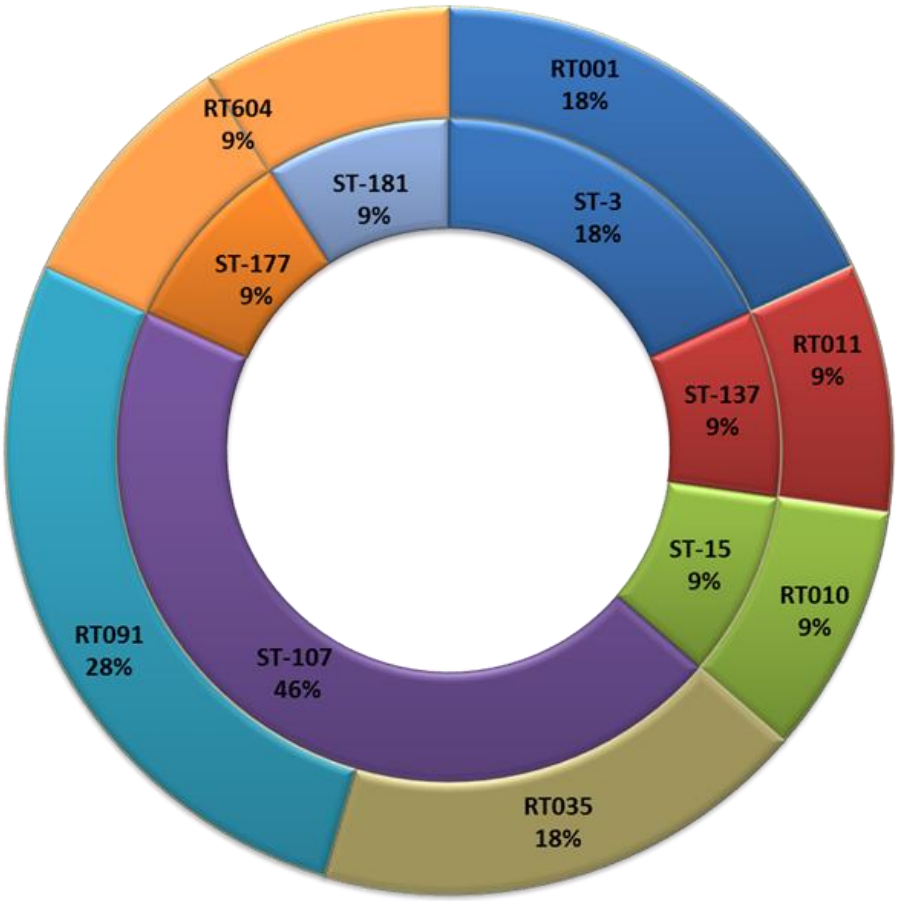


Figure 3-7. Doughnut Chart showing proportion of ribotypes and sequence types of the 11 isolates.

Outer layer of the chart showing the proportion of ribotypes, three isolates are R091; two isolates are R035, R001 & R604 and R011 and R010 are one isolate. The inner layer showing the proportion of sequence types of isolates. Five isolates are ST107 and two isolates are ST3, ST15, ST137; ST177 and ST181 are one isolate.

To explore genetic relationship among the isolates, single copy core genes of the 11new strains and 9 reference strains were identified Table 3-4. The reference strains were chosen because they are well characterized and are from a relatively diverse collection (He et al., 2010). The tree shows the genetic diversity of *C. difficile* strains in this collection Figure 3-8**.** In brief, Clade 1, 2, 4 & 5 are formed as described previously (Griffiths et al., 2010), in addition, the newly described clade (C-I), contained non-toxigenic strains only. The strains within clade C-1 are highly diverted from other non-toxigenic strains. The rest of the strains were located in clade 1Figure 3-8.

Table 3-4. *C. difficile* isolates and the references used in the phylogenetic analysis

Isolates	Country	Source	Ribotype	ST	Clade
CD630	Switzerland	Human	012	54	1
M680	Ireland	Human	017	37	4
CF5	Belgium	Human	017	37	4
M120	UK	Human	078	11	5
BI-9	USA	Human	001	3	1
R20291	UK	Human	027	1	2
CD196	France	Human	027	1	2
BI-1	USA	Human	027	1	2
CD2007855	USA	Bovine	027	1	2
CD105KSE1	Iraq	Sediment	091	107	1
CD105KSE2	Iraq	Sediment	091	107	1
CD105KSE3	Iraq	Sediment	001	3	1
CD105KSE4	Iraq	Sediment	001	3	1
CD105KSE5	Iraq	Sediment	035	107	1
CD105KSE6	Iraq	Sediment	011	137	1
CD105KSO7	Iraq	Soil	604	177	C-I
CD105KSO8	Iraq	Soil	604	181	C-I
CD105KSE9	Iraq	Sediment	010	15	1
CD105KSO10	Iraq	Soil	091	107	1
CD105KSE11	Iraq	Sediment	035	107	1

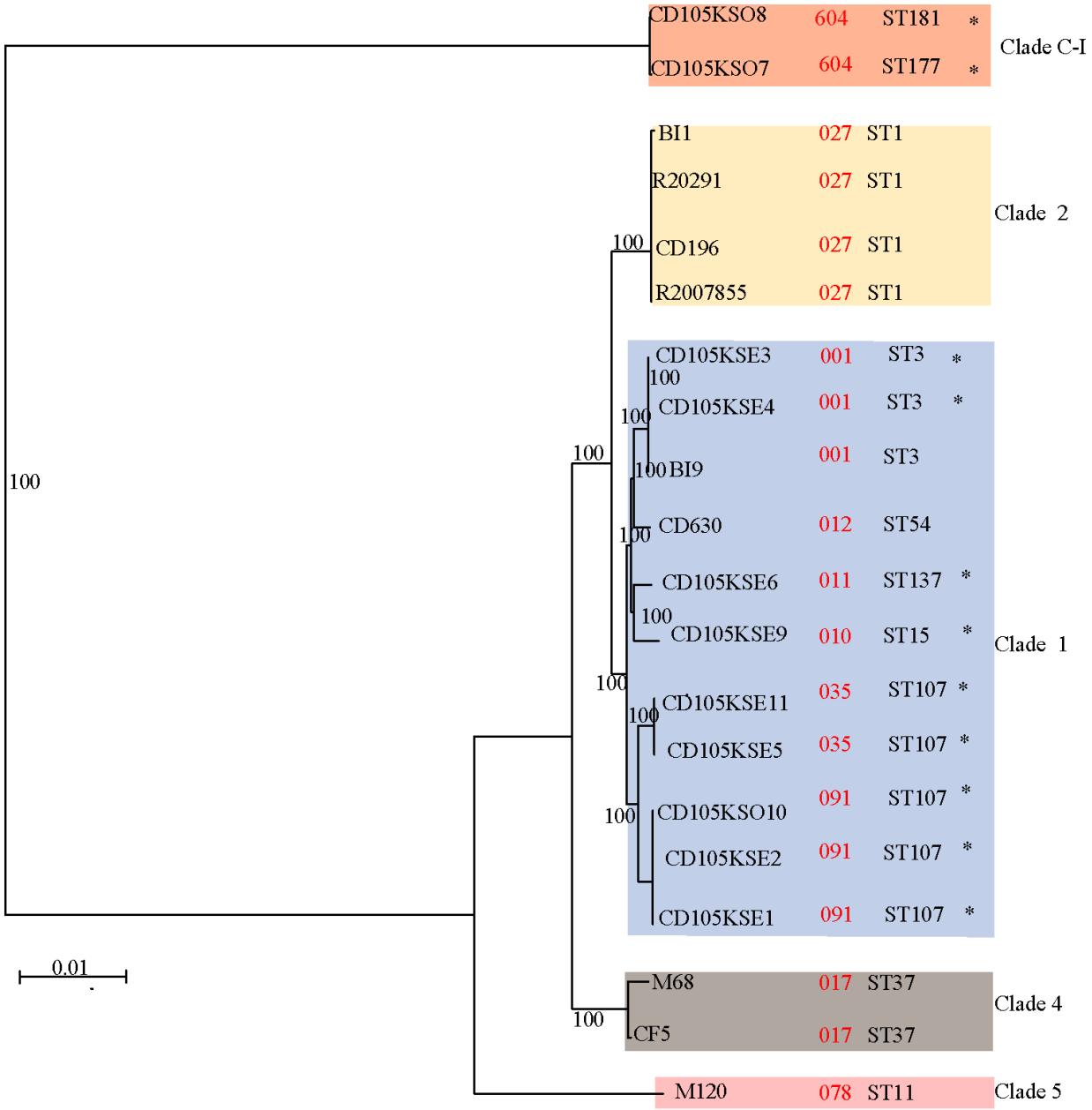


Figure 3-8. Phylogenetic tree of *C. difficile* isolates based on whole genome.

The Neighbour joining tree was constructed based on the core gene sets of 11 isolates and 9 reference stains. The ST and ribotypes (red numbers) are added after strain labels. All the genes concatenated into a single alignment prior using RaxML to carry out the construction of tree with 100 bootstrap replicate. The new isolated strains are labelled with the star.

3.3.7 **Whole genome comparison of the isolates**

Sequence comparisons were performed for clade 1 isolates against the reference genome CD630 (Sebaihia et al., 2006), and sequence similarity is shown by the intensity of shading. Several genes have been annotated and consist of regulatory genes, flagellin- related genes, motility and other virulence associated genes (Appendix 7). Obviously, a number of gene deletions occurred in all isolates compared to CD630, and are found where transposons and prophages are localized in CD630 Figure 3-9.

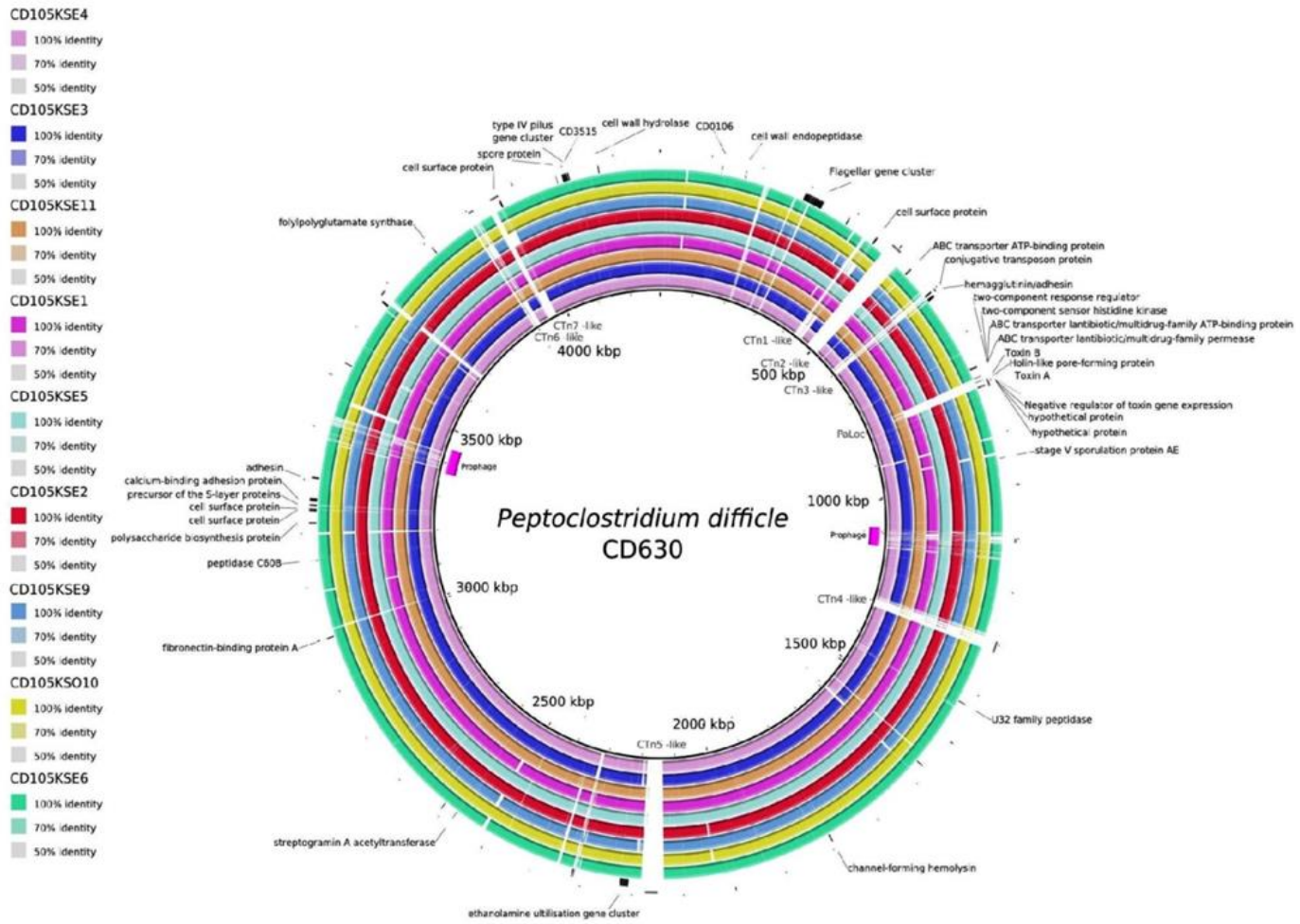


Figure 3-9. Whole genome comparisons between *C. difficile* strains of clade 1 and the reference strain CD630.

Multiple genome comparison figure generated using BRIG from performing a blastn analysis for each of the 9 isolates to reference genome CD630. On rings from the inside to the outside tracks are; CD105KSE4 (R001), CD105KS3 (R001), CD105KSE11 (R035), CD105KSE1 (R091), CD105KSE5 (R035), CD105KSE2 (R091), CD105KSE9 (R010), CD105KS010 (R091), CD105KSE6 (R011). Each genome has a unique colour with varying colour gradient according to the percent identity with the reference. The genome used as the reference (CD630) is indicated at the centre of the ring against the rest of the genomes, and gene features of interest from CD630 annotation are labelled. Gapped regions indicate the absence or low similarity between the genomes. Seven conjugative transposons are marked from the CD630 annotation and the two prophages which are highly divergence across the analysis.

3.3.8 **Multiple prophage carriage in *C. difficile* isolates**

Because *C. difficile* is known to be a highly lysogenic species, the genome of the 11 strains was examined using PhiSpy to determine the presence of lysogeny (Akhter et al., 2012). Since multiple regions in single strains have been predicted to harbour phage sequences, these regions were checked manually, and sequences similar to known *C. difficile* phages were observed throughout each genome. Each strain contains multiple prophage sequences that range from 2-6 prophages, and remarkably 6 phage sequences were observed in one isolate R604 that located in clade C-I, while three phage sequences were found in the genome of another isolate from the same ribotype (Table 3-5). Prophage carriage in isolates of clade one ranged from 2-3 prophages, R091 and R011 isolates encoded 3 prophages, and 2 prophages were found in the R010 strain. Dual prophage carriage was identified in one strain, R091, and the R001 and R035 strains each have isolates that encode between two and three prophages. Additionally, the genomes of R-type bacteriocins also known as phage tail-like particles (PTLPs) were detected in all isolates (Sangster et al., 2015, Hegarty et al., 2016).

Table 3-5. Lysogeny of the strains according to the morphology of phage particles as observed using TEM and from the genome.

Strains	RTs	Observed morphology by TEM: MMs	LTMs	SMVs	SVs	PTLP	Genome analysis
CD105KSE1	091	√M	–	–	–	√	3prophages + PTLP
CD105KSE2	091	–	–	–	–	√	3prophages + PTLP
CD105KSE3	001	–	–	√M	–	√	3prophages + PTLP
CD105KSE4	001	–	–	–	–	√	2prophages + PTLP
CD105KSE5	035	√N	√N	–	–	√	3prophages + PTLP
CD105KSE6	011	√N	–	–	–	√	3prophages + PTLP
CD105KSO7	604	–	√N	–	–	√	6prophages + PTLP
CD105KSO8	604	√N	–	–	–	√	3prophages + PTLP
CD105KSE9	010	√M, N	√M	–	–	√	2prophages + PTLP
CD105KSO10	091	–	–	–	√N	√	3prophages + PTLP
CD105KSE11	035	–	–	–	–	√	2prophages + PTLP

Abberviations: N/A = not assignable, MMs = medium myoviruses, LTMs = long tailed myoviruses, SMV = small myoviruse, SV=siphoviruse, PTLPs = phage tail-like particles, symbols "-" = negative, "√"= positive, M=Mitomycin and N= Norfloxacin.

3.3.9 **Temperate phage carriage in *C. difficile* isolates from the North of Iraq**

C. difficile isolates were induced with mitomycin C or norfloxacin to release prophages as genomic examination of strains indicated multiple prophage carriage. Thus, an overnight culture of the 11 strains in BHI was induced with antibiotic at a concentration of 3µg ml^{-1} for 24 hours. The result of the examination of induced lysates using TEM is comparable to the genome examination of strain R010, as both reveal the presence of two phage particles in addition to phage-tail-like particles (Figure 3-10 and Table 3-5). However, TEM results showed a limited number of prophages present in the other strains.

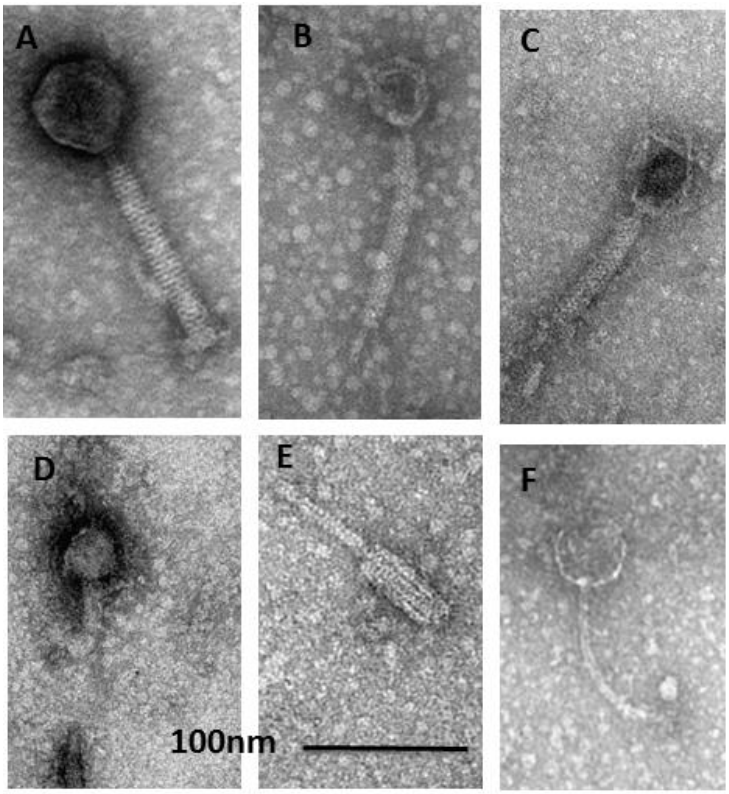


Figure 3-10 Morphological groups of temperate phages.

Morphology of phage particles observed from induced lysates either with norfloxacin or mitomycin C. Typical myoviruses (morphology A) was induced from five isolates belonging to five ribotype (R091, R035, R011, R604 and R010) alongside with defective myoviruses morphologies are C (induced from R010), E (induced from R011) An additional morphology of defective myovirus, morphology B Long tail myoviruse was induced from isolates of R035, R604 and R010, and morphology D small myoviruse was induced in isolate of R001. Typical siphoviruse (morphology F) was induced from one isolate belonging to RT091. Bar represent ~100 nm.

3.3.10 **Phage-tail-like particle**

R type bacteriocin or Phage-tail-like particles were observed in the tested lysates, Gebhart and the collegoues termed these particles as diffocins and showed that have bactericidal activity against other *C. difficile* isolates (Yamamoto et al., 1970). Two different kinds of these particles were recognized in the lysates of (Figure 3-11). Of these, one type (morphology H) has particles which resemble genuine phage-tails, are ~160 nm by 20 nm, and terminate in horizontal protrusions. These morphology H phage-tail-like particles were induced from 36% of the isolates as well as other morphologies, while tail particles (morphology G and I) were observed in 55% of induced lysates. Morphology G tail particles have tightly coiled spirals, approximately 20 nm in diameter, and are between 80-350 nm in length (Nale et al., 2012).

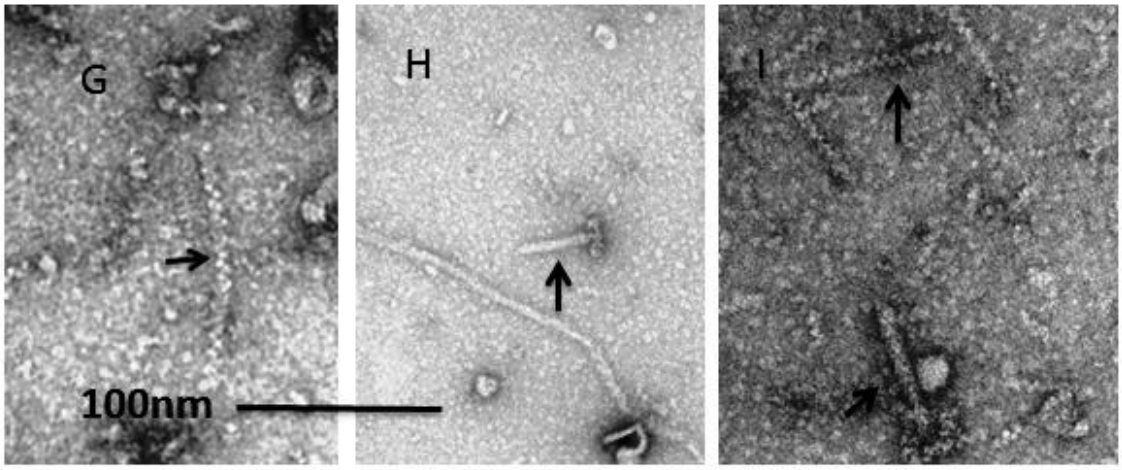


Figure 3-11. Phage tail-like particles observed from the induced lysates.

The phage tail-like particles G and H above were observed from the lysates of all the cultures along with other morphologies, I showing both morphologies G and H. Bar represent ~100 nm.

3.3.11 **Effect of antibiotics on the growth of the bacterial cultures**

To further understand if culture density is a sign of phage release, a growth of 11 isolates were monitored over 24 hours by measuring OD $_{550}$ nm before and after treatment with mitomycin C and norfloxacin, and then compared to non-treated controls (Figure 3-12). Data show decreased OD $_{550}$ nm in all the antibiotic treated cultures and in 73% of untreated cultures when compared to starting concentrations. Decreased OD $_{550}$ nm values were more frequent in inductions that used norfloxcin 82% than in those treated with mitomycin C 9%, while 9% of lysates induced with both antibiotics had the

same OD_{550}nm values. These data suggest that antibiotic treatment might not be a good indicator of phage release in this bacterium.

Furthermore, the induction efficiency of the two antibiotics were compared, and it was found that one isolate (9%) can be induced by either mitomycin C or norfloxcin , while 63.6% of the isolates were induced by the norfloxacin. However, mitomycin C induced 36.4% of the isolates.

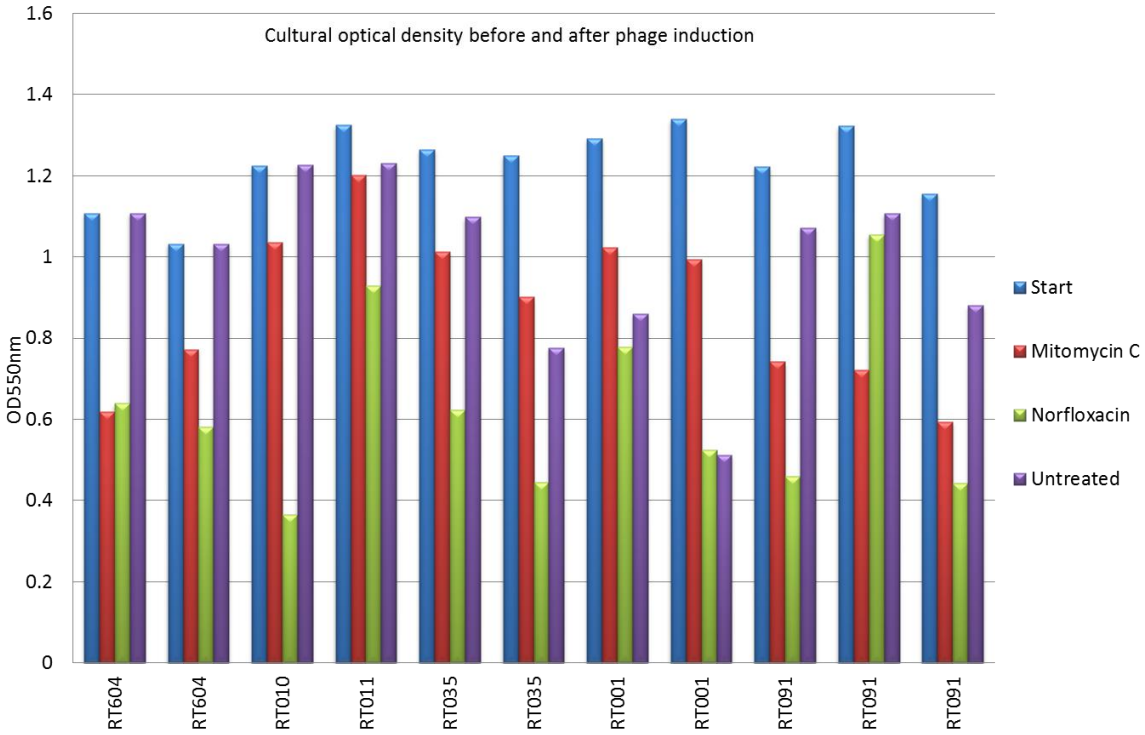


Figure 3-12. Optical density of *C. difficile* culture prior and after phage induction.

Cell density of the 11 strains was measured after overnight growth in BHI. The culture was subjected to phage release by treating with either mitomycin C (MC) or norfloxacin (NFX) for a period of 24 hour. The experiment repeated twice with three replicates.

3.3.12 *C. difficile* isolates encode multiple CRISPR arrays

Previous work has shown that the CRISPR system is active and functional in *C. difficile* (Boudry et al., 2015, Hargreaves et al., 2014a). Thus, to explore how diverse this system is in the isolates in this study, the genomes of 11 strains were examined using PILERC-CR (Edgar, 2007). Results show that all *C. difficile* isolates encode multiple CRISPR arrays, ranging from 3 to 12, and that they have variable spacer contents, ranging from 45 to 112 per strain (Figure 3-13 and Appendix 8). Isolates of

different ribotypes have different CRISPR contents, although differences between the same ribotypes were of the order of one or two arrays. Interestingly, both isolates of R604 have three CRISPR arrays.

C. difficile CRISPR spacers (n= 1054) of the 11 strains were examined for the sequence similarity using CRISPR target, 67 spacers were identified that are homologues to phage and plasmid sequences (Figure 3-14 and Appendix 9). Multiple strains have spacers that are homologous to the same phage, for example 8/11 strains contain spacers that match sequences in the genomes of phiCDHM3 and phiCD146. Spacers were also observed within the same CRISPR array target in the same phage, for example three multiple spacers in the R091 isolate match those in phiMMP02 (Figure 3-14), and strains of similar ribotypes have almost identical spacers (for example, strains of R035 have a similar spacer that is homologous to the CD630 plasmid).

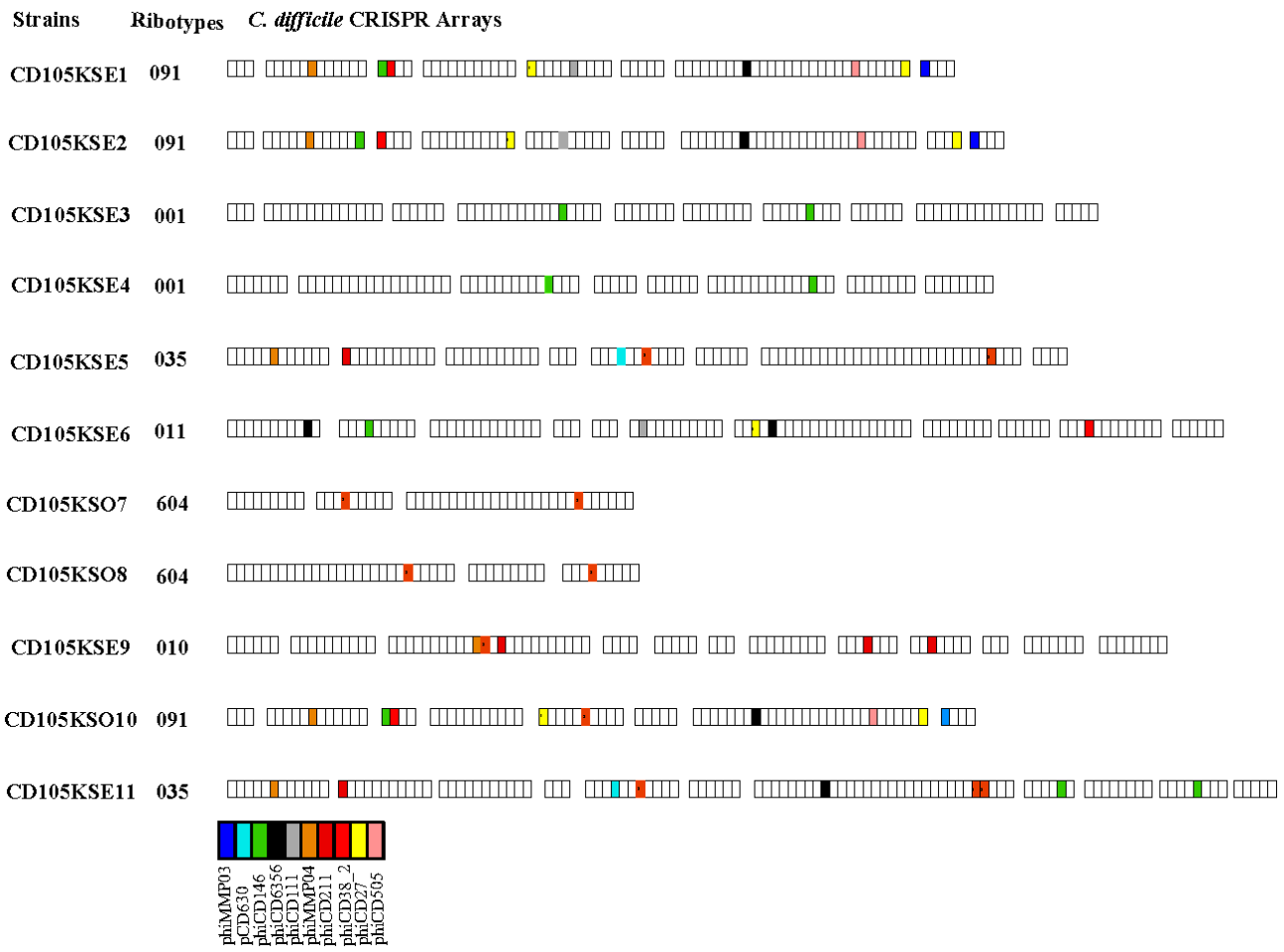


Figure 3-13. Graphical representation of *C. difficile* CRISPR arrays and homology between spacers and pahge sequences.

Homology between spacers and phage sequences are indicated by matching colours; each colour represents the similarity with particular *C. difficile* phage. Non- coloured square represent spacers have no homology to phage sequences. Spacers with multiple homologous to protospacers marked with numbers correspond to the number of protospacers that match Figure 3-14.

Bacteriophages	R091	R091	R001	R001	R035	R011	R604	R604	R010	R091	R035
CDKM15	2	2	0	0	1	1	0	0	0	2	1
CDKM9	1	1	0	0	0	0	1	1	0	1	0
CDKS6	2	2	0	0	0	0	0	0	1	2	0
phiCDHM1	0	0	0	0	1	1	0	0	0	0	1
phiCDHM3	2	2	0	0	1	1	1	1	0	2	1
phiCDHM4	0	0	1	1	1	0	0	0	0	0	1
phiCDHM5	2	2	0	0	0	0	1	1	0	2	0
phiCDHM6	2	2	0	0	0	0	1	1	0	2	0
phiCDHS1	0	0	0	0	0	0	0	0	0	0	0
CDKS7	2	2	0	0	0	0	0	0	1	2	0
CDKS8	2	2	0	0	0	0	0	0	1	2	0
phiMMP04	1	1	0	0	1	0	0	0	1	1	1
phiC2	0	0	0	0	2	1	0	0	0	0	2
phiCD111	1	1	0	0	0	1	0	0	0	1	0
phiCD146	1	1	2	2	0	1	0	0	1	1	2
phiCD211	0	0	0	0	1	0	0	0	3	0	1
phiCD27	2	2	0	0	0	0	1	1	0	2	0
phiCD38-2	1	1	0	0	0	1	0	0	1	1	0
phiCD505	1	2	0	0	0	0	1	1	0	2	0
phiCD6356	1	1	0	0	1	2	0	0	0	1	1
phiCDHM19	0	0	0	0	1	0	0	0	0	0	1
phiMMP02	3	2	0	0	0	0	2	2	0	2	0
phiMMP03	0	0	0	0	2	1	0	0	0	0	2
plasmid CD630	0	0	0	0	1	0	0	0	0	0	1

Figure 3-14. CRISPR spacer target *C.difficile* phage and plasmid sequences.

Spacer content of 11 strains was compared with the *C. difficile* phage genomes from the GenBank and default plasmid database using CRISPRtargets. Numbers and colours correspond with the identified spacers.

3.3.13 Phage sensitivity of *C. difficile* isolates

The phage susceptibility of 11 strains of *C. difficile* was tested against 12 *C. difficile* phages (Appendix 4), of which seven have been isolated previously in this laboratory (Hargreaves, 2012) from either environmentally enriched cultures or induced lysate environmental strains, and 5 phages were isolated from Kurdistan the same reservoirs, from soil and sediment enriched samples (This study). The results showed that strains of ribotypes 011, 091 and 604 could not be infected by the tested phages, whereas strain of ribotypes 001, 010 can be infected by multiple phages with different morphologies that were derived from different environments. However strains of ribotype 035 have narrow host ranges as can be infected by 3 phages belong to myoviridea family (Figure 3-15 and Table 3-6).

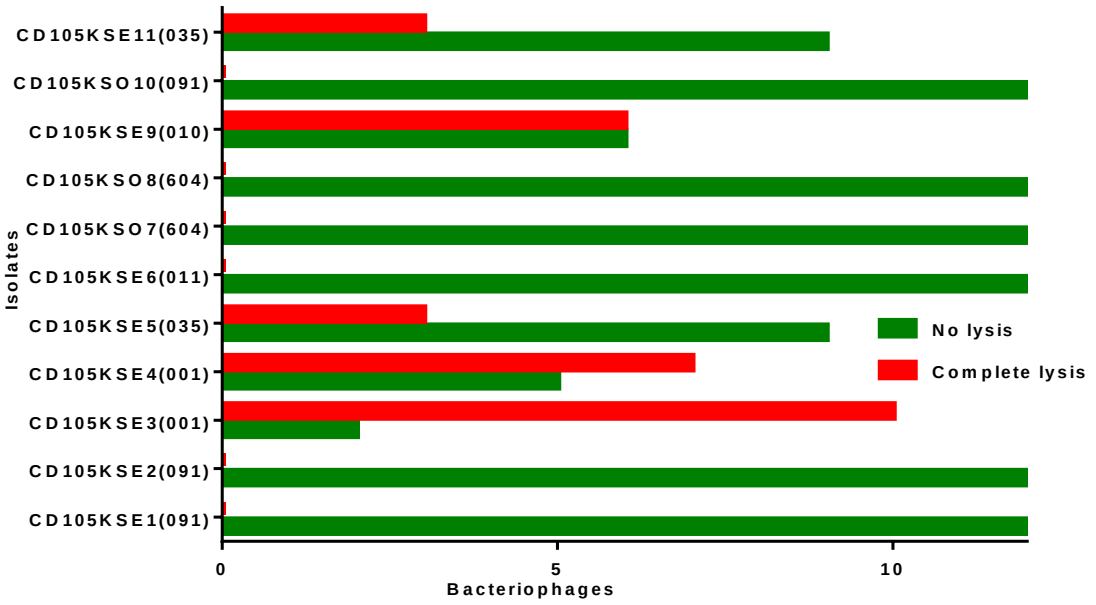


Figure 3-15. Sensitive and resistance activity of *C. difficile* strains to phages.

Sensitivity of the 11 strains against 12 phages was determined using a spot test, normalized with 10^8 PFU/mL of phage.

3.3.14 Correlation between CRISPR- Cas system and phage infection

In order to determine the mechanisms underlying phage resistance, and the processes constraining the narrow current range of hosts in our isolates, we analysed the relationship between phage sensitivity and CRISPR spacer matches with corresponding phage sequences. Results show that the CRISPR profile for each strain matched perfectly with its host range data Table 3-6. The existence of spacers homologs to with phage protospacer DNA is used to predict resistance to phages; this pattern is clear in strains of R091 and R001 that have not been infected with phages and that do indeed that have a copy, or multiple copies, of spacers that show homology to protospacers. However this was not the case in the R010 isolate where despite of the presence of spacers showing homology to phage DNA, this strain was sensitive to 'corresponding' phages.

Table 3-6. Phage sensitivity of the 11 strains and the number of spacers with similarity to phage sequences.

The sensitivity of 11 strains subjected to 12 phages. The black shading indicates complete lysis of phage infection, non-shading reflects the strains that are resistant to the phages. The numbers in the boxes indicate the number of spacers that each strain has with homology to phages sequences, 0 indicates no spacer with homology to the phages sequences present, 1 indicate presence of one spacer and 2 presence of two spacers homologues to phage sequence.

Strains	Bacteriophages											
	Myoviruses								Siphoviruses			
	phiCDHM5	phiCDHM6	phiCDHM1	phiCDHM2	phiCDHM3	phiCDHM4	CDKM9	CDKM15	CDKS7	CDKS8	CDKS6	phiCDHS1
CD105KSE1(091)	2	2	0	0	2	0	1	2	2	2	2	2
CD105KSE2(091)	2	2	0	0	2	0	1	2	2	2	2	2
CD105KSE3(001)	0	0	0	0	0	1	0	0	0	0	0	0
CD105KSE14(001)	0	0	0	0	0	1	0	0	0	0	0	0
CD105KSE6(011)	0	0	1	0	1	0	0	1	0	0	0	0
CD105KSE5(035)	0	0	1	0	1	1	0	1	0	0	0	0
CD105KSE11(035)	0	0	1	0	1	1	0	1	0	0	0	0
CD105KSO7(604)	1	1	0	0	1	0	1	0	0	0	0	0
CD105KSO8(604)	1	1	0	0	1	0	1	0	0	0	0	0
CD105KSE9(010)	0	0	0	0	0	0	0	0	1	1	1	1
CD105KSO10(091)	2	2	0	0	2	0	1	1	2	2	2	2

3.4 Discussion

In this study *C. difficile* was isolated from the river water and mountain samples in Kurdistan. Previous work in our laboratory revealed that this bacterium could be readily isolated from such environments and we were keen to explore novel geographical areas (Hargreaves et al., 2013). Again consistent with previous work, in this study the highest rates (75%) of the isolation were from sediment samples (al Saif and Brazier, 1996, Hargreaves et al., 2013, Pasquale et al., 2012, Pasquale et al., 2011), 25% were recovered from soil samples consistent with previous published data (Baverud et al., 2003, Hafiz, 1974, Kim et al., 1981, Simango, 2006, Simango and Mwakurudza, 2008); the fact that earlier workers failed to isolate *C. difficile* from soil samples (Hargreaves, 2012), may be due to the sporadic distribution of these bacteria in these environment, small sample sizes, and variations in methodologies (Kim et al., 1981). The presence of *C. difficile* in soil may be due to contamination by infected people and animals (Baverud et al., 2003, Kim et al., 1981, Simango, 2006, Simango and Mwakurudza, 2008). Furthermore studies have demonstrated that environmental strains can serve as a possible pool of toxigenic *C. difficile* (Baverud et al., 2003).

The ribotype analysis presented here shows prevalence of a set of mixed *C. difficile* strains in soil and sediments in the study area, including types that are either related to the environment, or to clinical sources, for example R010 and other strains associated with clinical R001(Hargreaves et al., 2013, Romano et al., 2012). Different ribotypes were observed at each sample site with the rare UK strains R604 and R091 commonest in the study area, at 31%, while strains of ribotype 001, 010, 014, and 081 were also identified and are common in the UK and Europe (Stubbs et al., 1999, Bakker et al., 2010, Koene et al., 2012, Hargreaves et al., 2013, Bauer et al., 2011). Isolates of R035 and R011, already in our ribotype collection, have been previously reported from Europe and the Middle East (al Saif and Brazier, 1996, Koene et al., 2012, Rotimi et al., 2003, Stabler et al., 2012, Terhes et al., 2006). In addition, the study area yielded the novel ribotype R691 that had not previously been observed, and a group of three strains that come from a novel *Clostridium* species (Rashid et al., 2016). Furthermore, we have shown that these isolates have different toxin profiles, include non-toxigenic ($A^- B^-$) and toxigenic strains with two toxin genes ($A^+ B^+$) as well as strains with just one toxin gene ($A^- B^+$). We also observed, that isolates of the same ribotype could have different toxin profiles, supporting the previous reports that the absence

of a *paloc* is not a feature of clonality and that the *PaLoc* can readily be lost (Dingle et al., 2014, Zhou et al., 2014) . These data suggest that numerous types of *C. difficile* might co-exist in the natural environment, and that the prevalence and distribution of specific ribotypes differs by location, affected by their ability to survive in the environment or by entry from contaminating sources (e.g., sewage) (Rotimi et al., 2003, Terhes et al., 2006, Hargreaves et al., 2013, Janezic et al., 2012). However, it remains unclear where these environmental strains come from, if they are active outside clinical settings, or just exist as spores (Hargreaves et al., 2013, Hargreaves et al., 2015).

Our motility assay showed that all of the strains were capable of motile growth with variability in the degree of motility. In contrast to previous studies (Hargreaves et al., 2013, Stabler et al., 2009), we did not observe any isolates with non-motile growth. This discrepancy could be due to the variability in flagellin genes (*fliC* and *fliD*) suggest that they undergo recombination events and/or are subjected to environmental selective pressure to develop (Lemee et al., 2004). However, the results of antibiotic sensitivity assays were comparable to previous studies which have reported resistance to clindamycin and ciprofloxacin in non-clinical isolates (Hargreaves et al., 2013, Norman et al., 2009), and suggest that the natural environment is a reservoir of resistance genes. The presence of fluoroquinolone resistant isolates may be of clinical concern if there is a flow from sediments back to human hosts (Hargreaves *et al*., 2013).

Further, phylogenetic analysis of whole core genome from 11 representative strains in our collection identified a lineage (C-I) that is extremely different from the four already established (supported by bootstrap) (Figure 3-8). Similar to previous reports, this clade includes just non-toxigenic strains (Dingle et al., 2014, Janezic and Rupnik, 2015), although others have shown that strains with unique toxin profiles (A^+B^-) fall within this clade (Monot et al., 2015). Similar again to earlier reports, clade 1 is diverse in RTs and STs, where toxigenic strains are combined with non-toxigenic one (Stabler et al., 2012), and MLST and PCR ribotype analysis were also comparable in their discriminatory power; RTs with multiple STs and *vice versa* reveal that the *C. difficile* genome is constantly diverging (Griffiths et al., 2010, Stabler et al., 2012). Additionally, the genome comparisons for strains in clade 1 reveal divergence across isolates with respect to the reference genome CD630 and demonstrate that major regions of divergence occur at sites where transposons are located. These regions

encode mobile genetic elements, phage proteins, surface layer proteins, and proteins for antibiotic resistance. All of these elements have a profound effect on the biology of an organism, including the transfer of antibiotic resistance and other factors that allow survival in challenging environments (Mullany et al., 2015).

The prevalence of prophage carriage in bacteria varies from 4% to almost 100%, dependent on species, source of isolates, and the method used to assess proportions (Coetzee and De Klerk, 1962, Jiang and Paul, 1998). Genome examination showed that all 11 strains harbour multiple, and diverse, phage sequences; to the best of our knowledge ours is the first report that shows the presence of six *C. difficile* prophages in a single host, signifying that phage infection plays an essential part in host biology, and that these temperate phages have access to diverse strains in the environment (Hargreaves et al., 2013). Previous studies on *C. difficile* strains in the environment (Hargreaves et al., 2013), as well as clinical and infant isolates (Fortier and Moineau, 2007, Nale et al., 2012, Shan et al., 2012), reported multiple and diverse prophage carriage but in the lower ratios. In agreement with earlier studies, we also observed different prophage carriage with strains of the same ribotype (Nale et al., 2012, Hargreaves et al., 2013), but there are discrepancies with the results of other research as strains of same ribotype released morphology identical prophages (Fortier and Moineau, 2007). In addition, TEM examination of induced lysates was not comparable with genome examination with the exception of the R010 strain, as both analyses showed the presence of two phage particles beside phage-tail-like particles. This discrepancy might be due to the nature of prophage release by different inducing agents, although using TEM to assess the frequency of prophages is likely to underestimate populations as it depends on the titre of phages present in a sample (Shan et al., 2012, Hargreaves et al., 2013). Consistent with previous studies we also observed dual phage carriage in the genome of one strain, two morphologically and genetically distinct phages encoded in a single *C. difficile* (Fortier and Moineau, 2007, Goh et al., 2007, Hargreaves et al., 2013, Nale et al., 2012), although from the induced lysate just the siphovirus particle was observed in addition to PTLP.

Phage-tail-like particles are thought to be bacteriocins encoded by the bacterium (Fortier and Moineau, 2007, Hargreaves et al., 2013, Nagy and Foldes, 1991, Nale et al., 2012, Shan et al., 2012). Two type of phage-tail-like particles observed in all our induced lysates have been reported previously in other induced lysates, including those represented by morphology

H (Fortier and Moineau, 2007, Nale et al., 2012, Hargreaves et al., 2013, Hegarty et al., 2016, Yamamoto et al., 1970) and phage tail particles morphology G (Nale et al., 2012) that have the ability to kill *C. difficile in vitro* and thus might represent a potential therapy for CDI (Sangster et al., 2015). Multiple prophages in these strains may explain bacterial resistance by preventing phage infection through super infection with related phages (Govind R, 2011).

One potential mechanism used by *C. difficile* to resist phage infection and gene transfer via horizontal gene transfer is the CRISPR- Cas system (Hargreaves et al., 2014a). Examination of the genome of our strains shows that this system is diverse and functions as in other examined species (Seed et al., 2013, Soutourina et al., 2013). Thus, because strains of the same ribotype might either have similar, or divergent, CRISPR contents, we propose that these arrays undergo prompt evolution within a ribotype (Hargreaves et al., 2014a). In line with earlier studies, we observed multiple *C. difficile* spacers that match known phage and plasmid sequences, with multiple elements shared between strains (Figure 3-13 and Figure 3-14). The presence of a spacer that matches to phage sequences suggests that bacteria have been exposed to this phage, or its relatives, while the absence of a spacer match suggests that there has been no obvious exposure as a 100 % match between spacer and proto-spacer sequences is required to provide immunity (Bolotin et al., 2005, Horvath et al., 2009, Stern et al., 2012, Hargreaves et al., 2014a). In addition, the CRISPR profile for each isolates matches perfectly to host range data (Table 3-6). Generally, spacers with homology to phage DNA are expected to provide resistance to matching phages (Boudry et al., 2015, Hargreaves et al., 2014a), but this is not the case in the R010 isolate. This strain was infected by phages in spite of the presence of a spacer copy that matches to corresponding phages; publication on the *C. difficile* phage host range has led to the suggestion that the interface between phage-host does not depend exclusively on the CRISPR/Cas system but on other factors as well including differences in absorption count (Ramesh et al., 1999). A similar observation was seen with the *C. difficile* phage CD38-2 when it was used to infect strains of CD196 and R20291 (Sekulovic et al., 2014) that have a copy of spacer with similarity to the sequence in its genome, suggesting that this phage has the facility to escape the CRISPR/Cas system (Hargreaves et al., 2014a).

3.5 Conclusion

Taken together, our data provide new insights into the prevalence and diversity of the *C. difficile* population in natural environment in the Middle East. A large collection of environmental strains has been established and characterised in the area including a novel ribotype as well as novel species of *Clostridium*. Their genomic features were explored as well and contained abundant prophage carriage and multiple functional CRISPR arrays that 0020 contained diverse spacers.

Chapter 4 Isolation and characterization of *C. difficile* phages from the natural environment for therapeutic purpose

4.1 Introduction

Clostridium difficile is an opportunistic bacterium that can cause infection in patients after an alteration in the intestinal microbiota following the use of antibiotics. *Clostridium difficile* infection (CDI) is one of the most reported and reoccurring healthcare associated infections in industrialized countries (Depestel and Aronoff, 2013, Magill et al., 2014). The high prevalence of CDI leads associated increases in the burden on healthcare and society (Avila et al., 2016). This increase of CDI incidence is due to the rapid emergence of hypervirulent strains that raised concerns regarding the evolution of its pathogenicity and antibiotic resistance (Knight et al., 2015, Lessa et al., 2012). In an attempt to treat this problematic pathogen, novel ways are being explored, and phage therapy is one alternative approach.

Bacteriophages (phages) are viruses that specifically infect bacteria and are usually specific to a single species or individual strains (Clokie et al., 2011). Phages are the most diverse entities on earth, and are abundantly present wherever bacteria are found, so in soil, rivers, lakes and sewage (Kutter and Sulakvelidze, 2004). When plaque assay counts are compared to the abundant phage-like particles that are seen by counting under epiflourescence or electron microscopes, it would appear that majority of phages are yet to be isolated (Hendrix, 2002). Phage that infect *Clostridium* species were first isolated by Cowles in 1934 and they targeted *C. tetani* from sewage (Cowles, 1934). Since then, phages with both lytic and lysogenic life cycles have been isolated (Dolman and Chang, 1972, Roseman and Richardson, 1969, Nieves et al., 1981, Takumi et al., 1980, Shimamura et al., 1974). Some Clostridial phages have been found to infect different species, such as the *C. sordellii* phages that infect *C. difficile* (Schallehn, 1985) but most are specific to one species (Hargreaves and Clokie, 2014).

The first *C. difficile* phage isolated was in 1983 and it was used for bacterial typing purposes (Sell *et al*.1983), since then several *C. difficile* phages have been isolated and described in the literature (Bacon et al., 1988, Dei, 1989, Goh et al., 2005b, Govind et al., 2006, Hargreaves et al., 2013, Hawkins et al., 1984, Horgan et al., 2010, Mahony et al., 1985, Meessen-Pinard et al., 2012, Nagy and Foldes, 1991, Nale et al., 2012, Ramesh et al., 1999,

Sekulovic et al., 2014, Sekulovic et al., 2011, Shan et al., 2012, Wittmann et al., 2015, Fortier and Moineau, 2007). Different methods have been used to isolate *C. difficile* phages, including: Lysogenic induction, direct screening of faeces, soil, sediment and river samples and environmental sampling following the enrichment of soil, sediment and river samples with nutrient rich media (Adams, 1959, Goh et al., 2005b, Hargreaves, 2012, Nale et al., 2012, Shan et al., 2012).

The induction of *C. difficile* strains with antibiotics mitomycin C and norfloxacin is the most popular method to isolate *C. difficile* phages, so most of the phages that are described in the literature and listed above have been isolated in this way. Due to the high frequency of prophage carriage in *C. difficile* strains, 90% of clinically relevant strains of *C. difficile* were found to carry one or more prophage (Shan et al., 2012). No *C. difficile* phages have been isolated yet through the direct screening of samples, all attempts have been unsuccessful (Goh et al., 2005b). However, some studies successfully isolated *C. difficile* through the enriched samples from sites inhabited by the bacterium itself such as soil, sediment and stool samples (Hargreaves, 2012, Meessen-Pinard et al., 2012). Despite this variety of isolation protocols, all the *C. difficile* phages that have been identified so far appear to be temperate phages, meaning that they encode integrases and other genes associated with lysogeny and thus have the potential to integrate their genomes within the host bacterial genome to form a lysogen. Despite this potential and in some cases demonstrated access to the lysogenic life-style, studies have shown that temperate phages can infect *C. difficile* lytically (Goh et al., 2005b, Hargreaves et al., 2014b, Sekulovic et al., 2014, Sekulovic et al., 2011).

Regardless of phage life styles, studies have shown that there is significant variation in the degree categorization of *C. difficile* phages, and include one or more of the following types of data:, host range studies, morphology information, growth dynamics, whole genome sequences, and comparative genetics (Hargreaves and Clokie, 2014). In general, phage host range refer to the bacterial genera, species and strains it can lyse (Kutter, 2009). Host range studies of *C. difficile* phages have shown that the phages can infect across ribotype groups; for example, phage phiCD6356 and phiCD38-2 could infect 35% and 47% of the strains tested, that belonged to 5 and 11 ribotypes, respectively (Horgan et al., 2010, Sekulovic et al., 2011). However, 79/207 of the strains tested against phiCD38-2 belonged to one type of strain. The largest host range survey published so far included a panel of 80 strains belonging to 21

ribotypes that was tested against 9 phages (Nale et al., 2015a). PhiCDHM3 and phiCDHS1 infected 39% and 38% of tested isolates belonging to 12 and 11 ribotypes respectively. The ribotypes included in the screening panel were from three different continents (Nale et al., 2015a). The isolation of phages able to infect clinically relevant prevalent strains is of key importance and may need to be constantly modified in order to determine the most effective combination of phages for therapeutic purposes.

In Prof Martha Clokie's laboratory, several *C. difficile* phages were isolated by a previous PhD student, Katherine Hargreaves, from sediment samples and these phages were shown to infect many strains of clinically relevant *C. difficile*. However, only a few phages can infect the clinically relevant and increasing problematic ribotype 078. *C. difficile* ribotype 078 has become a predominant type in cattle and causes CDI in humans with a similar severity to CDI due to ribotype 027 (Janezic et al., 2014). The isolation of new phages with a broad range of *C. difficile* strains that can infect multiple clinically relevant ribotypes might be readily isolated in the natural environment where its prevalence has not been addressed before.

4.2 Aims and objectives

1. The initial aims of the study were to identify novel *C. difficile* phages, preferably those with no lysogenic potential, so they would therefore be putative candidates for use as therapeutic agents against CDI.
2. Isolate new phages that can infect *C. difficile* strains of multiple clinically relevant ribotypes including strains of R078 in a lytic manner, to expand in the diversity of phage in our laboratory catalogue (University of Leicester, UK).
3. Characterise the isolated phages based on their
    a. Morphology using transmission electron microscopy.
    b. Genome size using pulsed-field gel electrophoresis.
4. Investigate the therapeutic potential of the phages through host range studies using spot test.

4.3 Results

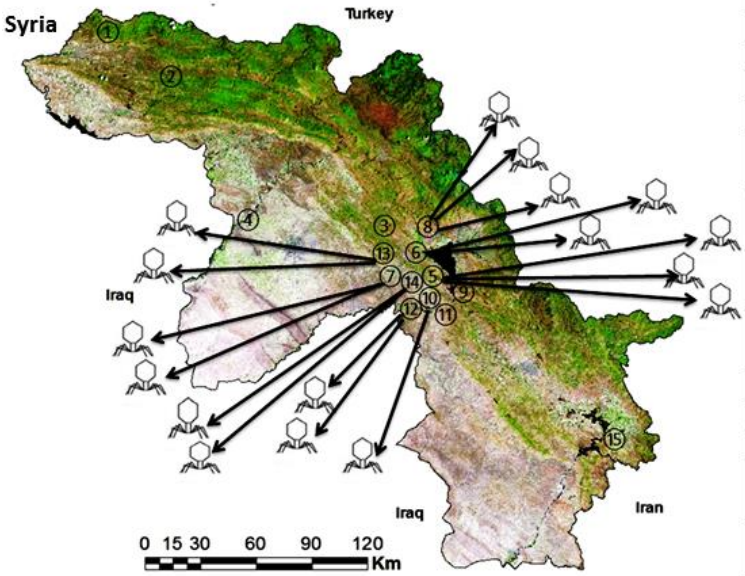
4.3.1 Isolation of lytic *C. difficile* phages

Bacteriophages are present wherever bacteria reside, in the previous chapter we isolated genetically diverse set of *C. difficile* strains in Kurdistan the northern of Iraq. No previous studies have addressesd prevalence of this pathogen and its bacteriophages in this area. Accordingly, we used the soil and sediment samples described that previously in section 2.1.1 to isolate *C. difficile* phages.

A panel of soil and sediment samples (n=15) obtained from 8 sample sites in the studied area was used as a first round for lytic phage screening (Figure 4-1). The phage extracts were prepared from the environmental samples as described in section 2.18.1 and were spotted directly on lawns from a panel of *C. difficile* strains (n=14) using the spot test. However, none of these tests yielded plaques. Further unsuccessful attempts occurred when 9 soil and sediment samples collected from 7 sample sites were tested.

4.3.2 Isolation of *C. difficile* phages from the enriched samples

As no phages were isolated using the direct method, enrichment of soil and sediment samples method was performed. A volume of 10 µl of the supernatant taken from a 10-day culture was used in spot assay on different environmental and clinical *C. difficile* isolates of 4 ribotypes (n=14). Phage plaques were observed for 10/15 enriched samples on strains AIU, K15 and HB of R027, strain T6 and R10 of ribotype 076 and 010 respectively. A third round of plaque purification was carried out and this resulted in the isolation of 16 phages: CDKS2, CDKS3, CDKS4, CDKS5, CDKS6, CDKS7, CDKS8, CDKS9, CDKS10,CDKS12, CDKS13, CDKS14, CDKS15, CDKS16, CDKM9, CDKM15, which were purified on 5 *C. difficile* strains from different sources. Some samples (n= 4) each yielded two phages and were bulked on two different hosts, and one enriched phage sample yielded three phages that were purified on 3 indicator strains. Lastly, five enriched phage samples yielded single phage particles. The second round of phage screening using 9 enriched samples resulted in isolation of phage CDKM4, using the same set of *C. difficile* strains as used for the first round of phage isolation, with an extra 4 strains of one ribotype that were isolated in this project. In total, 58.8 % of the phages were isolated from soil samples and 41.2% of the phages were isolated from sediment samples.



Sites	Source	Notes	Bacteriophages
1	Dalal Bridge	Sediment	-ve
2	Duhok Valley	Sediment	-ve
3	Safien Mountain	Soil	-ve
4	Kalak River	Sediment	-ve
5	Chnarok Mountain	Soil	CDKS4/CDKS5/ CDKS14
6	Hizope River	Soil/ Sediment	CDKS8/CDKM9
7	Hamamok Resort	Soil/ Sediment	CDKS15/CDKM15
8	Bestana	Soil/ Sediment	CDKS6/CDKS7/ CDKS16
9	Dokan Resort	Sediment	-ve
10	Khalakan	Soil	CDKM4
11	Kalkasmaq	Soil	-ve
12	Taq Taq River	Soil/ Sediment	CDKS2/CDKS9
13	Jalee River	Soil/ Sediment	CDKS10/CDKS12
14	Haibat sultan Mountain	Soil	CDKS3/CDKS13
15	Sirwan River	Sediment	-ve

Figure 4-1. Attendance of *C. difficile* phages arcos Kurdistan the North of Iraq.

Left: map of sampling sites alongside rivers and mountains. Sites are numbered to correlate with the table on the right. Phage Symbols indicate where the *C. difficile* phages were isolated. Right: table of sampling sites, sample type and the bacteriophages. Numbers of sites correspond to map on the left and are referred to the text.

4.3.3 **Transmission electron microscope (TEM) analysis of the isolated phages**

The results from the TEM analysis of the 17 phage lysates showed the phage particles are belong to either *Myoviridae* or *Siphoviridiae* family within the order *Caudovirales* (Figure 4-2). The myoviruses have contractile tail sheaths and siphoviruses have non-contractile tails. The phages can be assigned morphotypes as described by Ackermann (2001), 14/17 are siphoviruses belonging to B1 morphotype and 3/17 are myoviruses of morphotype A1, with tail lengths of 250nm long and capsid diameter 60nm. There were also phage tail-like particles (PTLPs) observed in the examined lysates, which were distinguishable from broken myovirus tails as they did not have an inner sheath when contracted or a clear neck region. Figure 4-2 shows the morphology of 16 phages as the TEM examination of CDKS16 showed only broken tail.

4.3.4 **Phage nomenclature**

The phages were named in such a way as to provide information on the species, the place of isolation from and the family they belong to. They were then assigned a number in succession. For example 'CDKM9' or CDKS2, where CD is *C. difficile*, K is Kurdistan, M is myoviruse and S is siphovirus, and they are the 9^{th} and 2^{nd} phage that meets those criteria respectively.

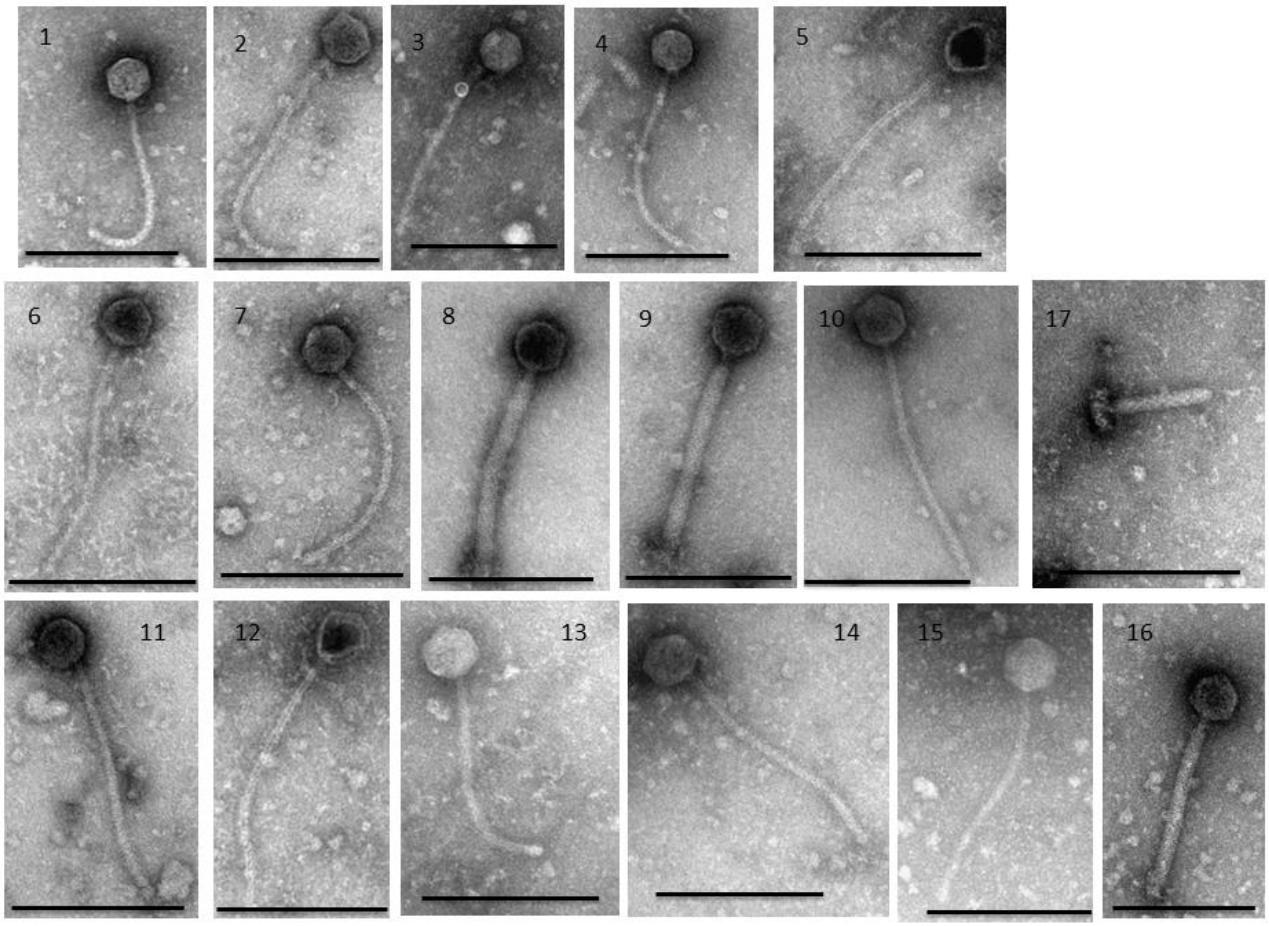


Figure 4-2. Morphology of isolated phages in natural environment form Kurdistan the North of Iraq.

The phage samples 1: CDKS2, 2: CDKS13, 3: CDKS14, 4: CDKS6, 5: CDKS7, 6: CDKS4, 7: CDKS3, 8: CDKM9, 9: CDKM15, 10: CDKS8, 11: CDKS9, 12: CDKS5, 13: CDKS15, 14: CDKS12, 15: CDKS10, 16: CDKM4, 17: PTLP. TEM images of the 16 phages and their structures reveal that 13 of these belong to the Siphoviridae family and 3 belong to Myoviridae family. The scale bar represents ~200nm.

4.3.5 **Analysis of phage genome size using Pulsed Field Gel Electrophoresis**

The genome sizes for 8 of the 17 phages were obtained using PFGE analysis, shown as a composite image from two seperate gels (Figure 4-3). The phage genome sizes correlate with the phage's particle morphology. The three myoviruses (CDKM4, CDKM9 and CDKM15) have similar sized genomes ranging from 50-55 kbp. The siphoviruses (CDKS2, CDKS5, CDKS7, CDKS9, CDKS10, CDKS10, CDKS12, CDKS13 and CDKS16) also have similar also sized genome at ~ 41 kbp. Producing a phage stock with high enough phage titre to use as a plug to determine the size was

difficult and had to be repeated 4 times. Therefore, the intensity of the bands was varied correspondingly to the titre of phages. Some phages failed to produce the bands indicating that this was not a high enough titre. These phages were, CDKS15, CDKS4, CDKS3, CDKS6, CDKS8 and CDKS14.

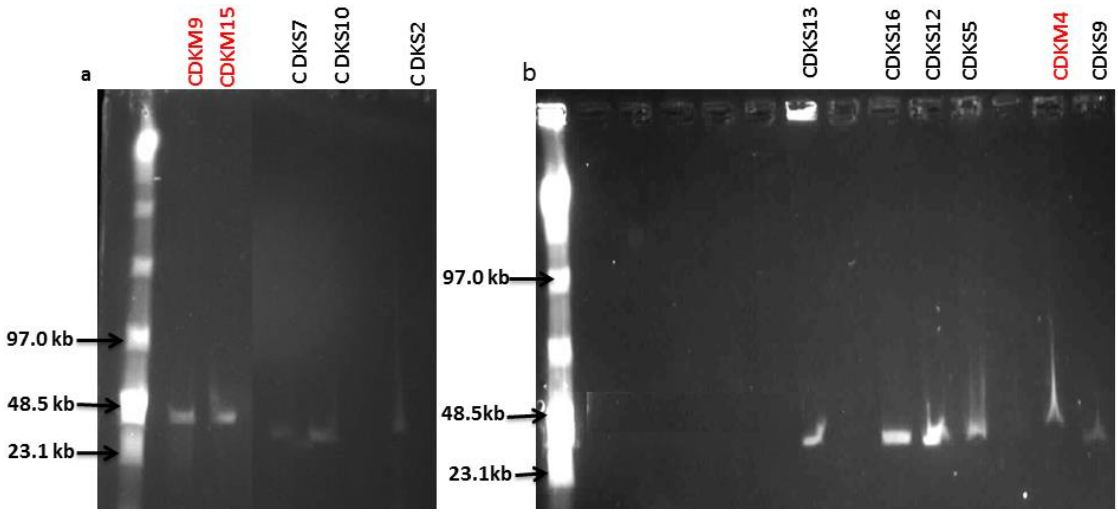


Figure 4-3. Phage genome analysis using Pulsed-field gel electrophoresis analysis showed similar sizes.

Phage DNA was run on PFGE, gel stained with ethidium bromide and visualised under UV light. A low range PFGE ladder was included and the genome sizes of the phages ranged from 45 to 55 kb.

4.3.6 **Host range analysis of the isolated phages**

Host range analysis was conducted for the purified phages (n= 17) to determine if they had narrow or broad host ranges. To do this, the phages were tested against a diverse panel of *C. difficile* strains (n= 96) that belong to 23 ribotypes, including clinical (n= 49) and environmental isolates (n= 47). Of these environmental strains, 23 strains were isolated from the local environment from which the phages themselves were isolated and 24 strains were isolated from the UK (Appendix3).

The host range analysis showed that the phages could infect multiple strains of *C. difficile* of both clinical and non-clinical significance. All phages infected clinically relevant strains such as epidemic isolates of ribotype 001; 16/17 of the phages infected hypervirulent strains of ribotype 027 and 13/17 of phages infected strains of R078. In addition, all phages infect strains from the local environment and strains from distant

geographical areas. Interestingly, however, the proportion of the strains from the local environment that could be infected was lower (35%; 8/23) than the proportion of strains from the UK environment (63%; 15/24). The phages could not infect strains of 9/23 ribotypes that tested for, which included: 002, 011, 012, 017, 021, 085, 091, 604 and 691 (Table 4-1 and Appendix 10).

Over all, the siphoviruses had highly similar host range patterns; they infected strains of same ribotypes 6/23, with differences only in the number of strains infected within the same ribotypes. CDKS6, CDKS7 and CDKS8 had the widest host range, as they could infect 37% of all tested isolates belonging to 6/23 ribotypes Figure 4-4. The myoviruses CDKM9 and CDKM15 both had a broad host range, infecting 18% of isolates of 9/23 and 7/23 ribotypes, respectively. The phages with the narrowest host ranges were the siphoviruses CDKS14 and the myoviruse CDKM4 that infected 26% and 8% of isolates of 5 and 6 ribotypes including epidemic strains of R001, respectively Figure 4-4 .

Remarkably, the two myoviruses CDKM15 and CDKM9 were capable of infecting 2/18 strains of R027, only one study has reported a myovirus (phiCD481-1) that could infect strains R20291 of R027, but this has different morphology and smaller size (~30 kbp) (Sekulovic et al., 2014). So CDKM9 and CDKM15 are first long tailed myoviruses could infect hypervirulent strains of R027.

In addition, extensive intra-ribotype variation was observed in phage infection, For instance, the 7 strains of R010 could be differently infected by 14/17 of the tested phages, and both strain of R14/020 can be infected by 3/17 phages. Variation in phage infection were also observed in the host range analysis, in which same phage could infect the isolates differently (confluent lysis and turbid). This was the case for CDKS3 produced confluent clearance on 6 isolates of R001 and turbid zone on 3 isolates of same ribotype. Likewise, single host could be subjected to confluent lysis from some phages and turbid clearance from others. This suggests that the dynamic of phage and host interaction controls the outcome of infection in *C. difficile*.

Table 4-1. The number of ribotypes infected by 32 *C. difficile* phages isolated

Ribotype	No. of isolates	CDKS2	CDKS3	CDKS4	CDKS5	CDKS6	CDKS7	CDKS8	CDKS9	CDKS10	CDKS12	CDKS13	CDKS14	CDKS15	CDKS16	CDKM4	CDKM9	CDKM15
001	10	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
002	2																	
005	2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			
010	7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			
011	2																	
012	3																	
014	2															✓	✓	✓
015	1																✓	
017	2																	
014/020	2															✓	✓	✓
021	1																	
023	2	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			
027	18	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓
031	2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓
035	2																✓	
076	1															✓	✓	✓
078	16	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓			
085	1																	
091	6																	
106	4															✓	✓	✓
220	4															✓	✓	✓
604	3																	
691	2																	

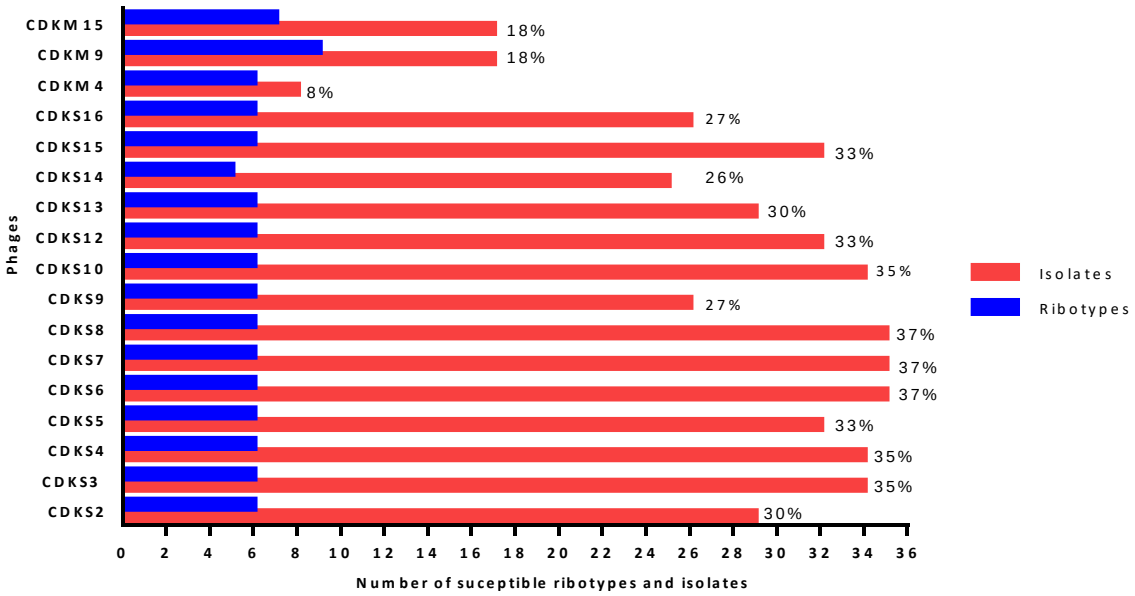


Figure 4-4. Host range analysis of the seventeen phages against 96 strains and 23 ribotypes.

The host range of the phages was determined using spot test of 10 µl of 10^8 PFU/ml of phage stocks on agar overlays containing cultures of the hosts. A, Bars represent total number of susceptible ribotypes and strains infected by the phages.

4.3.7 **Phage plaquing efficiency**

Throughout the host range analysis, it was noticed that some phages formed different numbers of plaques on certain isolates compared to others. To further examine this observation, the efficiency of plaquing (EOP) was examined for 4 phages (2 siphoviruses and 2 long-tailed myoviruses) on 24 isolates of 6 ribotypes (Appendix3). The two myoviruses CDKM9 and CDKM15 were isolated on the same host while the siphoviruses CDKS6 and CDKS8 had the same host as well. Multiple isolates were included for each ribotype, with one isolate being infected by at least one phage. Table 4-2 shows phage efficiency of plaquing on different hosts which can be compared between the strains as the same phage stock with a titre of 10^8 PFU was used for all infections.

Confluent lysis, which refers to a high EOP, was noticed on at least one isolate for every phage (n=1-2), while plaque formations were observed for all the phages in the

spot test. Turbid zones indicative of lysogeny were noticed for the two myoviruses. The results suggest that the efficiency of infection differs for each phage among different strains even of the same ribotype, regardless of morphological similarity.

Table 4-2. Phage plaquing efficiency for four phages on 24 strains of *C. difficile* belonging to 6 ribotypes.

Colour shade indicates different phage infection parameters: Orange= individual plaquing, green= confluent lysis and red = turbid zones.

		Phages (SV)		LTV	
Ribotype	Isolate	CDKS6	CDKS8	CDKM9	CDKM15
001	H8				
	H18				
	H4				
	CDNCTC1120				
005	H5b				
	K18				
	H5				
	H1b				
010	H3				
	I				
	R				
	H1B				
002	P				
	0				
	H15				
	S				
220	Y				
	K				
	H12				
	K12				
106	CD106				
	R40V0106				
	R6106				
	K10				

4.4 Discussion

To date the majority of the phages infecting *C.difficile* are from industrialized countries where many studies have been conducted on *C. difficile*; no *C. difficile* phages have been isolated in Middle East, especially in the North of Iraq. Therefore, our knowledge about diversity of *C. difficile* phages in Middle East is lacking. Here we describe isolation and characterization of *C. difficile* phages in this area.

No genuinely 'lytic' phages were isolated in this project, despite screening of 24 soil and sediment samples using the direct method. However, previous studies also failed to isolate free phages via the direct method, despite testing a wide variety of samples, including soil, untreated sewage, stool samples from healthy people, inflammatory bowel disease (IBD) patients and healthy pigs, as well as pig caecal samples and slurry (Horgan et al., 2010, Goh et al., 2005b). One study has failed to isolate lytic phages from screening 400 stool samples from CDI patients (Shan et al., 2012).

The fact that no virulent phages have been isolated to date may be explained in terms of the presence of *C. difficile* as spores in the environment due to it being an obligate anaerobe, and for effective phage infection the host must be in an active form of growth and the lack of cell surface structures that function as phage receptors in a spore (Goh et al., 2005b). This is showing that the lysogenic cycle is apparently appropriate for phages of spore-forming anaerobes, as prophages are not restricted by the accessibility of metabolically active anaerobic cells (Goh et al., 2007, Hargreaves and Clokie, 2014).

Despite the inefficiency of the direct method in isolating phages, 17 phages were isolated from the prior enrichment of soil and sediment samples. Phages isolated by the enrichment method may be virulent phages due to the ability to propagate on the naturally present culture of *C. difficile* or might be prophages released by these strains, as these samples were positive for *C. difficile* isolates. Additionally, the enrichment medium contained antibiotics as a selective agent that probably induced the prophage release (Hargreaves, 2012). These 17 phages require genome sequencing and characterisation to confirm if they are temperate phages or virulent phages.

In accordance with our result, one study successfully isolated phage particles using the same method, in which four myovirideae phages were isolated as free viral particles

from the enrichment of pooled patient stool samples due to *in vivo* prophage spontaneous release (Meessen-Pinard et al., 2012). Another two independent studies isolated *C. difficile* phages from screening river water, soil and sediment samples using the enrichment method (Hargreaves, 2012, Thanki, 2016). However, others failed to isolate phages using the same method from screening soil and animal faeces, untreated sewage, faecal samples from healthy humans, pig caecal and slurry samples, despite optimizing the isolation technique.

The outcome of host range analysis showed some interesting data. For instance, the phages could be identified as generalist phages as they infected multiple and diverse strains of *C. difficile*. Generalist phages are desirable for potential therapeutic use in order to successfully kill a larger number of diverse *C. difficile* strains. Therefore, the phages are an ideal candidate to be used in phage therapy. Although the host ranges for the siphovruses appeared to be more closely related, the most effective is CDKS8 which killed and infected 37% of the strains tested and was the only phage that infected 5/ 17 strains of ribotype 078. Adding phages that effectively infect strains of ribotype 078 to the phage catalogue will increase the diversity of phages available for the development of a therapeutic approach to combat *C. difficile* infection in the future. Also of particular interest, CDKM9 and CDKM15 comprise the first long tailed myoviruses that infected strains of R027. Previous *C.difficile* host range studies revealed that strains of R027 can be infected only by siphoviruses, with the exception of strain R20291 found to be sensitive to the medium myoviruse phiCD481-1 (Sekulovic et al., 2014). In addition, we also observed phages with narrow host range in our collection, such as CDKM4 which might be useful in therapeutic approach to combat CDI based on a combination of phages. The ability of phages to infect a broad range of *C. difficile* strains is a key attribute deemed necessary for their development for therapeutic use.

Another significant point that was noticed through the host range analysis is the intra-ribotype variation in which multiple strains of a single ribotype were not all infected by the phages, signifying that the inherent variation between strains within a ribotype group results in variation in phage invasion (Hargreaves and Clokie, 2014). This observation is in accordance with previous studies (Sekulovic et al., 2014, Sekulovic et al., 2011). Strains within a single ribotype have different phage specificity profiles, this can be due to the expression of various phage receptors at the cell surface, prevalence of different prophage carriage that provide phage-specific immunity, and the

incidence of various CRISPR spacers, or any other phage-defence mechanism (Sekulovic et al., 2014). This is providing an explanation for the narrow host range of the local strains, as the phages could only infect 35% of the local strains. The CRISPR spacer analysis of these strains revealed homology between the genome sequences of the tested phages, besides the presence of multiple prophage carriage within these strains conferred protection to their host.

Another observation that was made during the host range analysis with regard to phage infection pertained to the dissimilarities between the plaquing efficiency that a particular phage had on different strains. Hosts have different resistance strategies that can alter the outcome of phage infection and correlate to bacterial mechanisms of phage resistance. These include adsorption resistance which is known as ‘’restriction’’, phage-genome uptake blocks, restriction modification, CRISPR/cas systems and abortive infection (Hyman and Abedon, 2010). These mechanisms that are behind decreasing EOP without complete inhibition of phage infection are well understood across bacterial populations.

The variation in EOP is the outcome of the phage host interaction rather than the viability of phages in the stock. Difficulties in propagating *C. difficile* phages are well documented and usually require technique optimization to isolate and propagate them in the laboratory (Goh et al., 2005b, Horgan et al., 2010, Mahony et al., 1985).

To conclude, we have isolated 17 phages from enriched cultures of soil and sediment samples that could be temperate or virulent. Host range analysis demonstrated that the phages could infect strains from environmental sources and from hospital settings. They also showed that they were generalist phages that could infect strains from local and more distant environments, the siphoviruses being particularly closely related due to their similar host range trend. We also showed that the strains of a single ribotype had different phage susceptibility, suggesting variation in the genetic basis between strains of one group ribotype, in phage gene resistance which may be horizontally transferred between *C. difficile* populations, and in the prevalence of prophage carriage that provide phage homo immunity to the host as well as presence of a CRISPR/cas system. The host range assessment showed that the two myoviruses were exceptional, being the only two myoviruses that could infect the strain of R027. In the next chapter we will explore the genomic characterisation of CDKS8, CDKM9 and

CDKM15, to determine whether these phages are temperate or virulent, It is also necessary to ensure that their genome does not encode for putative virulence factors. Genome analysis will reveal their genetic content and their relatedness to known *C. difficile* phages through phylogenetic analysis and consideration of their diversity.

Chapter 5 Genomic characterisation of *C. difficile* phages isolated in the north of Iraq

5.1 Introduction to *C. difficile* phages

To date there are several *C. difficile* phages and prophages that have been described in the literature and their complete genome sequences are available in the databases (Govind et al., 2006, Goh et al., 2007, Mayer et al., 2008, Hargreaves et al., 2014a, Hargreaves et al., 2014b, Wittmann et al., 2015, Hargreaves and Clokie, 2015, Horgan et al., 2010, Meessen-Pinard et al., 2012, Sekulovic et al., 2011). These phages are all temperate phages isolated either by lysogenic induction or enriched samples from different geographical areas in Europe and North America, but some places remain less well defined in terms of genetic diversity across phage and related prophages sequences. This project is the first study to exploit the genetic diversity of *C. difficile* phages in the Middle East, especially in the north of Iraq.

The *C. difficile* phages from the studies listed above are all members of the Caudovirales order that belong to either the *Myoviridae* or *Siphoviridae* families. Previous characterisation of *C. difficile* phage genomes suggest that the myoviruses could be grouped according to particle morphology, tail length and capsid diameter as follows; long tailed myoviruses (LTM) with capsid diameter between ~60 to ~70 nm and tail lengths between 150–260 nm, medium myoviruses (MM) with capsid diameters between ~60–70 nm and tail lengths of 110–130 nm, and small myoviruses (SM) with capsid diameters of 40–60 nm and tail lengths of 105–110 nm (Hargreaves and Clokie, 2015). All phage genome sizes fall within 30-57 kbp in length, with GC% from 28.4% to 30.8%, with the exception of the two “jumbo" myoviruses (phiCD211 and phiCDIF1296T) their genomes being 131kbp in length, and the latter has a GC% of 26.4%. These two phages have the largest genome sizes of a temperate *C. difficile* phage recognised so far (Wittmann et al., 2015).

5.2 *C. difficile* genes in phage genomes

In order to obtain more information in the biology of some of the phages isolated in this study, the genome of phages CDKM9, CDKM15, CDKS7 and CDKS8 were sequenced. To use the phages therapeutically it is important to ensure that they do not contain toxins or genes that could impact on therapeutic use. Phage genome architecture, packaging and genes of interest that have been found in *C. difficile* phages are discussed below.

5.2.1 **Phage genome architecture**

In order to understand the dynamics of phage-host interactions, the diversity of phages may be assessed as a way to relate host range to the species or type of phage with the ability to infect certain strains (Koskella and Meaden, 2013). To date, taxonomic grouping of the *C. difficile* phages includes the species phicd119likevirus and a proposed species ''phiMMP04likevirus'' (Hargreaves and Clokie, 2015, Lavigne et al., 2009), but some phages do not fall into either genus.

This variation in part could be due to the fact that phage genomes are usually mosaic, due to the horizontal gene transfer of single genes or groups of genes, often caused by to the recombination events between temperate phages (Hendrix et al., 2000, Simon et al., 1971). Mosaic architecture is a feature of Caudovirales genomes in general, where phages have a distinctive core set of units with different evolutionary histories arising from their exchange amongst phages (Hatfull, 2008, Hendrix et al., 2000, Pope et al., 2011). Moreover, phages are also characterised by a rapid evolution in which mutations, recombination between different phage genes, and horizontal gene transfer result in low phylogenetic signals (Nilsson et al., 2011). While functional genes such as endolysin have been used as phylogenetic markers for phages, the utility of such genes in any one dataset is linked to their biological roles. The utility of using endolysin as a novel antimicrobial has been explored due to the capacity of lysins to target and lyse susceptible bacteria, as shown for *C. difficile* (Kumar et al., 2016, Mayer et al., 2011, Mayer et al., 2008) and thus phages, either whole or phage-encoded enzymes, can inform approaches to combat serious infectious bacteria.

5.2.2 **Phage packaging**

Large terminase protein sequences are found to be another useful marker that researchers have employed to construct phylogeny and decode evolutionary relationships among phages belonging to different families (Casjens et al., 2005). Phages in common have established diverse approaches for packaging DNA into their capsids, reliant on their mechanism of replication, head size and its consequent capacity to hold a few more than one whole phage genome, which frequently result in circularly permutated genomes (Thanki, 2016). The DNA packaging process usually includes three components: the portal protein that threads the viral DNA through the capsid, the large terminase motor protein that generates power for packaging and the small terminase that initiates packaging (Rao and Feiss, 2008). By achieving coordination of

the three components, phage DNA is translocated via the headful packaging system, in which the capsid is filled to full capacity.

In general, most tailed dsDNA bacteriophages use a similar DNA packaging mechanism; however their DNA replication strategies and the resulting nature of ends of the packaged DNAs are not the same for all. There are six well-studied types of virion chromosome ends, which are classified by the presence of (i) single-stranded cohesive ends, (ii) circularly permuted direct terminal repeats, (iii) short, several hundred base pairs exact (non-permuted) direct terminal repeats, (iv) long, several thousand base pairs exact (non-permuted) direct terminal repeats, (v) terminal host DNA sequences, and (vi) covalently bound terminal proteins (Casjens and Gilcrease, 2009). DNA packaging styles other than those listed above are possible, as phages have not been considered to the point that we have classified all of their molecular lifestyles in detail.

5.2.3 ***PaLoc* genes**

So far no close homologs of the *C. difficile* toxin genes have been found in any of the phage genomes, other than an ORF found in phiC2 that has a low level of homology to *tcdB* at the amino acid (aa) level (Goh et al., 2007). This ORF is located within a proposed lysogenic conversion module, on the antisense-strand. Homologs are also found in the other MM and LTMs with the variation in the location and gene content (Goh et al., 2007, Hargreaves and Clokie, 2014). Sequence similarity of the holin gene to the *tcdE* gene encoded on the *PaLoc* is also found in phiC2 genome, which leads researchers to propose that it may have a phage origin or have been transferred via phages (Goh et al., 2007). This hypothesis of phage mediated bacterial gene transfer is further supported by the identification of a homolog of *cdu1*, Orf 46, which also encodes a Penicillinase R protein domain in the genome of phiC2 (Goh et al., 2007). The ORF is located downstream of the lysogenic conversion module, on the sense strand, and may have a regulatory function. Homologs are also found in other myoviruses such as phiCD119, phiCDHM1 and phiMMP02 (Hargreaves et al., 2014b, Mayer et al., 2008, Meessen-Pinard et al., 2012). The impact of these genes is not known, and it is also not clear whether they would have the same role as the genes they are shown to be similar too (Hargreaves and Clokie, 2014).

5.2.4 **Genes associated with *C. difficile* defence: Abortive infection**

Another unusual feature in phage genomes is the presence of the host gene, an ORF with a predicted accessory function, Orfs37, which has sequence similarity to AbiF found in phiC2 genome (Goh et al., 2007). AbiF is a predicted phage abortive infection protein (Labrie et al., 2010), similarity to this gene is not detected in any of the other *C. difficile* phages. *Abi* genes are usually encoded by the bacterial hosts, so it is usually carried on a plasmid and confers phage resistance to bacteria, causing an abortive phage infection at the level of phage DNA replication, although the mechanism by which AbiF performs its function is not fully known (Goh et al., 2007, Hargreaves and Clokie, 2014). The benefit of carrying such genes is unclear and has not been identified on other phage genomes; it has been proposed that they might prevent secondary phage infection or control phiC2 replication (Hargreaves and Clokie, 2014).

5.2.5 **Genes associated with *C. difficile* defence: CRISPR/Cas system**

The isolation of novel phages is also pertinent when considering the ability of bacteria to evolve phage resistance. Bacteria have a multitude of ways to prevent phage infections, but little is known about resistance in *C. difficile* and its phages. The best characterized mechanism is arguably its CRISPR/Cas system (Boudry et al., 2015, Hargreaves et al., 2014a, Sebaihia et al., 2006). These clustered regularly interspaced short palindromic repeats (CRISPRs) and the CRISPR-associated (cas) genes have been described in many prokaryote genomes (Barrangou et al., 2007), and provide interference to incoming phage DNA. *C. difficile* strains appear to have a functioning and diverse repertoire of CRISPR/Cas elements (Boudry et al., 2015, Hargreaves et al., 2014a, Soutourina et al., 2013). They include CRISPR arrays carried on mobile genetic elements (MGEs), including prophages with implications for secondary phage resistance and the spread of resistance between strains on these MGEs (Boudry et al., 2015, Hargreaves et al., 2014a).

5.3 Aims of this chapter

1 Select the most interesting *C. difficile* phages from Chapter 4and sequence them to determine genetic structure and content. As we are particularly interested in phages that could be used in therapy we chose three phages that infected clinically relevant strains/ribotypes.

2 Determine the phylogenetic relationship with other *C. difficile* phages isolated from different geographical areas by constructing phylogenetic trees based whole genome sequencing and specific genes such as the endolysin gene and TerL gene.

5.4 Results

5.4.1 Genomic annotation of Myoviruses

5.4.1.1 Genome of the long tailed myovirus CDKM9

Phage CDKM9 is a temperate long tailed myovirus with a dsDNA genome isolated with CD105HE1 from R076. This phage was selected as it is highly unusual for a myovirus to target R027. Assembly of the sequencing data resulted in a contig 49,822bp long. Seventy five ORFs were predicted, the majority of the ORFs (66/75) are encoded on the sense DNA strand (Figure 5-1 and Appendix 11). The ORFs on the antisense strand are mainly clustered between Orfs 41-44 and encode for the integrase gene, a transcriptional regulator gene, and those between Orfs 35-37 encode uncharacterised proteins. Of the 75 ORFs, 32 (42.7%) could be assigned putative functions and 43 (57.3%) have similarity to proteins with uncharacterised functions in bacteria or phages. No obvious termini could be located in the genome.

5.4.1.2 Detailed Analysis of the whole genome

The modular organisation of the CDKM9 genome follows the well-known arrangement that is described in previous *C. difficile* phages and other temperate phages infecting low GC bacteria (Sekulovic et al., 2011, Goh et al., 2007, Horgan et al., 2010, Mayer et al., 2008). This include gene modules with putative products that are involved in DNA packaging, virion assembly, host cell lysis, lysogeny control and DNA replication Figure 5-1.

Predictably Orfs 1 and 2 code for the terminase enzymes that homologous to terminase small subunit proteins in phiCD27 and *Clostridium* species in the case of the terminase large subunit protein (Appendix 11). Terminase proteins are required for packaging of the phage genomic DNA into the preassembled empty capsid shells. (Rentas and Rao, 2003). Orf 4 and 5 show a sequence similarity to phage portal proteins and head morphogenesis in multiple species of *Clostridium* and *C. difficile* respectively. Portal proteins are found that may participate in procapsid assembly during phage morphogenesis (Govind et al., 2006).

Similar to other double stranded DNA bacteriophages, a cluster of genes located in the tail morphogenesis regions encode for the component structures such as tail sheath, LysM, base plate J family protein, tail protein with tail fibre protein, and Orf 17 is the

largest ORF in CDKM9 and is the predictable code for tail tape measure protein, which is assumed to determine tail length in tailed phages (Katsura, 1987). In addition, 4 ORFs display sequence similarities to genes XkdN, XkdS , XkdX and XkdM , the latter found to code for the core tail (Govind et al., 2006).

The lysis module is located between the structural module and the lysogeny module and consists of Orfs 31 and 32 that encode for holin and *N*-acetylmuramoyl-L-alanine amidase. The module is typical in the comparative positions of the cell lysis genes as these are usually next to each other. Both are responsible for cell lysis and the release of phage progeny. Most double-stranded DNA phages need the combination of a holin and an endolysin to attain host lysis (Govind et al., 2006). The putative holin gene is found to be highly conserved in *C. difficile* phages and aligns well with the holin of another phage (Shan et al., 2012).

The noticeable feature of the lysogeny module is the presence of integrase gene encoded by Orf 41 with homology to the integrase gene in phiC2. Interestingly, the gene is located in between two copies of Orfs 39 and 40 downstream that encode for BlaI/MecI/CopY family transcriptional regulator and Orfs 43 and 44 upstream that encode for XRE family transcriptional regulator, which likely play a role in the maintenance of lysogeny (Hargreaves and Clokie, 2014), These are followed by ORFs 45 and 46 with putative repressor and anti-repressor respectively.

Orfs 62, 63, 64 and 75 in the DNA replication, transcription and gene regulation model displayed sequence similarity to genes in *C. difficile* phages phiCD27, phiCDHM19, phiC2, and CDHM1 respectively. In addition, Orf 73 encodes putative KilA-N/ORF6 family anti-repressor homologues to multispecies of *Clostridium.*

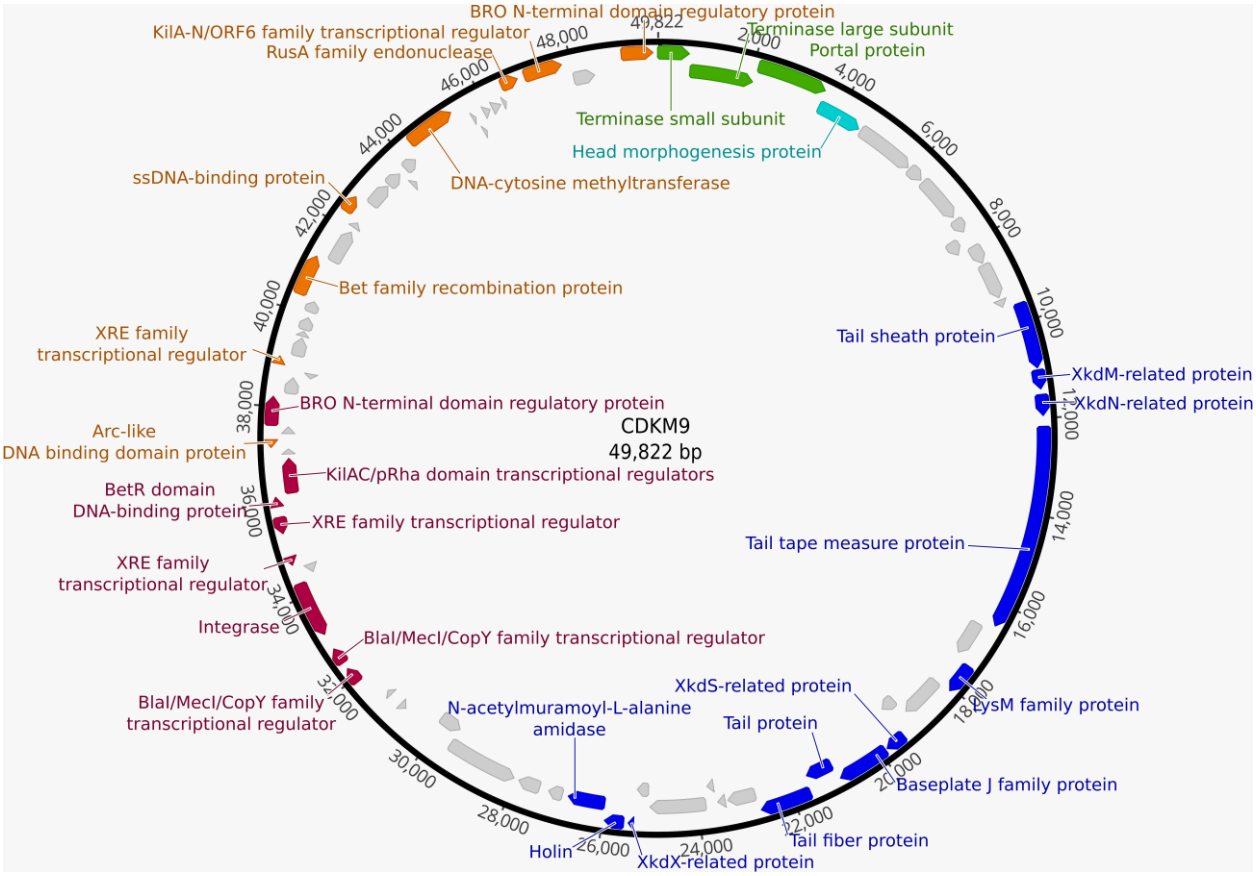


Figure 5-1. Genome map of CDKM9.

Circular double stranded genome map and modular regions of the genome. The predicted Orfs are marked with arrows. Different colours of genes indicate functional modules: head packaging (green); aquamarine (head) tail and lysis (blue); lysogenic conversion (purple), and DNA replication (orange). Light grey are ORFs with no predicted function assigned.

5.4.2 **Genomic annotation of the long tailed myovirus CDKM15**

Based on genome analysis, CDKM15 is a dsDNA molecule. Assembly of the sequencing data resulted in two contigs that were joined into one contig by PCR as described in section 2.4.4, which has a length of 50,605bp. The phage has average GC contents of 28.98%. Analogous to CDKM9, the physical terminus of the genome was not identified, even though searched for by PCR using specific designed primers. In total, CDKM15 had 79 ORFs predicted, with 73 on the sense strand and 6 on the antisense. Comparison against Intproscan, NCBI conserved domain and Protein BLAST databases enabled the assignment of a predictive function of 34 ORFs in 79

(43%) (with a BLASTp e-value of 1e-10 as a cut-off). The 45 ORFs (57 %) in the genome code for unknown function. No tRNA genes were identified in the genome,

In general the genome organisation of CDKM15 appeared to be classical and closely related to CDKM9, with a cluster of genes coding for the distinct functional modules involved in DNA packaging, virion structure and assembly, host cell lysis, lysogeny control and DNA replication. The genome displays similar mosaicism at the sequence level to genomes of other *C. difficile* phages across the main functional modules based on the top BLASTp matches (Appendix 12). Five ORFs share similarity with the phiC2, phiMMP02, phiCD505 and multispecies of *Clostridium* as well as *C. difficile*. Similar to CDKM9, it also has an integrase gene in the lysogeny control module, located between two copies of repR a predicted regulatory protein and a penicillinase repressor family protein domain (Pfam PF03965) (Figure 5-2 and Appendix 12).

Although the characteristic feature of genome CDMK15 is that it carries a predicted CRISPR array, the presence of arrays in *C. difficile* prophages and released virus particles has been described previously (Hargreaves et al., 2014a, Minot et al., 2013, Sebaihia et al., 2006). The two prophages of CD630 both contain CRISPR arrays and can be propagated through lytic infections; however, whether the CRISPR arrays were maintained during these cycles was unknown (Goh et al., 2007). The presence of this array in the genome of CDMK15 suggests such arrays are stable. The array consists of consensus direct repeats sequence, spacer content, and adjacent coding DNA sequences (CDSs). The CRISPR array is located in the structural region of the phage genomes, flanked by CDS encoding a protein containing a Bro N-terminal domain (Pfam 02498) and a CDS encoding a tail tape measure protein Figure 5-2. The array contains 6 spacers, 3 of which have perfect matches to the genome sequence of phage phiCD6356, 1 has an imperfect match to phiCD505 sequence (97% identity) and 1 has an imperfect match to phiMMP02 (95% identity). A putative PAM motif, CCT, was identified for each detected protospacer (advanced paper. In addition the phage does not encode any predicted Cas genes consistent with previous descriptions of these phage carried arrays (Hargreaves et al., 2014a).

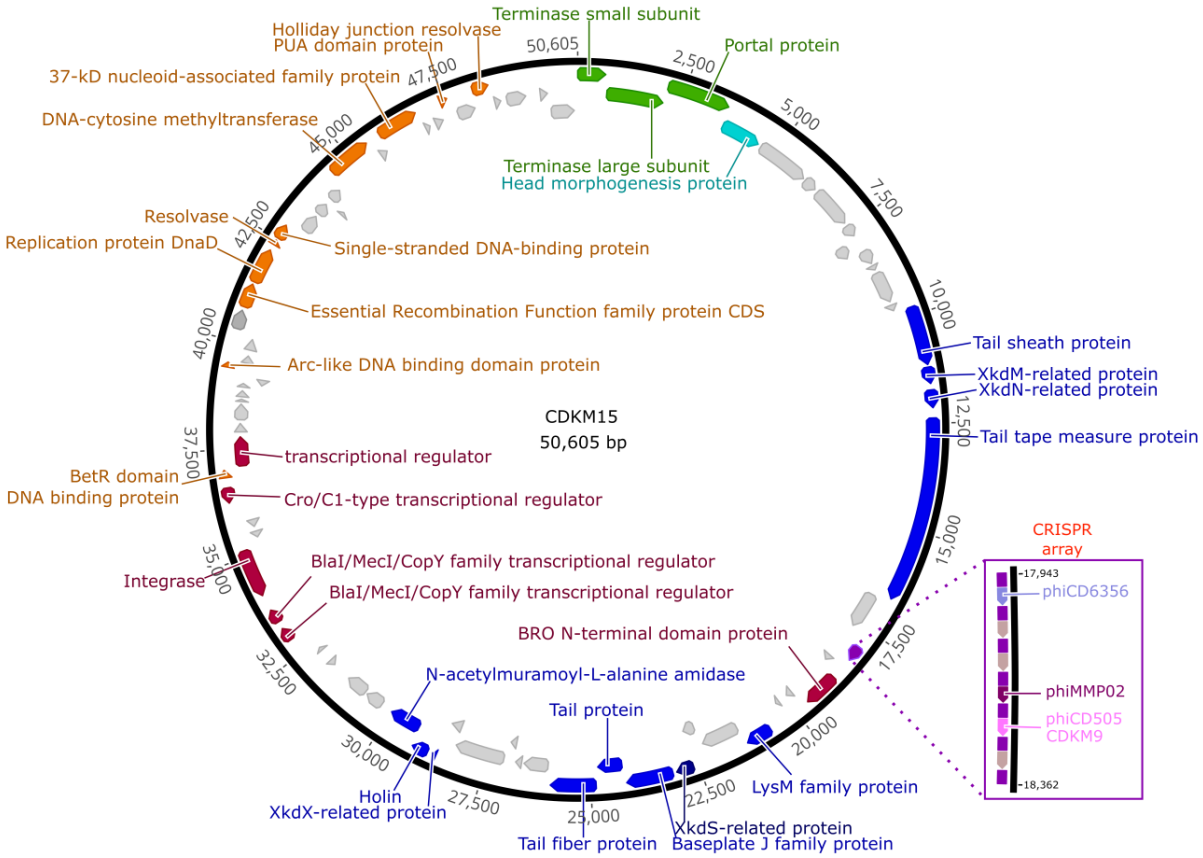


Figure 5-2. Genome organization of CDKM15.

Shows a circular double stranded genome map and modular regions of the genome. The predicted ORFs are marked with arrows. Different colours of genes indicate functional modules: head packaging (green); aquamarine (head) tail and lysis (blue); lysogenic conversion (purple), DNA replication, (orange) and CRISPR array (red). Light grey are CDSs with no predicted function assigned.

1.3.1 Genomic annotation of siphoviruses CDKS6, CDKS7 and CDKS8

CDKS6, CDKS7 and CDKS8 are siphoviruses with dsDNA with a genome in the range of 40 kb to 45 kb. The phages were isolated from enriched environmental samples on different hosts (AIU, K18 and HB) of the same ribotype R027. Whole genome sequencing resulted in the assembly of one contig for each phage. The results from the sequencing and annotation have shown that the siphoviruses CDKS6, CDKS7 and CDKS8 are almost identical (99.9 to 100%). It is commonly asked whether it is possible to isolate identical phages in nature; so far no two independently isolated phages have been found to have identical genome sequences (Abedon, 2008), as phages are considered the most common diverse entities on the biosphere in part due to their

abundance as well as the fluid nature of their genomes (Breitbart et al., 2002). However, and in keeping with the similarity of our phages, we noticed a highly level of similarity (99.9 to 100%) on the translated comparison and on the nucleotide level between the “jumbo" myoviruses phiCD211 and phiCDIF1296T based on the genome comparison we performed using Gegenees 2.2.1 (Figure 5-6 and Figure 5-7). This opens the question of whether it is possible to isolate identical phages in nature, which is a large challenge, and extensive investigation needs to be performed on these phages to show if they are independent phages in nature. Due to the time limit of the current study of research this is to be a consideration for future work.

Due to the genetic identity of the sequenced siphoviruses we analysed the genome of CDKS8 to represent the siphoviruses. CDKS8 has a genome of 41,639 bp length with a GC content of 30.8% that agrees well with the phiCD38-2 (30.83%). It is slightly larger but in the range with other sequenced *C. difficile* phages (28.4% -29.4%) (Goh et al., 2007, Govind et al., 2009, Mayer et al., 2008, Hargreaves et al., 2014b), and no tRNA genes were identified by tRNA scan-SE, which is consistent with the other published *C. difficile* phage genomes with the exception of the jumbo virus φCDIF1296T, as analysis with tRNAscan-SE showed one predicted tRNA for serine (Wittmann et al., 2015).

Overall the genome encodes for 53 of predicted ORFs (Appendix 13). Predicted coding sequences account for 87% of the genome and the majority of ORFs (45/53) are encoded on the sense DNA strand. Of the 53 ORFs, 24 (45%) have been assigned putative functions based on either sequence homology and/or identification of a protein domain, and 29 (54%) are assigned as hypothetical proteins with unknown functions (Figure 5-3). Based on the genome sequence similarity, CDKS8 is closely related to that of the other *C. difficile* siphoviruses, as noticed from the entries in the top BLASTp hits. However, the genome similarities differ on an individual gene basis; for example there are 13 ORFs with homology to genes in phiCD111, 11 ORFs have homology with phiCD146 and 5 with phiCD38-2. In addition, there are two ORFs with homology to *C. difficile* myoviruses phiCDHM11 and phiCDHM13, which suggests that the genome is mosaic.

A notable feature in the genome is that the presence of Orf15 in the structural regions encodes for bacteriophage GP15 family protein, a membrane protein with super

infection exclusion activity, which so far has been identified in coliphages HK97 (Sekulovic and Fortier, 2015). Studies have shown that temperate phage HK97 encodes for this moron element within its tail morphogenesis region, and they demonstrate that the HK97 prophage provides the host cell with resistance to infection by phages HK97 and closely related phages, an addition that has not been found in other related phages (Cumby et al., 2012, Sekulovic et al., 2015)

Usually, the lysogeny module is located between the cell lysis cassette and the DNA replication and regulation module. In general, repressors such as Cro/C1, transcriptional regulators, and antirepressors, as well as the integrase are encoded in this region. Of note is that the lysogeny module could not be defined clearly in CDKS8. Instead an uncharacteristic organisation of the lysogeny module was observed, in which integrase gene Orf 51and a putative phage repressor Orf 50 with Orf 37 Cro/C1 repressor were found scattered between DNA replication and regulation genes. This was detected only for two siphoviruses, phiCD38-2 and phiCD6356, and a myoviruse, phiMMP04 (Sekulovic et al., 2011, Horgan et al., 2010, Meessen-Pinard et al., 2012). Orf 31, located on the antisense strand, was found to have similarity to a putative plasmid partition protein, parA, and a domain search also indicated a ParA/ParB protein (NCBI conserved domain accession number: cd02042). ParA has been described in several *C. difficile* phages, usually functioning in association with another portioning protein (Hargreaves and Clokie, 2015, Horgan et al., 2010, Sekulovic et al., 2011).

In the predictive region of DNA replication, recombination and modification, only a few ORFs could be consigned a function. Based on a BLAST search Orfs 35, 40, 47 and 53 were predictively recognised as Helix DNA-binding protein, replication terminator like protein, single-stranded DNA-binding protein and recombinase/resolvase respectively.

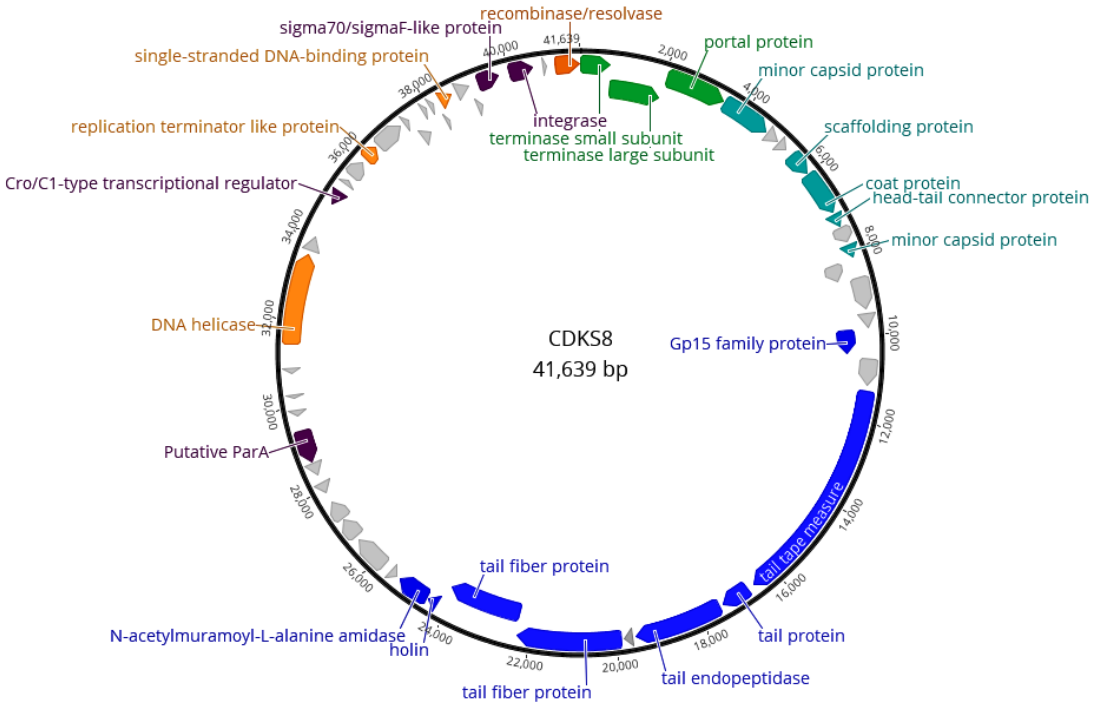


Figure 5-3. Genome organization of CDKS8.

Shows a circular double stranded genome map and modular regions of the genome. The predicted ORFs are marked with arrows. Different colours of genes indicate functional modules: head packaging (green); aquamarine (head) tail and lysis (blue); lysogenic conversion (purple), DNA replication (orange) and light grey are ORFs with no predicted function assigned.

1.3.6 Comparative genomics for CDKM9 and CDKM15

CDKM9 and CDKM15 were subjected to further analysis of genome comparison as preliminary data showed extensive mosaic structure of their genome. In order to visualise the patterns of sequence similarity across the genomes of phages CDKM9 and CDKM15, each were used as a reference to compare with 22 *C. difficile* phage genomes in BRIG 0.95 (Blast Ring Image Generator) (Goujon et al., 2010) (Figure 5-4 and Figure 5-5). This genome comparison tool uses BLAST to align sequences and calculates global pairwise similarity from the similarity and frequency of the local alignments (Alikhan et al., 2011).

In both analyses, the phages display the greatest level of sequence similarity to the three LTMs, (phiCD505, phiCD27 and phiMMP02). This is located throughout the packaging and structural modules. The similarity is lower in regions involved in lysogeny control and DNA replication, and these instead share similarity to sequences

of the medium myoviruses (phiC2, phiCDHM1, phiMMP03, phiMMP01, phiCDHM19 and phiCD119). As might be expected, the genetic variability of the modules varied to different degrees; the lysogeny control region is divergent across the genomes, while the cluster of genes responsible for lysis is more conserved across the family level. Notably, there is a region in the CDMK15 genome with no similarity to the other phage sequences and this is where the CRISPR array is located (Goujon et al., 2010) (Figure 5-5).

In addition to the mosaicism evident between the myoviruses are the conserved sequences present in a subset of genes across all phage genomes. The shared sequences are predominantly located in the endolysin gene as well as genes located in the lysis and attachment module of the phage genome. This suggests there may be sequences specific for interacting with the *C. difficile* cell surface or membrane proteins

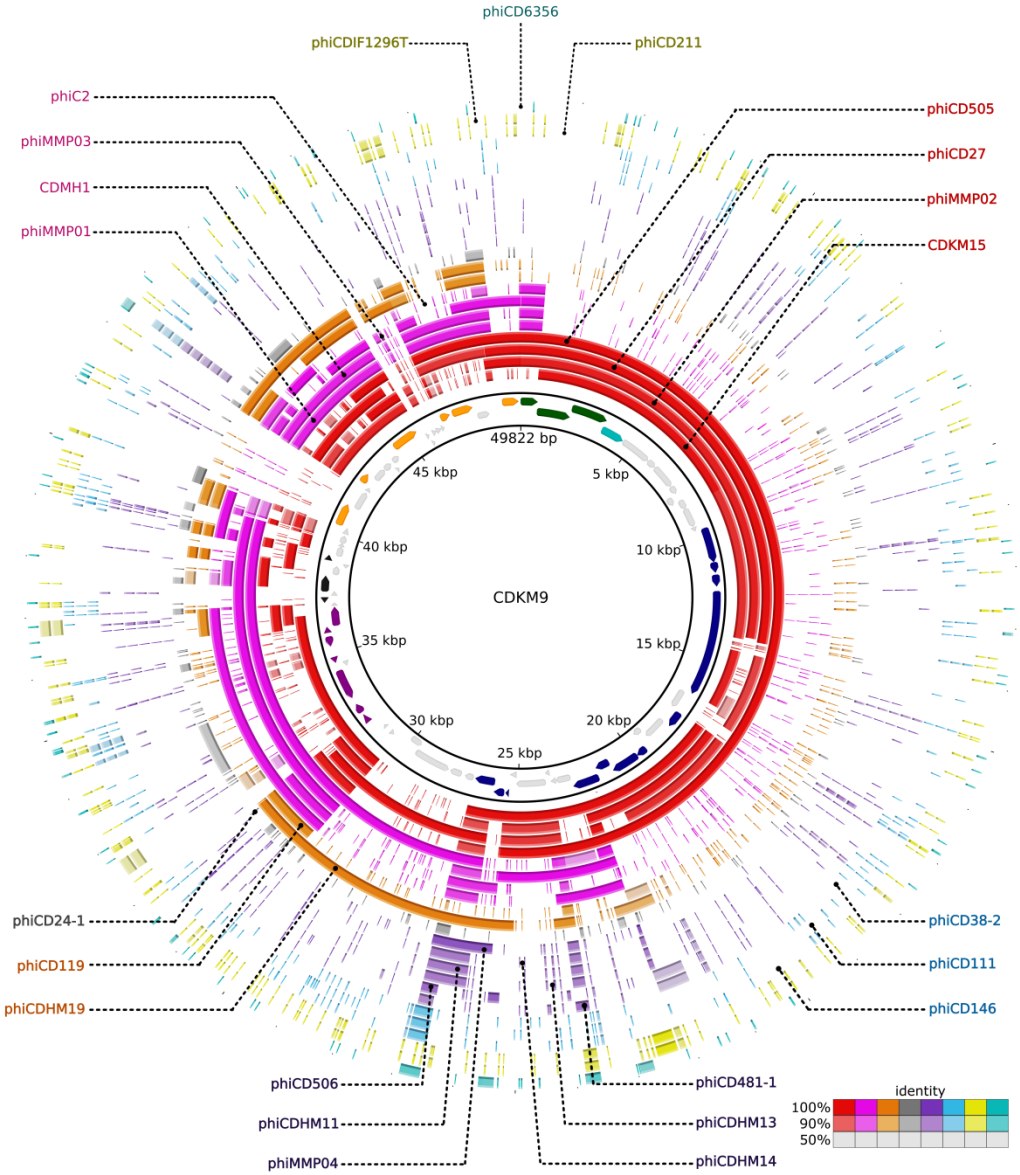


Figure 5-4. Whole genome comparison of phage CDKM9 and other *C. difficile* phages.

The local similarity of each phage is calculated based on BLASTn high scoring pairs and plotted against circular map of reference genome represented as the inner circle (in this case genome of CDKM9). Similarity to each of the 23 other *C. difficile* phages is shown as coloring intensity of consecutive rings.

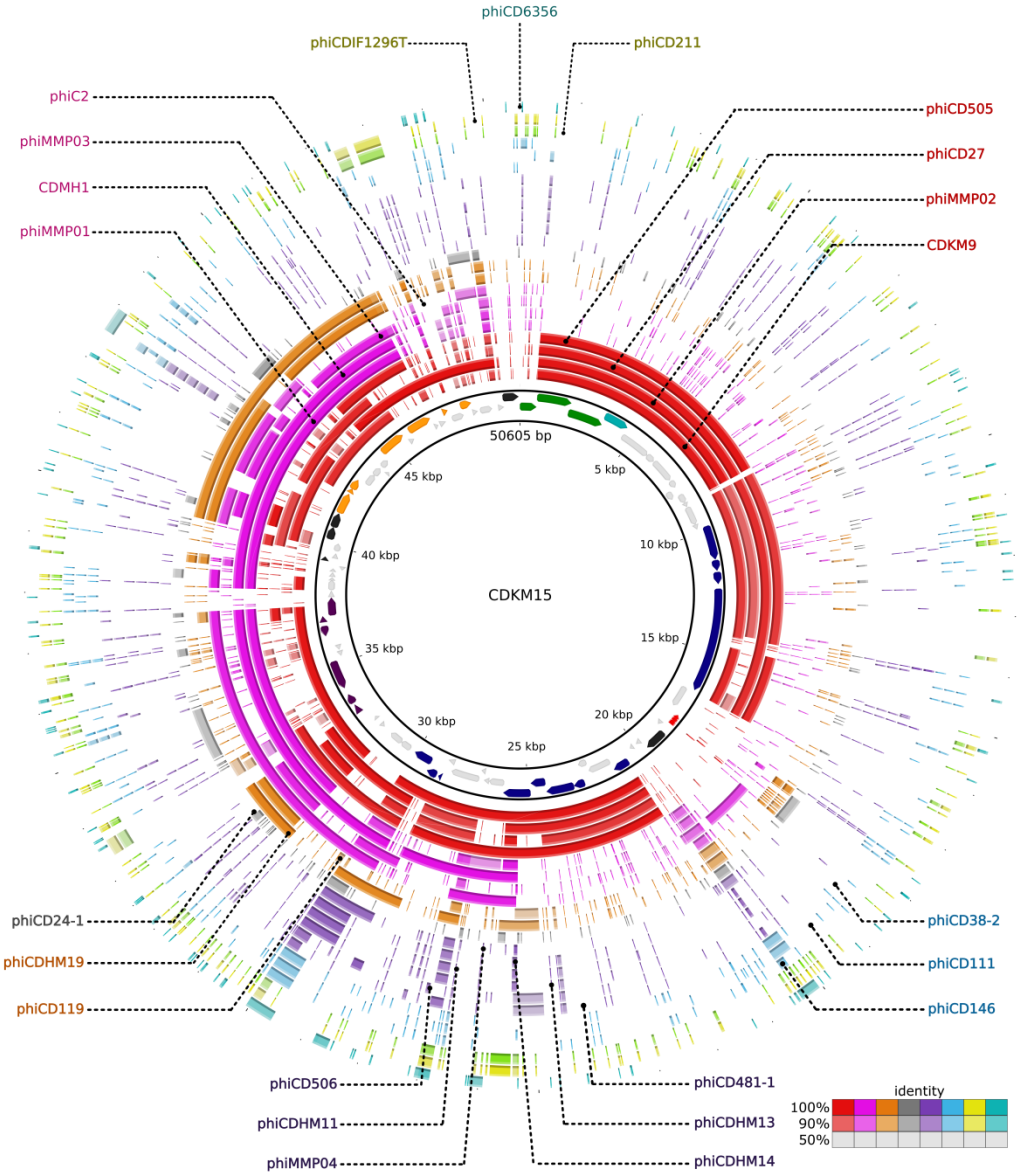
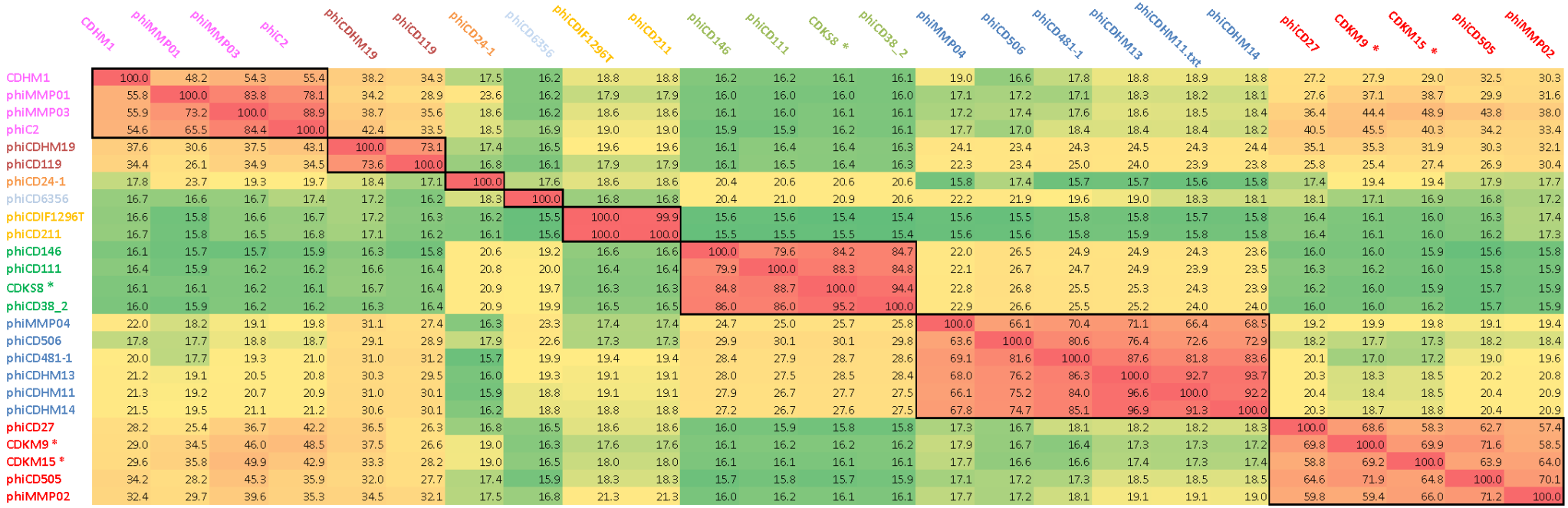


Figure 5-5. Whole genome comparison of phage CDKM15 and other *C. difficile* phages.

The local similarity of each phage is calculated based on BLASTn high scoring pairs and plotted against circular map of reference genome represented as the inner circle (in this case genome of CDKM15). Similarity to each of the 23 other *C. difficile* phages is shown as coloring intensity of consecutive rings.

5.4.3 **Phylogenetic analysis of CDKS8, CDKM9 and CDKM15 and published *C. difficile* phages**

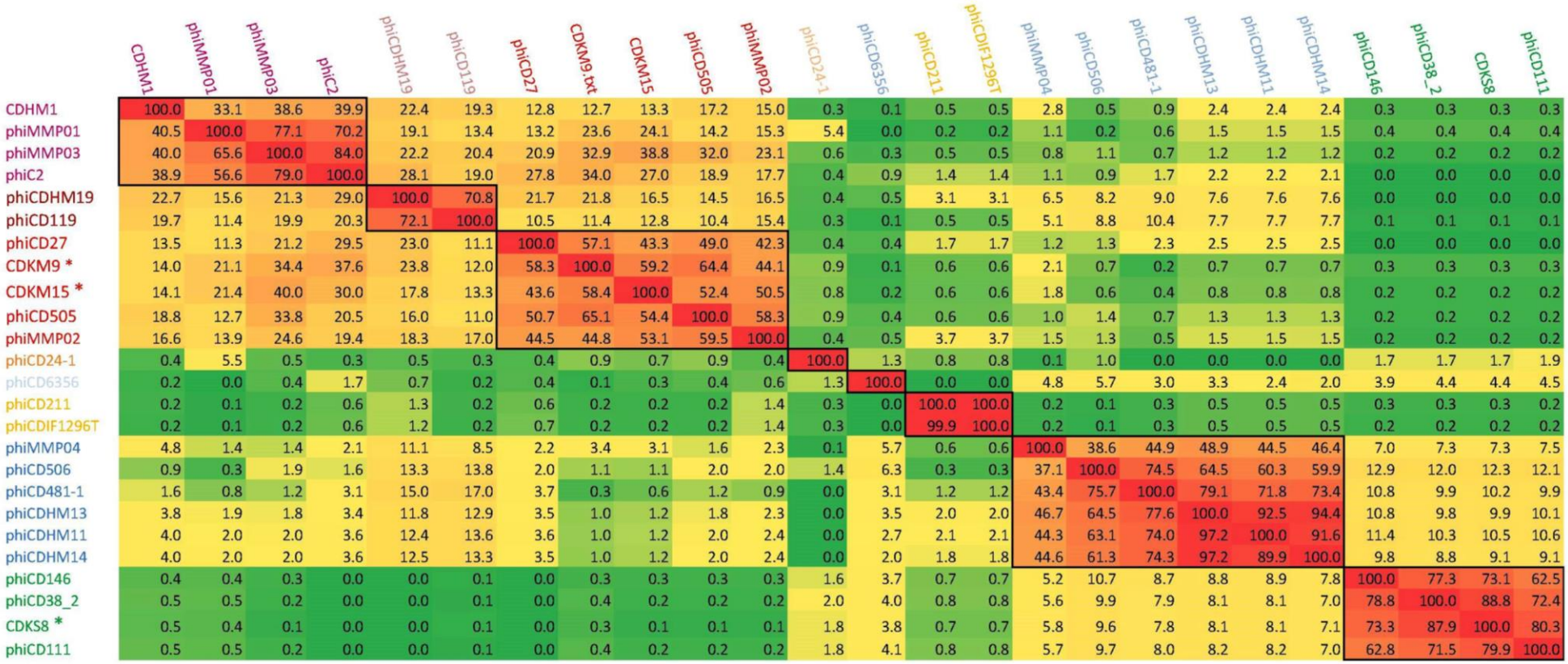
Phage whole genome comparisons, are a useful tool to classify new phages (Ågren et al., 2012). We constructed a phylogenetic tree from whole genome based on pairwise similarity scoring between CDKS8, CDKM9 and CDKM15 and 22 *C. difficile* phages (Appendix1-6) using Gegenees 2.0.0 (tBLASTx method, fragment size –200, step size –100)(Ågren et al., 2012), Figure 5-6 and Figure 5-7. Based on the result of similarity matrices, a Bio NJ phylograms generated using Splits Tree 4.13.1 (Huson and Bryant, 2006) (Figure 5-8). The tree showed the CDKM9 and CDKM15 group with the long tailed myoviruses, phiCD27, phiCD505 and phiMMP02. The tree also showed clearly a distant evolutionary relationship between the two clusters of medium myoviruses (one of them corresponding to currently accepted genus phicd1119virus), but overall similarity between the genomes of these groups fell to less than 40% on the nucleotide level and 50% for translated comparison (Figure 5-6 and Figure 5-7). The “jumbo" myoviruses (phiCD211 and phiCDIF1296T) were even more divergent and could not be classified into any of these groups. As expected CDKS8 groups with the siphovirues phiCD38-2, phiCD146 and phiCD111, but sub-clusters were observed for the phages could be split into phiCD38-2 group (CDKS8, phiCD111 and phiCD146), phiCD24-1 and phiCD6356.



	CDHM1	phiMMP01	phiMMP03	phiC2	phiCDHM19	phiCD119	phiCD24-1	phiCD6356	phiCDIF1296T	phiCD211	phiCD146	phiCD111	CDKS8 *	phiCD38_2	phiMMP04	phiCD506	phiCD481-1	phiCDHM13	phiCDHM11.txt	phiCDHM14	phiCD27	CDKM9 *	CDKM15 *	phiCD505	phiMMP02
CDHM1	100.0	48.2	54.3	55.4	38.2	34.3	17.5	16.2	18.8	18.8	16.2	16.2	16.1	16.1	19.0	16.6	17.8	18.8	18.9	18.8	27.2	27.9	29.0	32.5	30.3
phiMMP01	55.8	100.0	83.8	78.1	34.2	28.9	23.6	16.2	17.9	17.9	16.0	16.0	16.0	16.0	17.1	17.2	17.1	18.3	18.2	18.1	27.6	37.1	38.7	29.9	31.6
phiMMP03	55.9	73.2	100.0	88.9	38.7	35.6	18.6	16.2	18.6	18.6	16.1	16.0	16.1	16.1	17.2	17.4	17.6	18.6	18.5	18.4	36.4	44.4	48.9	43.8	38.0
phiC2	54.6	65.5	84.4	100.0	42.4	33.5	18.5	16.9	19.0	19.0	15.9	15.9	16.2	16.1	17.7	17.0	18.4	18.4	18.4	18.2	40.5	45.5	40.3	34.2	33.4
phiCDHM19	37.6	30.6	37.5	43.1	100.0	73.1	17.4	16.5	19.6	19.6	16.1	16.4	16.4	16.3	24.1	23.4	24.3	24.5	24.3	24.4	35.1	35.3	31.9	30.3	32.1
phiCD119	34.4	26.1	34.9	34.5	73.6	100.0	16.8	16.1	17.9	17.9	16.1	16.5	16.4	16.3	22.3	23.4	25.0	24.0	23.9	23.8	25.8	25.4	27.4	26.9	30.4
phiCD24-1	17.8	23.7	19.3	19.7	18.4	17.1	100.0	17.6	18.6	18.6	20.4	20.6	20.6	20.6	15.8	17.4	15.7	15.7	15.6	15.8	17.4	19.4	19.4	17.9	17.7
phiCD6356	16.7	16.6	16.7	17.4	17.2	16.2	18.3	100.0	16.8	16.8	20.4	21.0	20.9	20.6	22.2	21.9	19.6	19.0	18.3	18.1	18.1	17.1	16.9	16.8	17.2
phiCDIF1296T	16.6	15.8	16.6	16.7	17.2	16.3	16.2	15.5	100.0	99.9	15.6	15.6	15.4	15.4	15.6	15.5	15.8	15.8	15.7	15.8	16.4	16.1	16.0	16.3	17.4
phiCD211	16.7	15.8	16.5	16.8	17.1	16.2	16.1	15.6	100.0	100.0	15.5	15.5	15.5	15.4	15.6	15.6	15.8	15.9	15.8	15.8	16.4	16.1	16.0	16.2	17.3
phiCD146	16.1	15.7	15.7	15.9	16.3	15.8	20.6	19.2	16.6	16.6	100.0	79.6	84.2	84.7	22.0	26.5	24.9	24.9	24.3	23.6	16.0	16.0	15.9	15.6	15.8
phiCD111	16.4	15.9	16.2	16.2	16.6	16.4	20.8	20.0	16.4	16.4	79.9	100.0	88.3	84.8	22.1	26.7	24.7	24.9	23.9	23.5	16.3	16.2	16.0	15.8	15.9
CDKS8 *	16.1	16.1	16.2	16.1	16.7	16.4	20.9	19.7	16.3	16.3	84.8	88.7	100.0	94.4	22.8	26.8	25.5	25.3	24.3	23.9	16.2	16.0	15.9	15.7	15.9
phiCD38_2	16.0	15.9	16.2	16.2	16.3	16.4	20.9	19.9	16.5	16.5	86.0	86.0	95.2	100.0	22.9	26.6	25.5	25.2	24.0	24.0	16.0	16.0	16.2	15.7	15.9
phiMMP04	22.0	18.2	19.1	19.8	31.1	27.4	16.3	23.3	17.4	17.4	24.7	25.0	25.7	25.8	100.0	66.1	70.4	71.1	66.4	68.5	19.2	19.9	19.8	19.1	19.4
phiCD506	17.8	17.7	18.8	18.7	29.1	28.9	17.9	22.6	17.3	17.3	29.9	30.1	30.1	29.8	63.6	100.0	80.6	76.4	72.6	72.9	18.2	17.7	17.3	18.2	18.4
phiCD481-1	20.0	17.7	19.3	21.0	31.0	31.2	15.7	19.9	19.4	19.4	28.4	27.9	28.7	28.6	69.1	81.6	100.0	87.6	81.8	83.6	20.1	17.0	17.2	19.0	19.6
phiCDHM13	21.2	19.1	20.5	20.8	30.3	29.5	16.0	19.3	19.1	19.1	28.0	27.5	28.5	28.4	68.0	76.2	86.3	100.0	92.7	93.7	20.3	18.3	18.5	20.2	20.8
phiCDHM11	21.3	19.2	20.7	20.9	31.0	30.1	15.9	18.8	19.1	19.1	27.9	26.7	27.7	27.5	66.1	75.2	84.0	96.6	100.0	92.2	20.4	18.4	18.5	20.4	20.9
phiCDHM14	21.5	19.5	21.1	21.2	30.6	30.1	16.2	18.8	18.8	18.8	27.2	26.7	27.6	27.5	67.8	74.7	85.1	96.9	91.3	100.0	20.3	18.7	18.8	20.4	20.9
phiCD27	28.2	25.4	36.7	42.2	36.5	26.3	16.8	16.5	18.6	18.6	16.0	15.9	15.8	15.8	17.3	16.7	18.1	18.2	18.2	18.3	100.0	68.6	58.3	62.7	57.4
CDKM9 *	29.0	34.5	46.0	48.5	37.5	26.6	19.0	16.3	17.6	17.6	16.1	16.2	16.2	16.2	17.9	16.7	16.4	17.3	17.3	17.2	69.8	100.0	69.9	71.6	58.5
CDKM15 *	29.6	35.8	49.9	42.9	33.3	28.2	19.0	16.5	18.0	18.0	16.1	16.1	16.1	16.1	17.7	16.6	16.6	17.4	17.3	17.4	58.8	69.2	100.0	63.9	64.0
phiCD505	34.2	28.2	45.3	35.9	32.0	27.7	17.4	15.9	18.3	18.3	15.7	15.8	15.7	15.9	17.1	17.2	17.3	18.5	18.5	18.5	64.6	71.9	64.8	100.0	70.1
phiMMP02	32.4	29.7	39.6	35.3	34.5	32.1	17.5	16.8	21.3	21.3	16.0	16.2	16.1	16.1	17.7	17.2	18.1	19.1	19.1	19.0	59.8	59.4	66.0	71.2	100.0

Figure 5-6. Heatplot showing the result of translated genome comparison of *C. difficile* phage.

The similarity value were considered using Gegenees software based on pair wise translated comparison of the analysed sequences (tBLASTX method, fragment size 200/100. The heat map colours reflect sequence similarity, ranging from low (green) to high (red). The new isolated phages are labelled with the star.



	CDHM1	phiMMP01	phiMMP03	phiC2	phiCDHM19	phiCD119	phiCD27	CDKM9.txt	CDKM15	phiCD505	phiMMP02	phiCD24-1	phiCD6356	phiCD211	phiCDIF1296T	phiMMP04	phiCD506	phiCD481-1	phiCDHM13	phiCDHM11	phiCDHM14	phiCD146	phiCD38_2	CDKS8	phiCD111
CDHM1	100.0	33.1	38.6	39.9	22.4	19.3	12.8	12.7	13.3	17.2	15.0	0.3	0.1	0.5	0.5	2.8	0.5	0.9	2.4	2.4	2.4	0.3	0.3	0.3	0.3
phiMMP01	40.5	100.0	77.1	70.2	19.1	13.4	13.2	23.6	24.1	14.2	15.3	5.4	0.0	0.2	0.2	1.1	0.2	0.6	1.5	1.5	1.5	0.4	0.4	0.4	0.4
phiMMP03	40.0	65.6	100.0	84.0	22.2	20.4	20.9	32.9	38.8	32.0	23.1	0.6	0.3	0.5	0.5	0.8	1.1	0.7	1.2	1.2	1.2	0.2	0.2	0.2	0.2
phiC2	38.9	56.6	79.0	100.0	28.1	19.0	27.8	34.0	27.0	18.9	17.7	0.4	0.9	1.4	1.4	1.1	0.9	1.7	2.2	2.2	2.1	0.0	0.0	0.0	0.0
phiCDHM19	22.7	15.6	21.3	29.0	100.0	70.8	21.7	21.8	16.5	14.5	16.5	0.4	0.5	3.1	3.1	6.5	8.2	9.0	7.6	7.6	7.6	0.0	0.0	0.0	0.0
phiCD119	19.7	11.4	19.9	20.3	72.1	100.0	10.5	11.4	12.8	10.4	15.4	0.3	0.1	0.5	0.5	5.1	8.8	10.4	7.7	7.7	7.7	0.1	0.1	0.1	0.1
phiCD27	13.5	11.3	21.2	29.5	23.0	11.1	100.0	57.1	43.3	49.0	42.3	0.4	0.4	1.7	1.7	1.2	1.3	2.3	2.5	2.5	2.5	0.0	0.0	0.0	0.0
CDKM9 *	14.0	21.1	34.4	37.6	23.8	12.0	58.3	100.0	59.2	64.4	44.1	0.9	0.1	0.6	0.6	2.1	0.7	0.2	0.7	0.7	0.7	0.3	0.3	0.3	0.3
CDKM15 *	14.1	21.4	40.0	30.0	17.8	13.3	43.6	58.4	100.0	52.4	50.5	0.8	0.2	0.6	0.6	1.8	0.6	0.4	0.8	0.8	0.8	0.2	0.2	0.2	0.2
phiCD505	18.8	12.7	33.8	20.5	16.0	11.0	50.7	65.1	54.4	100.0	58.3	0.9	0.4	0.6	0.6	1.0	1.4	0.7	1.3	1.3	1.3	0.2	0.2	0.2	0.2
phiMMP02	16.6	13.9	24.6	19.4	18.3	17.0	44.5	44.8	53.1	59.5	100.0	0.4	0.5	3.7	3.7	1.5	1.3	0.5	1.5	1.5	1.5	0.2	0.2	0.2	0.2
phiCD24-1	0.4	5.5	0.5	0.3	0.5	0.3	0.4	0.9	0.7	0.9	0.4	100.0	1.3	0.8	0.8	0.1	1.0	0.0	0.0	0.0	0.0	1.7	1.7	1.7	1.9
phiCD6356	0.2	0.0	0.4	1.7	0.7	0.2	0.4	0.1	0.3	0.4	0.6	1.3	100.0	0.0	0.0	4.8	5.7	3.0	3.3	2.4	2.0	3.9	4.4	4.4	4.5
phiCD211	0.2	0.1	0.2	0.6	1.3	0.2	0.6	0.2	0.2	0.2	1.4	0.3	0.0	100.0	100.0	0.2	0.1	0.3	0.5	0.5	0.5	0.3	0.3	0.3	0.2
phiCDIF1296T	0.2	0.1	0.2	0.6	1.2	0.2	0.7	0.2	0.2	0.2	1.4	0.3	0.0	99.9	100.0	0.2	0.1	0.3	0.5	0.5	0.5	0.2	0.2	0.2	0.2
phiMMP04	4.8	1.4	1.4	2.1	11.1	8.5	2.2	3.4	3.1	1.6	2.3	0.1	5.7	0.6	0.6	100.0	38.6	44.9	48.9	44.5	46.4	7.0	7.3	7.3	7.5
phiCD506	0.9	0.3	1.9	1.6	13.3	13.8	2.0	1.1	1.1	2.0	2.0	1.4	6.3	0.3	0.3	37.1	100.0	74.5	64.5	60.3	59.9	12.9	12.0	12.3	12.1
phiCD481-1	1.6	0.8	1.2	3.1	15.0	17.0	3.7	0.3	0.6	1.2	0.9	0.0	3.1	1.2	1.2	43.4	75.7	100.0	79.1	71.8	73.4	10.8	9.9	10.2	9.9
phiCDHM13	3.8	1.9	1.8	3.4	11.8	12.9	3.5	1.0	1.2	1.8	2.3	0.0	3.5	2.0	2.0	46.7	64.5	77.6	100.0	92.5	94.4	10.8	9.8	9.9	10.1
phiCDHM11	4.0	2.0	2.0	3.6	12.4	13.6	3.6	1.0	1.2	2.0	2.4	0.0	2.7	2.1	2.1	44.3	63.1	74.0	97.2	100.0	91.6	11.4	10.3	10.5	10.6
phiCDHM14	4.0	2.0	2.0	3.6	12.5	13.3	3.5	1.0	1.2	2.0	2.4	0.0	2.0	1.8	1.8	44.6	61.3	74.3	97.2	89.9	100.0	9.8	8.8	9.1	9.1
phiCD146	0.4	0.4	0.3	0.0	0.0	0.1	0.0	0.3	0.3	0.3	0.3	1.6	3.7	0.7	0.7	5.2	10.7	8.7	8.8	8.9	7.8	100.0	77.3	73.1	62.5
phiCD38_2	0.5	0.5	0.2	0.0	0.0	0.1	0.0	0.4	0.2	0.2	0.2	2.0	4.0	0.8	0.8	5.6	9.9	7.9	8.1	8.1	7.0	78.8	100.0	88.8	72.4
CDKS8 *	0.5	0.4	0.1	0.0	0.0	0.1	0.0	0.3	0.1	0.1	0.1	1.8	3.8	0.7	0.7	5.8	9.6	7.8	8.1	8.1	7.1	73.3	87.9	100.0	80.3
phiCD111	0.5	0.5	0.2	0.0	0.0	0.1	0.0	0.4	0.2	0.2	0.2	1.8	4.1	0.8	0.8	5.7	9.7	8.0	8.2	8.2	7.0	62.8	71.5	79.9	100.0

Figure 5-7. Heatplot showing the result of nucleotide comparison of *C. difficile* phage.

The similarity value were considered using Gegenees software based on pair wise translated comparison of the analysed sequences (BLASTN method, fragment size 200/100. The heat map colours reflect sequence similarity, ranging from low (green) to high (red). The new isolated phages are labelled with the star.

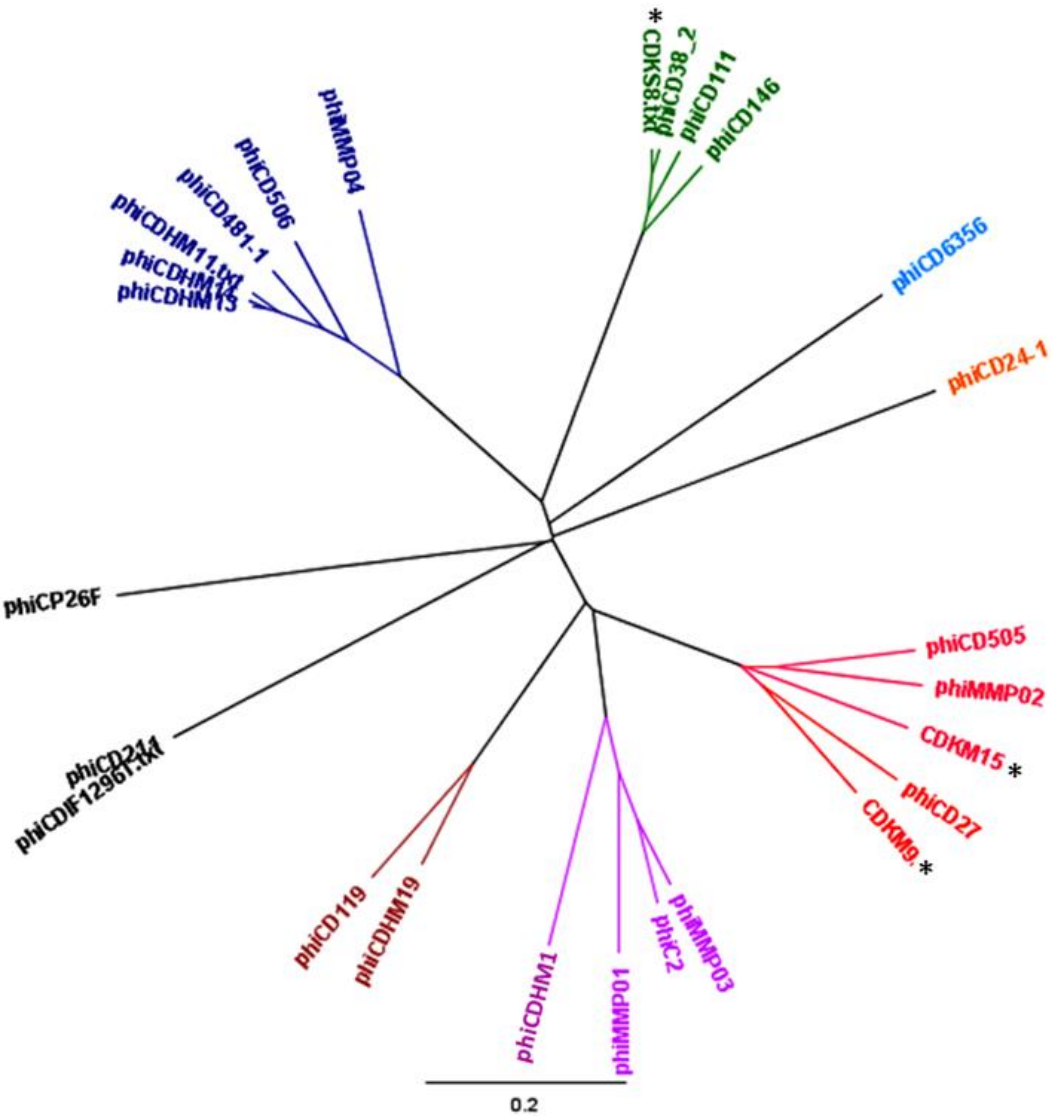


Figure 5-8. Phylogenetic tree based on whole genome comparison of *C. difficile* phages genome.

The similarity values were calculated based on translated pairwise comparison of the analysed sequences using Gegenees software. The phylogenetic tree was constructed with Splits Tree using the neighbour joining method. The analysis performed on 26 whole genomes of 19 myoviruses (CDKM9, CDKM15, CDMH1, phiC2, phiCD119, phiCD211, phiCD27, phiCD505, phiCD506, phiCDHM11, phiCDHM13, phiCDHM14, phiCDHM19, phiMMP01, phiMMP02, phiMMP03, phiMMP04, phiCD481-1 and phiCDIF1296T), 6 siphoviruses (phiCD111, phiCD146, phiCD24-1, phiCD38-2, phiCD6356) and one *C. perfringens* phage(phiCP26) as an out-group. The scale bar represents a 20% difference in average tBLASTx score. Branch colours correspond to clustering phage by amino acid sequence similarity. The new isolated phages are labelled with the star.

1.3.4 Phylogeny of the endolysin genes

To determine whether the genes responsible for host cell lysis follow the same evolutionary history of the overall phage genomes, a phylogenetic tree was generated using the neighbour-joining method at the amino acid level. The amino acid sequences of homologs for the endolysin gene in related phages were retrieved from GenBank using BLASTp and confirmed following protein domain identification using InterProScan and NCBI Conserved Domain Search databases (Appendix 5). After aligning the sequences with ClustalX2 (Larkin et al., 2007), evolution models were inferred using the Neighbor-Joining method and bootstrap of 1000 replicates, the analysis performed on MEGA 7.0 (Kumar et al., 2016), and the tree was visualized with Geneious 6.1.8 (Kearse et al., 2012).

The resulting tree reflected the taxonomic division between siphoviruses and myoviruses supported by a bootstrap value of 1000, in which 14 endolysin sequences (myoviruses) fell within clade 1 and 7 were clustered within clade 2. Of these, 6 are siphovirus and one is myovirus (Figure 5-9). In addition, there is a group of myoviruses that are clustered distantly from the myovirus clade and form clade 3, which includes the 'jumbo viruses' phiCDIF1296T and phiVD211, with phiCD119 and phiCD481-1. However, the phylogram is incongruent with the whole genome tree and subgroupings are not well reflected, as the myoviruses were sub-clustered regardless of the morphology. We inferred, then, that the evolutionary history of the endolysin genes does not reflect the phylogeny of the whole phage genomes.

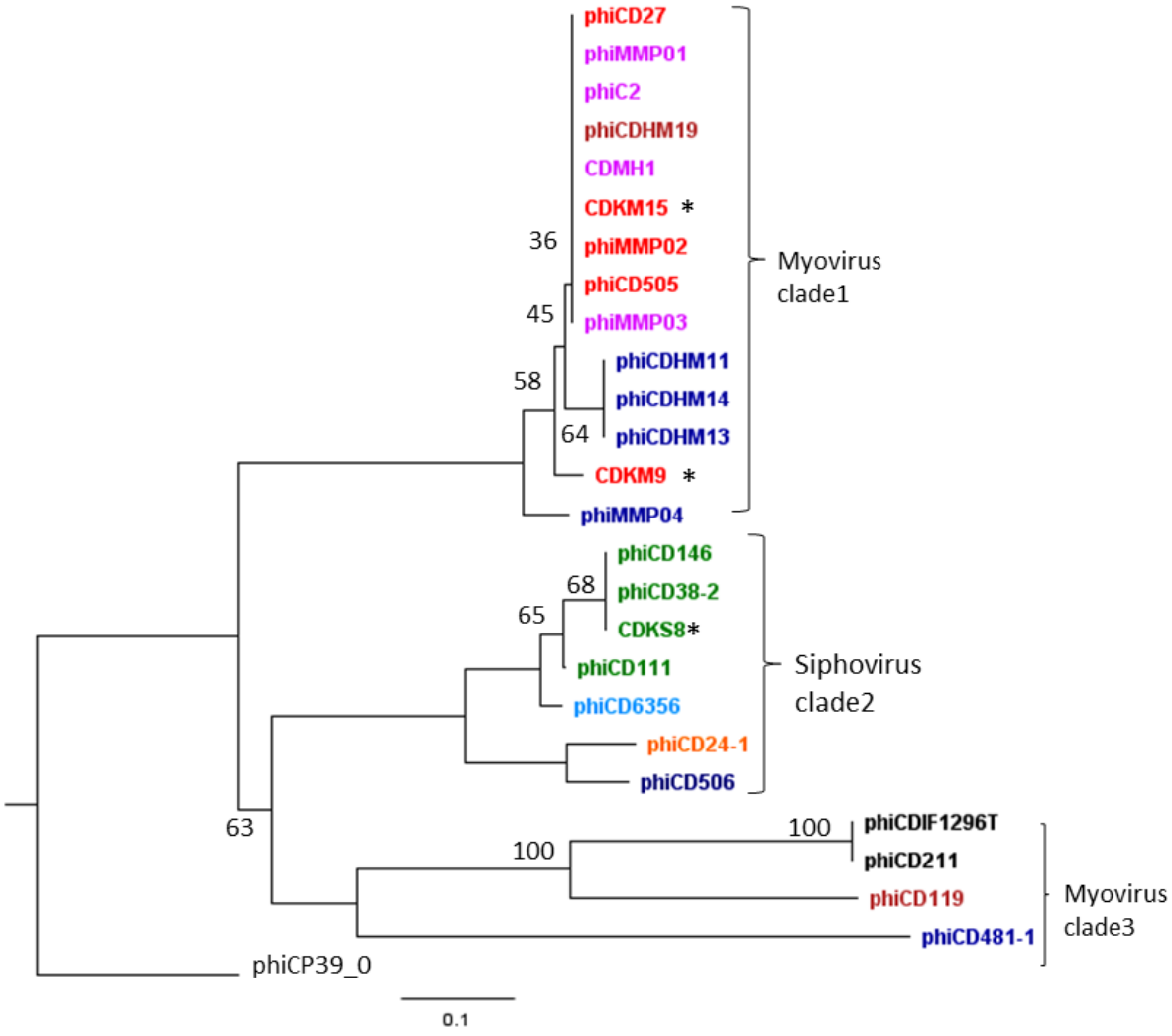


Figure 5-9. Phylogenetic analysis of the Endolysisn gene of *C. difficile* phages.

The evolutionary history was inferred using the Neighbor-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. The evolutionary distances were computed using the p-distance method based on their amino acid sequences. All positions containing gaps and missing data were eliminated. The analysis involved 26 amino acids of 19 myoviruses (CDKM9, CDKM15, CDMH1, phiC2, phiCD119, phiCD211, phiCD27, phiCD505, phiCD506, phiCDHM11, phiCDHM13, phiCDHM14, phiCDHM19, phiMMP01, phiMMP02, phiMMP03, phiMMP04, phiCD481-1 and phiCDIF1296T), 6 siphoviruses (phiCD111, phiCD146, phiCD24-1, phiCD38-2, phiCD6356) and *C. perfringens* phage (phiCP39_0) used as an out-group. Evolutionary analyses were conducted in MEGA7 (Kumar et al., 2016). The new isolated phages are labelled with the star. Colours correspond to phage clusters in Figure 5-8.

1.3.5 Phylogenetic analysis of TerL and the packaging strategy of isolated phage

As the physical terminus for the genome of the new phages CDKS8, CDKM9 and CDKM15 has not been identified, to predict the packaging strategy of analysed phages we therefore followed the method proposed by Casjens and Gilcrease (Casjens and Gilcrease, 2009). Briefly we aligned large subunit terminase genes from *C. difficile* phages to corresponding proteins from reference phages (Appendix 6) using ClustalX2. Maximum Likelihood tree was built and phylogenies were determined by bootstrap analysis of 1000 replicates in MEGA 7.0 version (Kumar et al., 2016).

The resulting tree revealed that the myoviruses CDKM9 and CDKM15 form a common clade with phiCD505, phiCD27 and phiMMP02, consistent with the whole genome analysis (Figure 5-10). However, this clade doesn't cluster with any phages with known packaging strategy. Corresponding to the whole genome similarity as well the siphovirus CDKS8 clusters with other related *C. difficile* siphoviruses in a known clade with DNA packaging strategies classified as P22 like (head full).

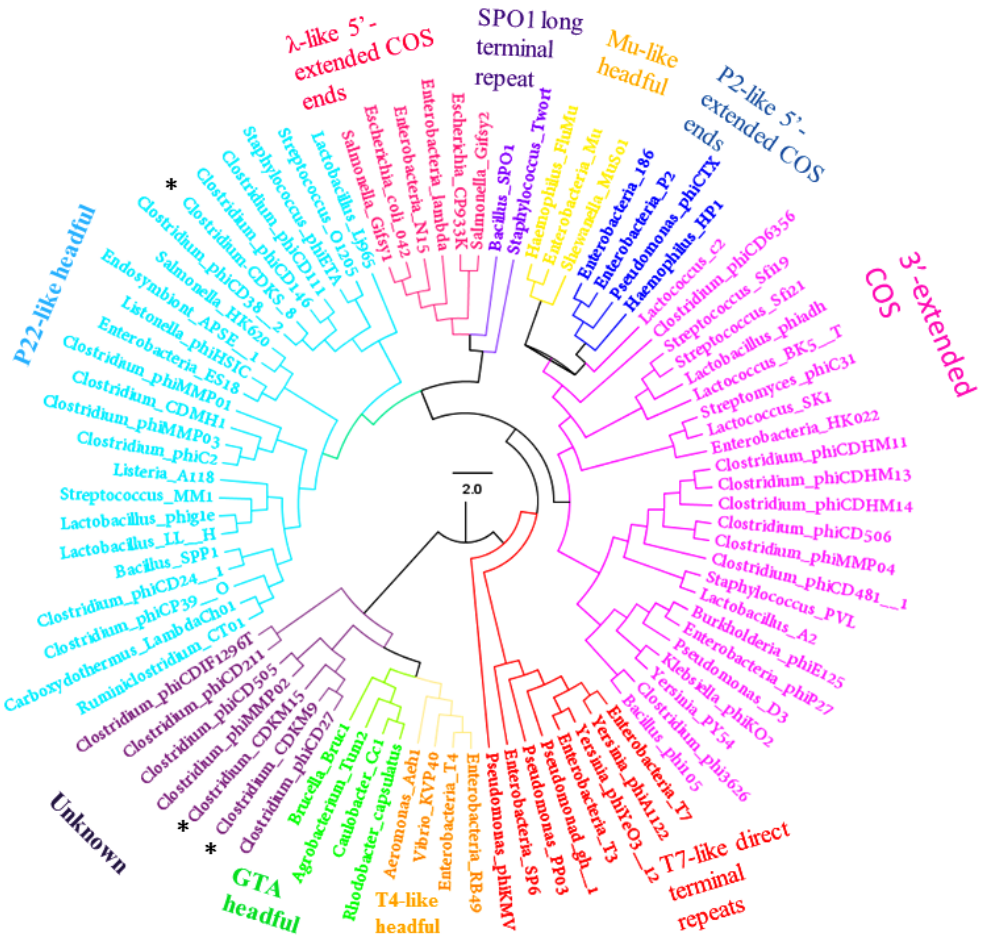


Figure 5-10. Phylogenetic analysis based on the alignment of large terminase proteins of different phages.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman Frequency model (Whelan and Goldman, 2001). The Initial tree (s) for the heuristic search was obtained by applying the BioNJ method to a matrix of pairwise distances estimated using a JTT model. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 86 amino acid sequences. The name of the phage or prophage is shown at each terminal node and packaging strategy for each group is given as well. The branches are coloured according to the DNA packaging strategy of the member phages: light pink (3′-cohesive ends), blue (P2 like 5′-cohesive ends), dark pink (λ-like 5′-extended COS ends), yellow (Mu like headful), dark purple (SPO1 long terminal repeat), light blue (P22-like headful), purple (unknown), green (GTA-headful), orange (T4-like headful) red (T7-like direct terminal repeats). All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA7 (Kumar et al., 2016), and the tree visualised with the FigTree.

1.4 Discussion

In the last decade there has been an exponential increase in sequenced *C. difficile* phage genomes that provides insight into the diversity of *C. difficile* phages isolated from sources in Europe and North America. In this study we fill a gap in the global picture of their diversity, having isolated and characterized phages from Northern Iraq. Seventeen phages were isolated and characterised; of these, three phages were selected based on their host range and particle morphologies as described in Chapter 5 to investigate the genomic diversity represented by a distant geographical region.

Complete genome sequences of the siphovirus CDKS8, and the two myoviruses CDKM9 and CDKM15 were determined in an effort to gain insight into their genetic diversity. No virulence factors or toxin genes could be promptly deduced from sequence analyses, which so far appears to be a general feature of this group of phages (Goh et al., 2007, Hargreaves et al., 2014b, Horgan et al., 2010, Mayer et al., 2008, Meessen-Pinard et al., 2012, Sekulovic et al., 2011). Based on the whole genome similarity between the new isolated phages and the known *C. difficile* phages, CDKM9 and CDKM15 are closely related to the long-tailed myoviruses phiCD27, phiCD505 and phiMMP02, and the siphovirus CDKS8 is closely related to phiCD38-2, phiCD146 and phiCD111.

The genome architecture of the *C. difficile* phages sequenced in this project all conform to a similar modular arrangement, by gene function. However, each phage has genes that are specific to their own genomes, and the distribution of these provides insight into the potential evolution during co-infection as well as across different host species. The genome feature of each phage is discussed in detail below.

5.4.4 *C. difficile* myovirus genome diversity and overall relationships

Genome sequence analysis of both CDKM9 and CDKM15 revealed that genome length and GC content of both phages are similar to that of the *C. difficile* myoviruses in the phiCD119virus species, but are at the lower range of 51-60 kb in genome size (Lavigne et al., 2009). Both myoviruses have putative genes that are involved in necessary functions, such as head packaging, morphogenesis, attachment, lysis, lysogeny control and DNA replication. In CDKM9 there is an ORF that encodes RusA ortholog which specifies that the phages ease their own recombination events and acquisition of new genetic material (Hargreaves, 2012). In addition, two copies of *repR*

encoding a predicted regulatory protein with a penicillinase repressor family protein domain (Pfam PF03965), were found in both myoviruses. Their position adjacent to integrase is significant, indicating that they likely play a role in the maintenance of lysogeny (Goh et al., 2007).

Additionally genome comparison for the two myoviruses with 22 known *C. difficile* phages revealed extensive mosaicism across the genomes (Figure 5-4 and Figure 5-5), which is highly characteristic for phage genomes. Mosaicism arises due to the continuing recombination of phage genomes and the mechanisms that drive these patterns of genetic exchange may differ in specific cases, for example illegitimate versus homologous or even site-specific recombination that facilitates these events (Casjens, 2005, Brüssow et al., 2004). The results here suggest the exchange of whole regions as well as CDS operons (e.g. those encoding the hypothetical proteins in the CDMK9 genome), in addition to conserved genes with sequence similarity across the phage genomes, e.g. the endolysin. Of note there is a region in the CDMK15 genome with no similarity to the other phage sequences. This is where the CRISPR array is located(Goujon et al., 2010).

The notable feature of *C. difficile* prophages is that several carry CRISPR (clustered regularly interspaced short palindromic repeats) arrays (Sebaihia et al., 2006, Hargreaves et al., 2014a), and CDKM15 was found to have a CRISPR array with 6 spacers. These phage carried arrays are located in the structural gene region of the genome and have nearby Bro N-terminal domain protein presumed to be involved in their processing (Hargreaves et al., 2014a) , which is consistent for the array in CDKM15. The bacterial CRISPR/Cas system acts as a bacterial "immune system" defending against invading phages and foreign DNA elements (Barrangou et al., 2007). The *C. difficile* phage carried arrays contain spacers that target other *C. difficile* phage sequences (Hargreaves et al., 2014a) and so presumably could have a role in phage defence. As this phage was isolated from geographically distant source, we wanted to determine if its spacer sequences could reveal past interactions with phages related to the published *C. difficile* phages. The array in CDKM15 had spacers that perfectly match to phiCD505, another long tailed myovirus that was induced from a dog isolate in Canada (Sekulovic et al., 2014). The protospacer in phiCD505 had a PAM motif CCT), which is also consistent with that previously identified (Hargreaves et al., 2014a), and suggests this could confer immunity to this related phage. The origin of the

CRISPR/Cas system from bacteriophages is still unknown, but could be from the host as there is abundant evidence for the continuous exchange of genetic materials between phages, bacterial genomes, and in bacteriophages, these provide in new insight into the strategy of co-evolution of phage and bacteria.

5.4.5 *C. difficile* siphovirus genome diversity and overall relationships

The result of the sequencing and annotation of CDKS8 revealed that its genome shares a high degree of similarity (94.4% to 95.2%) at the amino acid level and (87.9% to 88.8%) at the nucleotide level with phiCD38-2, despite the geographically distant location. Several highly conserved genes were noticed from the genome that share high levels of similarity to genes in other *C. difficile* siphoviruses, such as phiCD111 and phiCD146 (Sekulovic et al., 2014).

Despite the presence of genes with a high degree of similarity to other known *C. difficile* phages, CDKS8 also has unique genes. For example, ORF15 in the tail morphogenesis region that encodes bacteriophage gp15 family protein. Studies have evidenced that Gp15 signify a novel class of phage-encoded superinfection prohibiting proteins that seems have been derived from a large family of enterobacterial proteins of unidentified function (Cumby et al., 2012). So far the gp15 membrane protein has been found in the temperate phage HK97 within its tail morphogenesis region, and has been found to be absent in most closely related phages (Cumby et al., 2012, Juhala et al., 2000). They demonstrated that the HK97 prophage gives the bacterial cell immunity to infection by phages HK97 and closely related phages (Cumby et al., 2012, Sekulovic et al., 2015), and suggested that it entered the phages' genomes by horizontal transfer in a relatively recent evolutionary period (Juhala et al., 2000). The transfer of novel genetic material into the genomes of bacterial viruses (phages) has been widely documented (Hargreaves et al., 2014b). The presence of gp15 family protein in CDKS8 appears to provide an interesting example in which phages have adapted a host function and evolved it to their own benefit (Cumby et al., 2012). In addition, the presence of an atypical regulatory/lysogeny module in CDKS8 also provides an example of such an unusual genomic organization among *C. difficile* phages. However only two Siphoviridae phage genomes, phiCD6356 (Horgan et al., 2010) and phiCD38-2 (Sekulovic et al., 2011) have been sequenced so far, in addition this has recently been noticed in myovirus phiMMP02. The suggestion is that this might have occurred from a

past recombination event between a *Myoviridae* and a *Siphoviridae* phage (Meessen-Pinard et al., 2012).

5.4.6 **Taxonomy of these new phages**

A distance based phylogenetic tree was constructed based on pairwise similarity scoring between the three new phages and 22 *C. difficile* phages, which classified *C. difficile* phages into 5 taxonomic groups. CDKM9 and CDKM15 grouped into the phiCD119likevirus genus, and both clustered in the long tailed myoviruses sub-clade within this genus, and CDKS8 grouped with the siphoviruses that did not fall in either the phiCD119likevirus genus or proposed phiMMP02likevirus genus (Hargreaves and Clokie, 2015, Lavigne et al., 2009) and their sub-clades within this group of phages. The “jumbo" myoviruses (phiCD211 and phiCDIF1296T) were highly divergent and could not be classified into any of these groups as well. Accordingly, we propose three additional genera including phicd119virus, phiMMP04virus, “phicd38-2virus” containing phiCD38-2, phiCD111, CDKS8 and phiCD146, “phicd211virus’’ containing phiCD211 and phiCDIF1296T, and “phicd6356virus” including phiCD6356 and phiCD24-1. This is also supported by the protein cluster analysis of these phages. In the future it may be suitable to further divide the latter into separate genera depending on the discovery of further phages as they are distantly clustered.

5.4.7 **Phylogenetic analysis of these new phages**

The phylogenetic analysis of one of the most conserved genes, the endolysin gene revealed the taxonomic division of the siphoviruses and myoviruses. The result based on the comparison of the neighbour joining tree from the endolysin sequence analyses and the whole genome phylogeny, suggests that it may have been transferred horizontally between phages. This lateral transfer is exemplified by the endolsyin gene of the myovirus phiCD506 as this falls within the clade of siphovirus endolysin genes (Sekulovic et al., 2014). The possibility of its horizontal exchange is supported by studies which have proposed the evolution of modular endolysins through the transfer of phage and bacterial genes as single modules (Oliveira et al., 2013). In addition there is a small clade (clade3) that contains 4 endolysin genes of myoviruses (phiCDIF1296T and phiCD211, with phiCD119 and phiCD481-1) that are distantly located from the other 14 myoviruses. This is consistent with the result of protein analysis as the endolysin sequences in myoviruses were present in two conserved clusters (Goujon et

al., 2010). These were cluster no. 1, which includes 14 myoviruses and cluster no.296 which include phiCDIF1296T and phiCD211. However, phiCD119 and phiCD481-1 do not share any of these clusters, and this might possibly be due to the absence of the endolysin protein, only the putative N-acetylmuramoyl-l-alanine amidase domains are in the genome (Govind et al., 2006). It is of note, the endolysin genes are of biomedical interest because their specificity in action offers an alternative method to combat *C. difficile* without the destruction of effective GI tract microbiota (Mayer et al., 2008)

Whereas the endolyisn tree did not follow the same evolutionary history of the overall phage genomes, the phylogenetic tree based on *TerL* gene of the phages was consistent with whole genome. The generated tree showed the CDKS8 phage clustered with phicd38-2, phiCD211, phiCD146 and phages that their DNA packaging strategies classified as P22 like head full, as long phiCD 38-2 contains a pac site, we assume the same termini of the DNA molecule ends for CDKS8. However, the case was different with phages CDKM9 and CDKM15, as they clustered with phages with unclassified packaging mechanism. A second method to predict the DNA packaging mechanism is to identify the termini of the DNA molecule ends, as the cos and pac strategies result in different sequences (Casjens and Gilcrease, 2009), but in the case of the two phages no apparent termini could be located which is consistent with a previous investigation to identify (cos sites) of the related phage phiCD27 (Mayer et al., 2008). The DNA packaging mechanism used by this group of related phages therefore remains unknown. One of the considerations regarding the use of phages for therapeutic purposes is their capacity to facilitate horizontal gene transfer (Chen et al., 2015). The potential for phage mediated transduction of genetic material may be connected to the mechanism of DNA packaging into the virion (Rao and Feiss, 2015). Headful packaging phages (*pac* phages) provide a way in which generalised transduction may occur, but phages with cohesive end (*cos* phages) require sequence recognition in the packaging process. In *C. difficile* phages, phiCD38-2 has been found to contain a *pac* site, whereas phiCD6356 has a *cos* site identified (Horgan et al., 2010, Sekulovic et al., 2011). However, attempts failed to identify a cohesive ends (*cos* sites) for both phiCD119 and phiCD27 (Govind et al., 2006, Mayer et al., 2008).

To conclude, we isolated phages from a distinctly different geographical locations yet three phages were found closely resemble other *C. difficile* phages based on genome

analyses. Genome Comparisons revealed dissimilarity in discreet gene modules as well as the presence of a CRISPR array in the genome of CDKM15, and its array spacers are identical to one of the sequenced phages from the North America. This has implications for our understanding of the global evolution of this pathogen in these reservoirs and across the globe.

Finally, based on the results of whole genome phylogenic analysis and specific gene phylogenetic analysis, *C. difficile* phages can be classified into 5 taxonomic groups. This observation expands the current taxonomy, and is consistent with earlier analysis on fewer phage genomes (Hargreaves and Clokie, 2015). With this in mind, three additional genera can be proposed including phicd119virus, phiMMP04virus, "phicd38-2virus" containing phiCD38-2, CDKS8, phiCD111 and phiCD146, "phicd211virus'' containing ϕCD211 and ϕCDIF1296T, and "phicd6356virus" including phiCD6356 and phiCD24-1.

Chapter 6 Comparison of the growth physiology, toxin and spore production between environmental and clinical strains of *C. difficile*

6.1 Introduction

C. difficile causes epidemics and outbreaks in hospitals worldwide (Muto et al., 2005, Pepin et al., 2004). The disease known as *C. difficile* infection, or CDI can be caused by multiple strains of hundreds of types of *C. difficile*, and epidemiological studies have shown that strain, or ribotype prevalence varies both spatially and temporally (Freeman et al., 2010, Honda and Dubberke, 2014). The bacterium has also been found in environments outside clinical settings (Hargreaves et al., 2013, Zhou et al., 2014, Zidaric et al., 2010). Diverse strains of *C. difficile* have been found in the natural environment; and often multiple strain types have been shown to be present in close proximity to each other (Hargreaves et al., 2013, Zhou et al., 2014, Zidaric et al., 2010). The prevalence of the bacterium in food and meat products have been documented globally, signifying the potential for foodborne transmission of this pathogen to humans through intake of contaminated products (Gould and Limbago, 2010). This raises questions of where new strains come from, how they have evolved, and whether the natural environment functions as a source where new *C. difficile* strains can evolve, prior to the transmission to other hosts (Hargreaves et al., 2015).

In general, the existence of the bacterium and its persistence in the natural environment are frequently affected by environmental factors and bacteria can adapt their physiology (including their growth rate) according to their environment (Parkinson, 1993). In the estuarine environment, enteric bacteria are subject to several environmental stresses such as variations in salinity, oxygen and temperature (Bordalo et al., 2002, Burkhardt Iii et al., 2000).

Estuaries are dynamic and fertile areas at the transition between river (freshwater) and marine (saline, tidal) environments, and could represent a diverse and prolific reservoir of human pathogens. These pathogens are frequently associated with sediments and typically arise from contamination of fresh and marine waters with human sewage (Malham et al., 2014). A previous study that examined *C. difficile* in sites throughout Langstone and Chichester Harbour estuarine system revealed a diverse range of *C. difficile* strains including ribotypes related with both environmental and

clinical sources. This presented an opportunity for greater comprehension of the potential transmission mechanisms, persistence and ecology of *C. difficile* in Hampshire estuarine system (Hargreaves et al., 2013). The authors suggested that at least some of the *C. difficile* found in this system may have originated from sewage or agricultural runoff and that the bacterium remains viable within estuarine sediment Further research reported the isolation of diverse *C. difficile* strains that belong to different ribotypes, in addition to novel strains, from the marine water column (al Saif and Brazier, 1996, Hargreaves, 2012, Pasquale et al., 2011, Zidaric et al., 2010).

There are over 430 officially recognised *C. difficile* ribotypes (Patel, 2013). Specific ribotypes of *C. difficile* have been associated with strain phenotypes, such as increased disease severity linked with ribotype 027 due to more sporulation observed *in vitro* (McDonald et al., 2005, Warny et al., 2005). This has led to ribotype 027 frequently being referred to, as a 'hypervirulent variant' on the basis of studies which have identified two deletions found in the toxin regulator genes (Hubert et al., 2007, Morgan et al., 2008, Stabler et al., 2006). These deletions cause elevated toxin A and B production of approximately 16 and 23 times, respectively, in comparison with toxinotype 0 (Warny et al., 2005).

In *C. difficile* the major virulence factors are toxin A and toxin B, as described in section (2.4.10), and are responsible for the pathogenicity of *C. difficile*. The toxigenic strains of *C. difficile* express the toxins during the late exponential to early stationary phases of growth (Hundsberger et al., 1997). Researchers have observed that differences between the strains of different ribotypes in terms of their growth, toxicity and spore production (Vohra and Poxton, 2011). Additionally, the bacterium has other putative virulence factors, these include sporulation factors that are responsible for the ability to survive in the environment and transmit disease (Vedantam et al., 2012). *C. difficile* spores germinate in the colon to create the vegetative cells that initiate CDI. During CDI, *C. difficile* induces a sporulation pathway generating greater numbers of spores that are important for the prolonged existence of *C. difficile* in patients and for the horizontal transmission between hospitalised patients (Paredes-Sabja et al., 2014). Further, studies have confirmed a variation in sporulation rate between different strains, including strains of the same ribotype (Burns et al., 2010, Burns et al., 2011).

An additional feature that influences virulence is the flagellum, which has been demonstrated by studies (Rosey et al., 1996, Postnova et al., 1996) to be essential for motility, and thus for colonisation, in several enteric pathogens. Therefore, since *C. difficile* inhabits an environmental niche comparable to that of these guts, the flagella of *C. difficile* could have important role in colonisation and consequently pathogenicity (Stevenson et al., 2015). Recent research has suggested that *C. difficile* flagella influence virulence more directly than simply permitting force-driven motility towards gut nutrients and the flagella gene region is in fact involved in the modulation of toxin expression (Aubry et al., 2012, Baban et al., 2013, Stevenson et al., 2015).

Resistance to antibiotics is a phenotypic trait of *C. difficile* strains, including those associated with CDI, as CDI is linked to a certain range of antibiotic usage, such as clindamycin and ciprofloxacin (Gerding, 2004, Loo et al., 2005). Currently, vancomycin and metronidazole are consistently the standard treatment for CDI (Norén, 2010), although infection relapse has been stated for both of these drugs (Mizusawa et al., 2015, Perelle et al., 1997, Teasley et al., 1983). Minimum inhibitory concentrations (MICs) of a range of antibiotics had been ascertained and ranged from sensitive to fully resistant to clindamycin and ciprofloxacin among environmental strains of *C. difficile isolated* from an estuarine system. One isolate was observed with MIC suggestive to be of moderate susceptibility to vancomycin (Hargreaves et al., 2013).

Regardless of drug resistance, environmental strains of *C. difficile* isolated from the estuarine system were found to be also a genetically highly diverse and dynamic population and some of the ribotypes observed within these sediments have previously been associated with both clinical and non-clinical environments (Hargreaves et al., 2013). Ribotypes that cross over between clinical, animal and environmental strains of *C. difficile* have been documented elsewhere (J. S. Brazier 2001; Janezic et al. 2012). Information on the biology of *C. difficile* environmental strains and their adaptation for survival in environmental conditions has not examined before. Accordingly, a comparative study was conducted for a set of 3 isogenic pairs of genetically related strains of *C. difficile* isolated from marine environment and hospitals, to determine at least by common usage they are inherently different.

6.2 Aims and objective of the study

1- Compare growth curves of *C. difficile* strains isolated from hospitals and the environment alongside a range of physical factors, to see if the strains from the estuaries are better able to grow in 'environmental conditions' than strains isolated from a clinical setting.

2- Compare the pathogenic potential of the strains by:
   a- Measuring total toxin A & B production using the combined ELISA kit.
   b- Determine heat resistant spore production.
3- Compare the phenotype traits of environmental and clinical strains in terms of :
   a- Motility, using a motility agar stab.
   b- Antibiotic susceptibility, using Etest® strips.

6.3 Results

A collection of 12 *C. difficile* strains used in this study were previously isolated from either hospitals, or from and the natural environment (Table 6-1), and the main purpose of this work was to determine if strain origin had an impact on their physiology. For each ribotype there is at least one representative strain that associates with clinical and non-clinical sources, which allowed comparison between the strains within each ribotype. Growth for the representative strain of each ribotype from the environment and the hospital under a range of environmental factors was determined and compared with the simultaneous toxin A and toxin B production, total heat resistant spore as well as motility and antibiotic sensitivity.

Table 6-1. Summary of the environmental clinical strains and the source isolated from.

Ribotype	Isolate	Source
RT027	AIU	Toxin positive faecal sample
RT027	K15	Sediment
RT078	AML	Toxin positive faecal sample
RT078	ALL	Toxin positive faecal sample
RT078	ASS	Toxin positive faecal sample
RT078	H5c	Sediment
RT078	H17	Sediment
RT078	J	Sediment
RT010	BQR	Toxin positive faecal sample
RT010	BQT	Toxin positive faecal sample
RT010	K16	Sediment
RT010	H3	Sediment

6.3.1 Growth comparison at optimal conditions (37°C, pH 7 and salinity 0.5% w/v)

Growth comparisons of a representative strain from the clinical (ALL, AIU and BQT) and environmental sets (H17, K15 and H3) under optimal laboratory conditions was performed (Table 6-1). The strains had similar growth patterns over a 24 h period at optimal temperatures, pH and salinity. The lag phase lasted for about 2 h in all strains, then the strain stayed in the log phase for 6 h, followed by the stationary phase (Figure 6-1, a,b,c). There were no significant differences between (AIU and K15, ALL and H17, and BQT and H3) when compared at each individual time point ($p > 0.05$), however significant difference was noticed between the clinical strains of the three ribotypes and their environmental counterparts over a 24 h period ($p < 0.05$). The

growth curves for the strains in this condition were used as standard curves to compare with the growth curves observed from other ranges of the environmental factors described below.

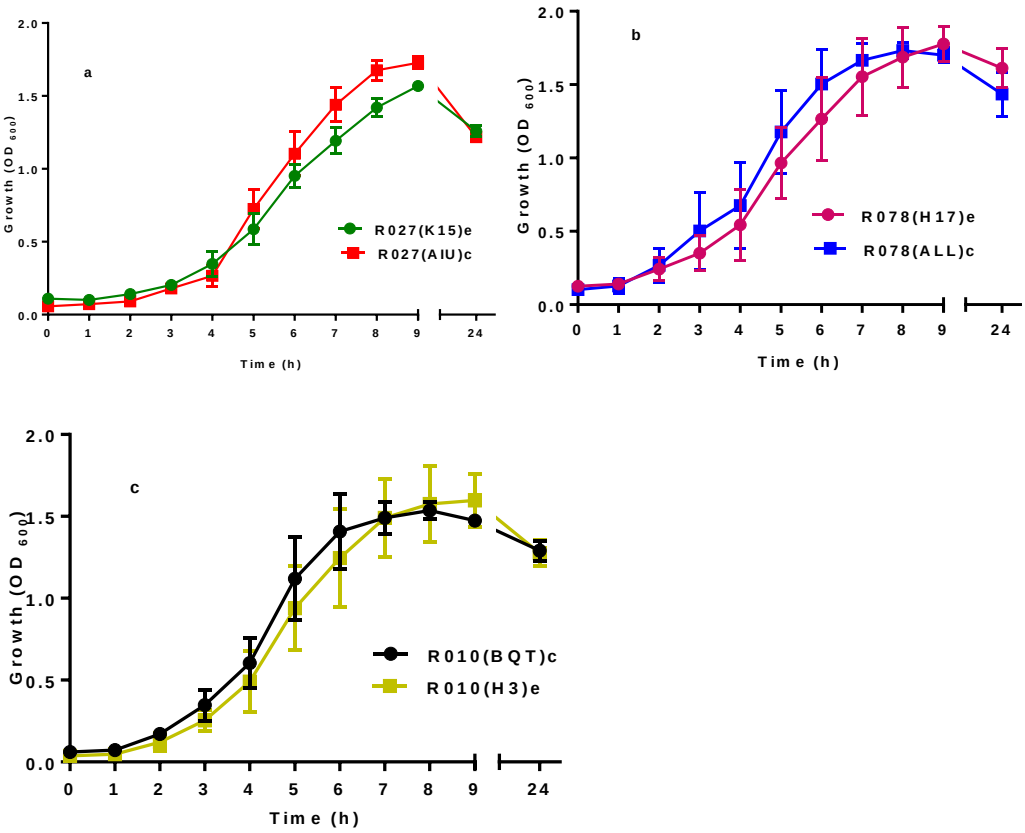


Figure 6-1. Growth curves comparison of environment and clinical strains of *C. difficile* over 24 h at 37°C.

(a) Growth curve comparisons of *C. difficile* strains of R027 at 37°C. (b) Growth curve comparisons of *C. difficile* strains of R078 at 37°C. (c) Growth curve comparisons of *C. difficile* strains of R010 at 37°C. The growth of the strains over 24 h was analysed using a Two - way ANOVA (Analysis of Variance) in GraphPad Prism 6 (Appendix 14). Values represent mean cell density and standard errors of three experiments conducted in three different occasions.

6.3.2 **Growth comparison at different temperatures**

The ability of the strains to grow under estuarine temperature (10°C) was examined. Prior to inoculating the pre-reduced BHI broth with 0.5ml of inoculum, the medium was calibrated to the desired temperatures the previous day. Non-inoculated BHI broth was

prepared and served as blank control. Neither the environmental nor the clinical strains grew at this temperature (as observed for 72 h [data not shown]). The culture did remain viable as was indicated by their growth when returned to 37°C. Consequently, the growth of the strains was examined at 23°C. The results showed that the environmental and clinical strains had different patterns of growth in 23°C from their growth curves in 37°C, all the strains had longer lag phase lasted for about 6 h (Figure 6-1,b). A significant difference were observed between the growth of clinical strains over the environmental strains over a 24 h period of growth ($p < 0.05$), this increase was observed at 9 h in case of the clinical strains of R078 and at 24 h period for the clinical strains of R027 and R010, when the clinical strains were compared to their isogenic pair from the environment at each individual time point.

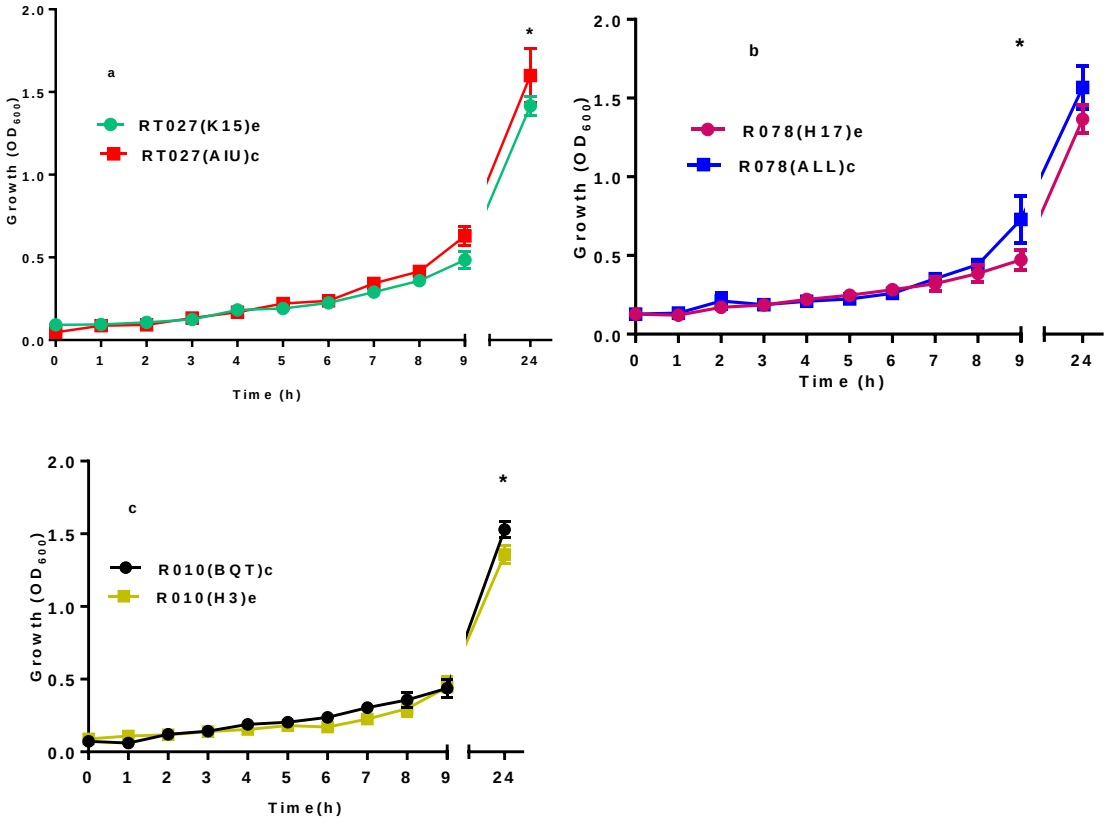


Figure 6-2. Effect of temperature on environment and clinical strains of *C. difficile* over 24 h.

(a) Growth curve of *C. difficile* strains of R027 at 23°C. (b) Growth curve *C. difficile* strains of R078 at 23°C. (c) Growth curve of *C. difficile* strains of R010 at 23°C. The growth of the strains over 24 h was analysed using a Two - way ANOVA (Analysis of Variance) in GraphPad Prism 6 and Appendix 15). Values represent mean cell density and standard errors of three experiments conducted in three different occasions.

6.3.3 **Comparison of growth curves of isolates at pH 8**

To determine the strains growth under estuarine pH, the pH of BHI broth was adjusted to pH 8. The growth curves obtained for all strains at pH 8 were almost comparable to their growth patterns in pH 7 (Figure 6-1), strains started log phase after 2h of lag phase. However, when the growth curve comparisons were performed for the strains of R072, R078 and R010, again statistically significant differences observed over a 24 h period between all clinical and environmental strains ($p < 0.05$) (Figure 6-3, a,b,c and Appendix 16), Although there were no differences between the strains of the three ribotypes at individuals time points ($p > 0.05$) beside reduced growth rate of R027 strains.

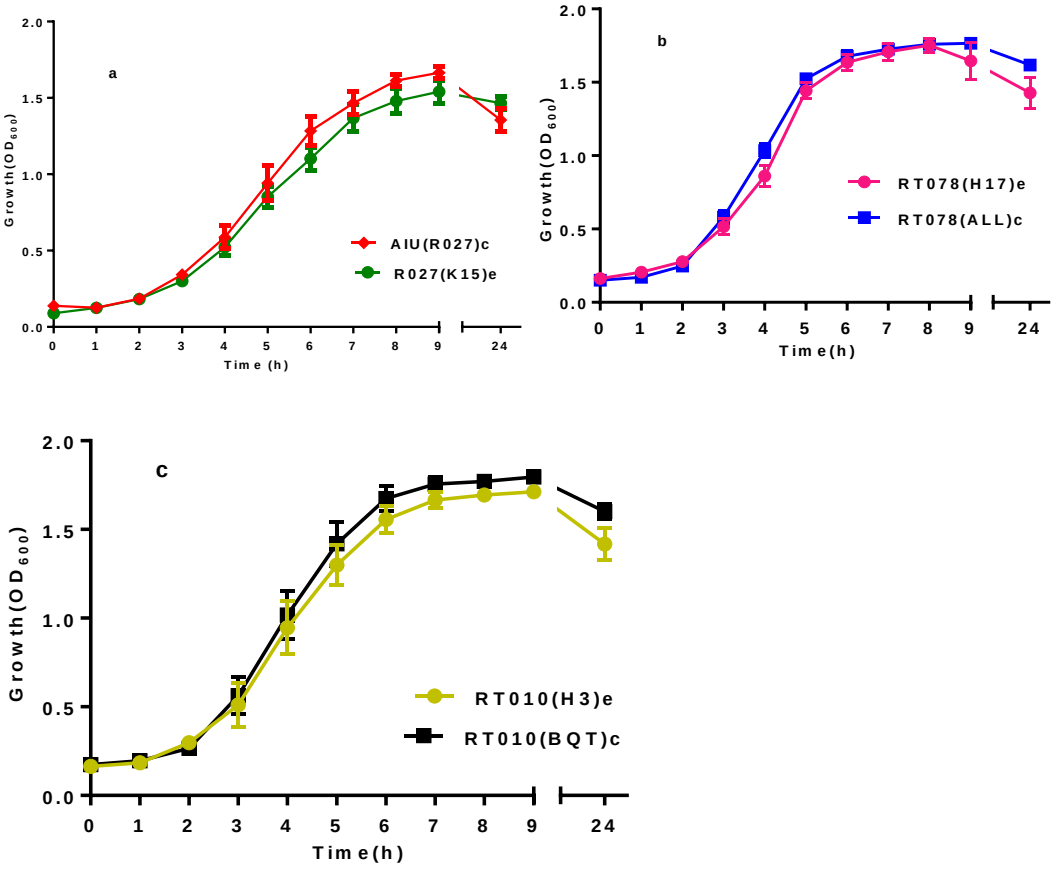


Figure 6-3.Effect of pH on environment and clinical strains of *C. difficile* over 24 h.

(a) Growth curve of *C. difficile* strains of R027 at pH 8. (b) Growth curve *C. difficile* strains of R078 at pH 8. (c) Growth curve of *C. difficile* strains of R010 at pH 8. The growth of the strains over 24 h was analysed using a Two - way ANOVA (Analysis of Variance) in GraphPad Prism 6 and Appendix 16). Values represent mean cell density and standard errors of three experiments conducted in three different occasions.

6.3.4 **Comparison of growth curves of isolates at salinity 2% (w/v).**

The two groups of the isolates were grown in BHI broth supplemented with salt concentrations 2% (w/v) that is comparable to the salinity in an estuarine environment. The strains grew in BHI broth supplemented with 2% (w/v) of sodium chloride (Figure 6-2). Growth comparison over a 24 h period was carried out between environmental and clinical strains. Substantially increased cell densities were noticed between the clinical strain of the three ribotypes and their isogenic pair from the environment ($p < 0.01$), however no differences in growth rate of clinical strains and environmental counterpart were observed between the strains of all ribotypes at individual time pointe ($p > 0.05$).

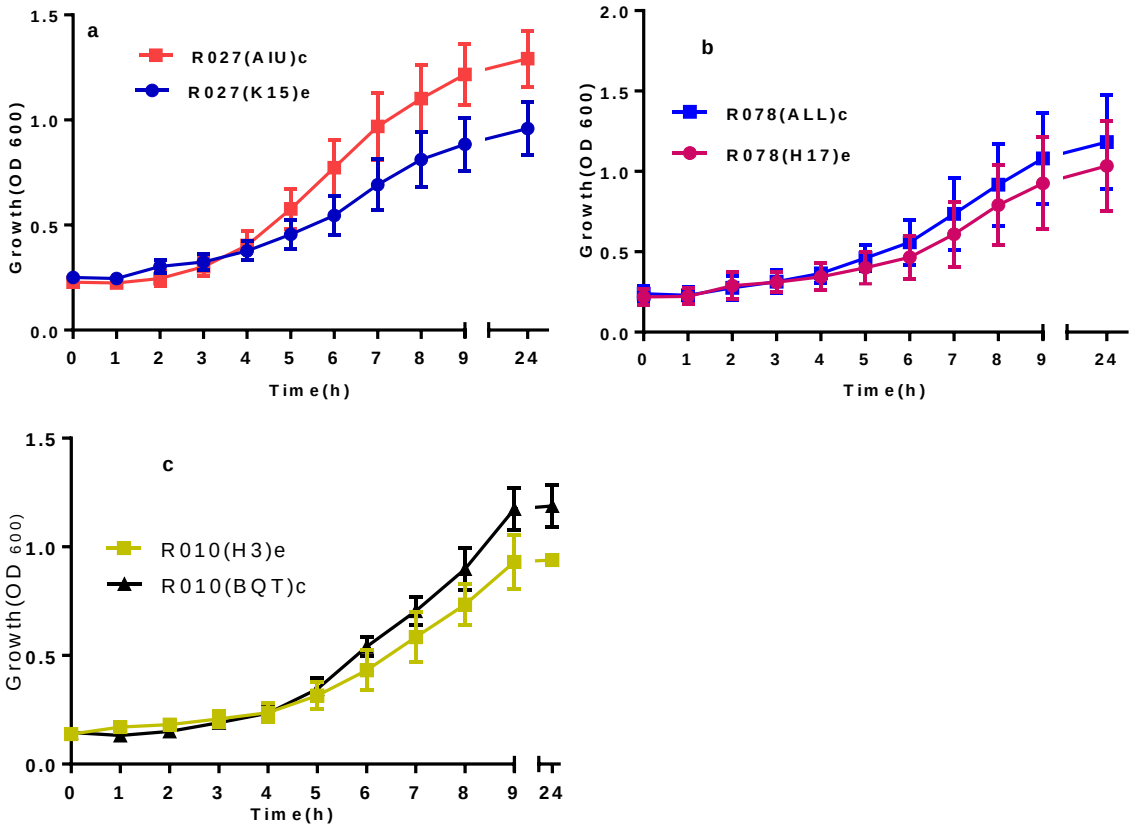


Figure 6-4. Effect of salinity on growth of environmental and clinical strains of *C. difficile* over period 24 h.

Growth of *C. difficile* strains at pH 8 (a); the patterns of growth were similar in all strains. (b) Growth of 6 the strains of *C. difficile* in BHI supplemented with 2% (w/v) of sodium chloride. Two-way ANOVA GraphPad Prism 6 was used to analysis the growth of the strains over 24 h in (Appendix 17). Values represent mean cell density and standard errors of three experiments conducted in three different occasions.

6.3.5 **Comparison of total toxin A & B production in *C. difficile* isolates**

Total toxin (A+B) production was measured for the strains from hospitals and the strains from the environment of strains using *C. difficile* TOX A/B II™ kit (TechLab, Inc). As previous studies have shown that strains of ribotype 010 are non-toxigenic, strains of R027 and R078 were examined for toxin production (Hargreaves et al., 2013, Otieno, 2013) .

All tested isolates displayed toxins at the stationary phase (24 h), with a significant amount of toxin produced by the clinical strain of ribotype 027 ($p < 0.0001$) compared to its counterpart (Figure 6-5**)**. The clinical strains of ribotype 078 (ALL and ASS) displayed a significant amount of toxin at 28 h ($p < 0.05$) and 32 h ($p < 0.01$), respectively when compared to their environmental strains (H5c & H17). At 32 h both clinical strains of ribotype 078 produced a higher amount of toxins ($p < 0.0001$) in comparison to their environmental strains, and these differences continued to 48 h when compared to strain H17 only. While at this time point the environmental strain of ribotype 078 (H5c) produced the same amount of toxin as the clinical isolate (ASS), less than ALL but not significant, and more significant than their environmental peer H17 ($p < 0.05$) .

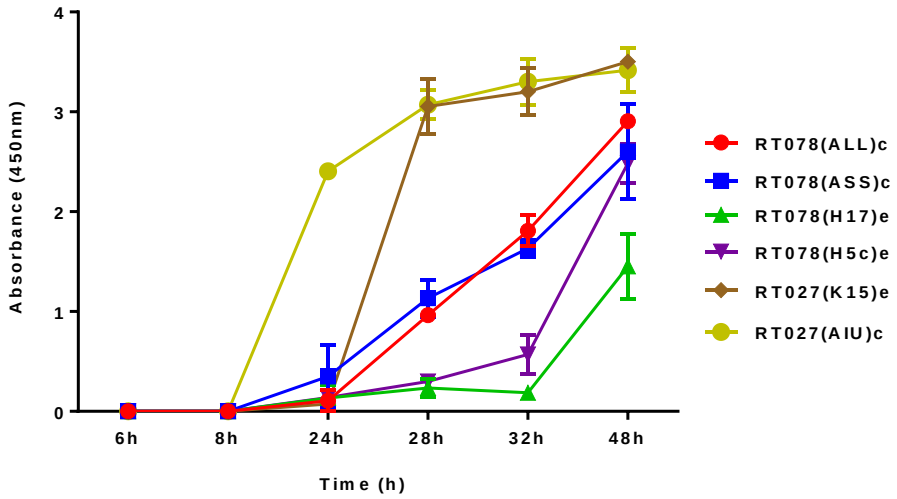


Figure 6-5. Total TcdA and TcdB production by *C. difficile* isolates.

Total toxin production of environmental and clinical strains of ribotype 078 and 027. The amount of produced toxins by the isolates was measured over 48 h and analysed using two - way ANOVA (Analysis of Variance) in GraphPad Prism 6 (Appendix 18). Bars represent mean and standard deviation of three experiments conducted in three different occasions.

6.3.6 **Comparative heat resistance spore counts of clinical and environmental *C. difficile* strains**

The sporulation rate of environmental and clinical strains were determined and compared. To initiate sporulation from the isolates, an equal amount of (40 µl) of 10^5/ml titre of spore sample was used for each isolate. This is to ensure that the same amount of starting spore material was used to initiate the sporulation assay for the isolates and to allow equitable comparisons of the total spore counts at the conclusion of the experiment. Prior to the spore initiation experiment, the spore samples were allowed to germinate in BHI broth supplemented with L-cysteine and sodium taurocholate for 18 h. This germinating starter was used to inoculate the sporulation medium culture followed by incubation for seven days. A 200 µl aliquot of the sporulation medium was taken for each strain at 0 h and heated for 3h at 65°C then plated on BHIS (supplemented with L-cysteine and sodium taurocholate) followed by incubation for 24 h. This is crucial for sporulation studies to ensure that the number of spores in the culture at the stage of initiation is minimal, in order not to misinterpret the consequent spore counts (Burns et al., 2011). At the end of 168 hours of incubation, the total spores produced by strains of ribotype were measured after heat treatment at 65°C for 3 h to kill vegetative cells but not spores. The spore suspension for all isolates was serially diluted, and plated onto BHIS agar to induce germination and enhance the recovery of *C. difficile* spores. The plates were incubated for 24 h then the CFU recovered from heat resistant spores were counted by determining their average viable counts from three replicate plate counts.

The accumulation of spores over a total of 168 h for the clinical and their counterpart from the environment were enumerated compared within each ribotype. The enumerated spores were statistically analysed for strains of R078 and R010 using Ordinary one-way ANOVA, and Paired t test was used to analysis total spore count produced by R027 strains. The analysis showed that the clinical isolates produced more spores than the environmental strains in all three ribotypes ($p<0.05$) Figure 6-6 (Appendix 19).

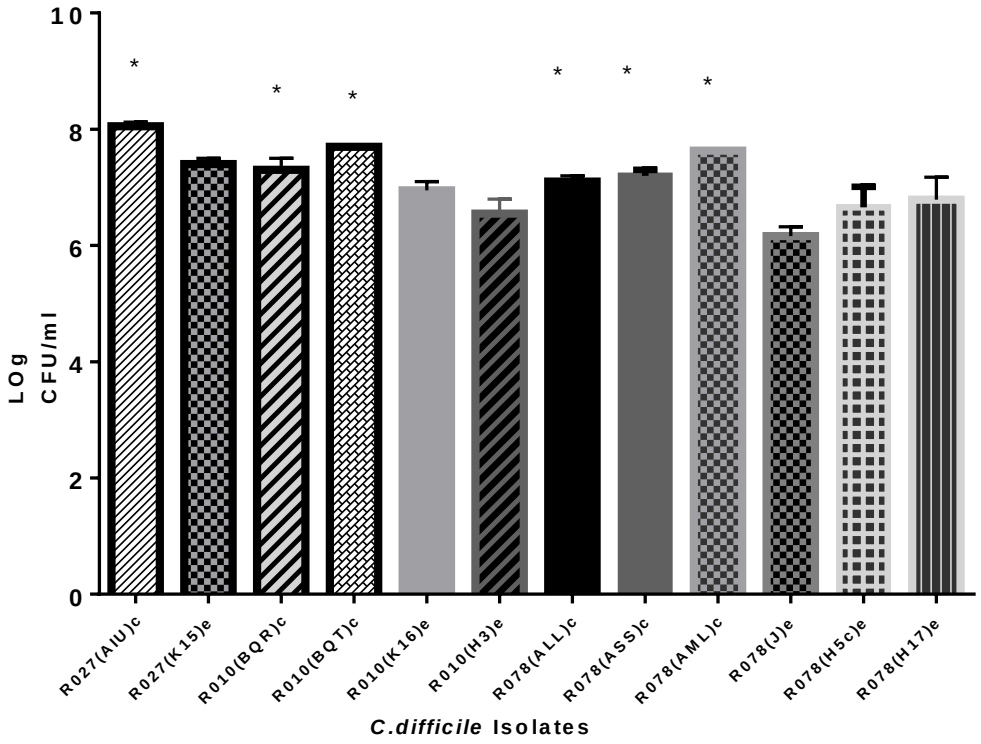


Figure 6-6. Total spores produced by environmental and clinical isolates of *C. difficile* after 168 h incubation.

The total spores produced by the 12 *C. difficile* isolates, at the end of 168 h incubation and heat treatment at 65°C for 3 h. The heat resistant spores were enumerated by determining their viable counts from triplicate plate counts on BHI supplemented agar. The experiment was carried out twice. The total spore counts were analysed and compared using one-way ANOVA and Paired t test GraphPad Prism (Appendix 19)

6.3.7 Motility assay comparative of *C. difficile* strain

Flagella enhance bacterial adaptation to their biological niches, and are encoded by flagella genes. Flagella were found to play a role in the pathogenesis of *C. difficile* (Ramos et al., 2004). A comparative assay was performed to characterise the motile growth of environmental and clinical strains. The environmental strains exhibited diverse motility, from no motility, low-motility, or high-motility growth, meanwhile the clinical strains showed no or low- motility (Figure 6-7). It was found that 5/6 of the environmental strains were motile, with one highly-motile, and 4/6 of the clinical strains had low-motility growth. The assay was carried with duplicate technical repeats and with triplicate biological repeats. There were too few isolates available within groups for statistical analysis.

Ribotype	Isolates	Source	Motility
078	ASS	clinical	-
	AML	clinical	-
	ALL	clinical	+
	H17	environmental	++
	H5c	environmental	-
	J	environmental	+
027	AIU	clinical	+
	K15	environmental	+
010	BQT	clinical	+
	BQR	clinical	+
	K16	environmental	+
	H3	environmental	+

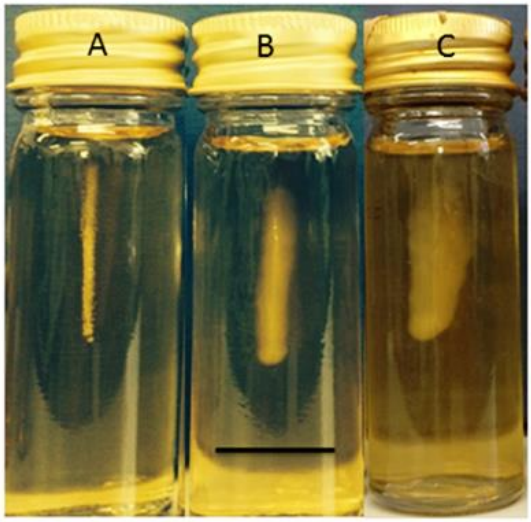


Figure 6-7. Comparative motility assays for *C. difficile* strains.

Left: ribotype and types of isolates with the result of motility assays. A negative symbol (−) indicates no motility or low motility, a positive (+) indicates motility observed and double symbol (++) indicates high levels of motility. Right: Photograph shows examples of the variable degree of motility observed, A non-motile growth, B motile growth and C highly motile growth. Hazy growth indicates *C. difficile* motile dispersion through BHI 0.175 % (w/v) agar media. Bar represents ~2 cm.

6.3.8 MIC determination of selected antibiotics for *C. difficile* clinical and environmental isolates

The environmental microbial community are generally assumed to serve as a resistome, which consists of all antibiotic resistance genes that are transferable to pathogenic bacteria (Davies, 1994, D'Costa et al., 2006, Zidaric et al., 2010). Using 12 isolates the MIC of four antibiotics was determined and compared, including those linked to CDI: vancomycin (VA), metronidazole (MZ), clindamycin (CM) and ciprofloxacin (Cl) (Alonso et al., 2001, Gerding, 2004, Marra and Ng, 2015). All isolates were sensitive to metronidazole (MIC < 8µg ml^{-1}) and vancomycin (MIC < 4µg ml^{-1}) (Figure 6-6 and Table 6-2). One isolate from the environment (J) was resistant to clindamycin, while one clinical strain (BQR) was borderline resistant to this antibiotic. Two isolates from the environment (H5c and J) were found to be resistant to ciprofloxacin (328µg ml^{-1}), and another two (H17 and K16) were borderline resistant (6 µg ml^{-1}). Of the 6 clinical

isolates from hospitals, AML showed resistance and BQR was borderline resistant to this antibiotic.

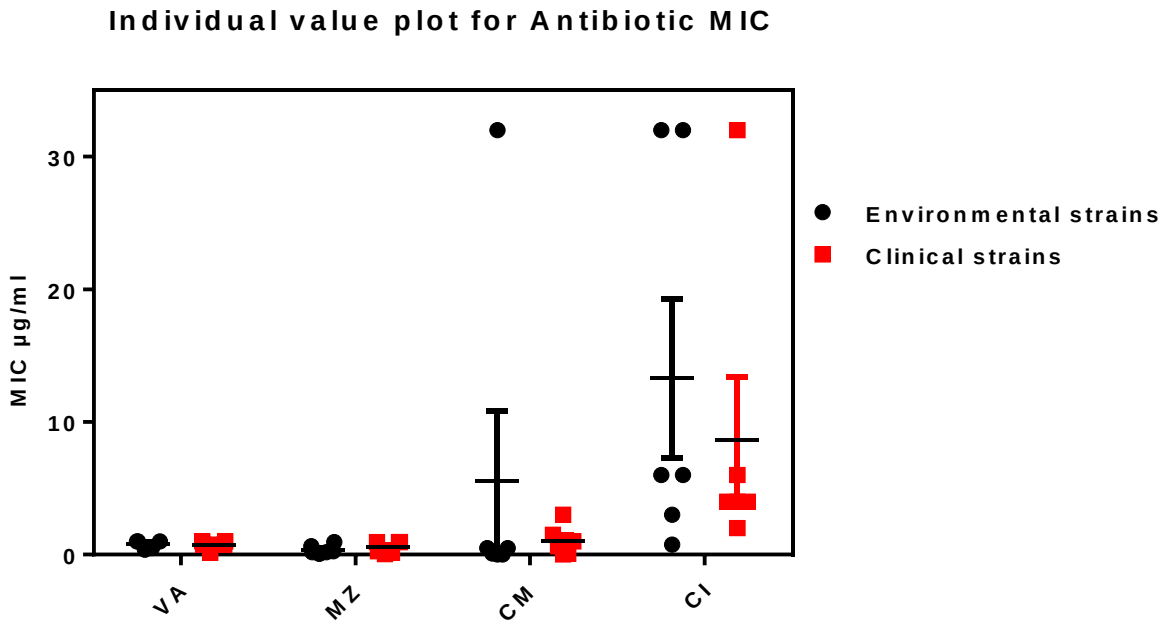


Figure 6-8. Individual value plot of MICs.

MICs (on the y axis) for each antibiotic (on x axis) of environmental and clinical isolates were plotted in an individual value plot. Lines and bars show mean and standard error of mean. Abbreviations: VA, vancomycin; MZ, metronidazole; CM, clindamycin; CI, ciprofloxacin. MICs for isolates resistant to clindamycin and ciprofloxacin are 32 µg ml^{-1}; in addition, there is greater variation in the MICs of these two antibiotics than in those of vancomycin and metronidazole.

Table 6-2. E-Test results for environmental and clinical strains of *C.difficile.* Abbreviation e: environmental strains and c: clinical strains.

		MIC in µg ml−1			
Ribotypes Isolates		Vancomycin	Metronidazole	Clindamycin	Ciprofloxacin
	(H17)e	1	0.94	0.5	6
	(H5c)e	0.5	0.64	0.5	32
RT078	(J)e	1	0.19	32	32
	(ALL)c	0.75	0.25	1	4
	(ASS)c	0.19	0.047	0.064	2
	(AML)c	0.75	0.94	0.64	32
RT010	(K16)e	1	0.19	0.016	6
	(H3)e	1	0.064	0.016	3
	(BQT)c	0.75	0.125	1.5	4
	(BQR)c	1	0.94	3	6
RT027	(K15)e	0.38	0.25	0.094	0.75
	(AIU)c	0.16	0.94	0.016	4

6.4 Discussion

Isolating *C. difficile* from the natural environment led to the suggestion a new reservoir had been formed for this bacterium and that was involved in the bacterium's abundance in the environment, origin, transmission and persistence. It is not known if their presence in the environment is due to contamination from infected hosts, sewage effluent or if they are active in these populations (Hargreaves et al., 2013, Hargreaves et al., 2015). Thus, examining physiological properties of strains from marine estuaries and hospitals, and determining their pathogenic potential is of clear interest in order to understand the relative threat which the spread of the pathogen and evolution of new strains may constitute.

In this study, it has been shown that *C. difficile* strains isolated from different sources differ in their pathogenic potential and some phenotypical aspects. The growth curve comparisons showed growth advantage of all clinical strains of over the environmental strains under optimal conditions, and this growth advantage was observed under all the environmental factors such as low temperature (23°C), alkalinity (pH 8), and salinity (2% w/v). The reasons behind this variation are not clear but could be due to the fact that there is strain-to-strain variation in growth rate, sporulation, toxin production in *C. difficile* (Siani, 2014). However, previous studies showed lack of variation between strains growth, although theses strains were all form clinical sources (Vohra and Poxton, 2011, Merrigan et al., 2010). No published work has previously examined the growth of strains originating for the environment. Genome examination of these strains highlighted genomic variation between strains isolated from the environment and hospital (Hargreaves et al., 2015).

The virulence of a pathogen is reliant on a distinct set of genetic factors and their well-regulated expression (Chakraborty et al., 2000). In *C. difficile tcdA* and *tcdB* genes are the major virulence factors. The result of total toxin A and B production assay revealed that both groups of the *C. difficile* strains produce the toxins, with an increased amount of toxins produced by the clinical strains at specific time. Clinical strains of ribotype 027 produced significant amount of toxins at 24 h ($p<0.0001$) compared to the environmental counterpart of the same ribotype (K15). Although previous genome examination indicated that both strains have all of the 5 *Paloc* genes (Otieno, 2013, Hargreaves et al., 2015), this discrepancy could be due to the variation in the levels of transcription of the *PaLoc* genes (Vohra and Poxton, 2011). The scenario was same

with both clinical strains of ribotype 078 (ALL & ASS), as both produced significant amount of toxin compared to their environmental strains. This matched the *Paloc* region analysis of the strains, which have been shown to have partial deletion of both *tcdA* and *tcdB* genes in both environmental strains (H5c & H17) (Hargreaves et al., 2015, Otieno, 2013). Studies demonstrated that the presence and expression of critical virulence genes in environmental strains of other pathogens such as *V. cholera*, may be critical and appear to constitute an environmental reservoir of virulence (Chakraborty et al., 2000).

Associations between toxin and spore production have previously been reported (Kamiya et al., 1992), and from the data presented here, clinical stains of the same ribotype (R027) displayed a significant number of heat resistant spores ($p<0.05$) compared to the isogenic-pair from the environment, while clinical strains of R078 and R010 displayed a higher amount but not significant different compare to their environmental isogenic-pairs. This is comparable to an earlier study which showed a significant variation in the amount of the spore production by genetically different strains, and also highlighted that hypervirulent strains produce more toxin than non-hypervirulent strains, both producing more spores than environmental strains (Wilcox and Fawley, 2000). While other researchers have stated that the sporulation in *C. difficile* is variable between strains rather than type associated (Burns et al., 2011, Burns et al., 2010). One of the key limitations of this study was the fact that the sample size within each ribotype was small, especially within R027 as a single isolate was used from each source. Others proposed more than seven *C. difficile* strains in one group, for the sake of determining variation in sporulation accurately between different groups (Burns and Minton, 2011). It has also been shown that the high toxin-producing strains sporulated more than others (Vohra and Poxton, 2011). Conversely, previous studies have observed an inverse correlation between toxin levels and spore counts (Åkerlund et al., 2006, Merrigan et al., 2010) suggesting that sporulation early in the stationary phase results in fewer toxins.

Toxin transcription in *C. difficile* is modulated by the flagella regulon, which provides proposals for involving regulatory networks for virulence genes (Aubry et al., 2012). The result of the motility assay demonstrated that environmental strains of a single ribotype have variable degrees of motility compared to clinical strains. This might be result of horizontal gene transfer between these strains in the natural environment that is mediated by mobile genetic elements (Hargreaves et al., 2015). The

four flagella genes (FliC, FliN, FlgM, FLiW) were entirely absent in the non-motile strains (ASS, ALL and H5c) (Otieno, 2013). Additionally, Hargreaves et al. (2013) showed variation in motility within the environmental strains, with no clear link detected between ribotype and motility. Both antibiotic resistant and sensitive strains are present among the strain from both sources, with slightly more resistant among environmental strains. This higher rate of antibiotic resistance in the environmental strains, could suggest the possibility of antimicrobial genes being transferred via mobile elements among this microbial community. A greater sample size could better define the variation between isolates.

6.5 Conclusion

Clostridium difficile is fascinating pathogen. Different strains with different properties exist and the reason behind its evolution and where they originate from is still not clear. Studying the properties of a large number of strains from both sources might reveal where the variations lie and that this is not simply strain to strain variation. In this study we have shown that clinical strains grew better than the environmental strains, even under a range of environmental stresses. Clinical strains produced more toxins and spores than the environmental strains. We also showed that the strains are different in terms of motility, with environmental strains being more motile. More antibiotic resistant strains are present within the environmental strains. *C. difficile* strains in the natural environment are genetically diverse.

Chapter 7 Summary

Clostridium difficile is a common hospital-acquired pathogen that causes infectious diahorrea. Despite its pathogenic capacity, it can be carried asymptomatically and, notably, this has an effect on disease transmission. The natural environment is a reservoir for genetically diverse *C. difficile* strains and these have been isolated from various environmental sources, including soil, river water, sea water, and estuarine sediments.

However, knowledge of the prevalence and diversity of *C.difficile* strains in certain regions is still limited. It is important to note that a relatively small body of work has been devoted to the investigation of *C. difficile* strains originating in the Middle East and, furthermore, nothing is known about the strains from this area that are found in environmental settings. To address this, soil and sediment samples were collected throughout Kurdistan in northern Iraq and examined for the presence of *C. difficile* and its bacteriophages. From the samples, a diverse strain set of 61 strains *C. difficile* were isolated and shown to have several novel features (details are in Chapter 3). Additionally, three further isolates were identified as novel species of *Clostridium* most closely related to *C. saccrolyticum* (Rashid et al., 2016).

Of the *C. difficile* strains, 7% were identified as a novel ribotype R691 (given this assignation during this study by the *Clostridium difficile* ribotyping network (CDRN) network, Leeds). Whole genome analysis revealed that the two strains examined from the most dominant ribotype were highly diverse compared to most known strains, and they can be classified into the newly described clade C-I which previously contained only non-toxigenic strains (Dingle et al., 2014). Interestingly, strains with different toxin profiles were identified that belong to R011. This is unusual as typically toxin profiles are consistent within a ribotype (Dr Catherine Eckert pers. comm.).

In terms of the physiological characteristics, variation in motility was observed within strains that were attributed to one ribotype (Chapter 3). Similarly in R091 the strains had different antibiotic sensitivity profiles with respect to clindamycin, where strains were resistant, sensitive, or borderline resistant. This can possibly be attributed to the horizontal gene transfer that is facilitated by mobile genetic elements and prophages. The presence of multiple prophage carriage within these strains suggests the important role of phages in genetic exchange within this reservoir. Typically *C. difficile*

strains encode one or two prophage (Hargreaves and Clokie, 2014), but remarkably the genome examination of one of these strains (CD105KSO7) identified 6 prophages. In addition to complete phages, further phage tail-like particles were detected in all the genomes and these may have an impact on the physiology of the host and the evolution within the microbial community. Consistent with this key role for phages, abundant and variable CRISPR arrays were found within strain genomes, with variation in the number of arrays between strains of single ribotypes which is suggestive of recent activity. The phage sensitivity analysis for these strains showed that phages (both from this study and from previous work in our laboratory) have limited access to strains with matching CRISPR arrays. Thus the CRISPR system does appear to be important for phage defence in this species.

Chapter 4 describes the isolation of novel phages. One of the motivating factors for this study was to determine if *C. difficile* phages isolated from a novel geographical area could infect key ribotypes including 078, for which few phages exist. A set of 17 new *C. difficile* phages that infect the bacterium were isolated following enrichment without the addition of strains. The characterisation of these phages revealed that all belong to order *Caudovirales* and, furthermore, all were either of the siphoviruses or myoviruses. Host-range analysis revealed that the phages are promising candidates for therapeutic purposes as they can infect a different spectrum of hosts than was observed with existing phages in our laboratory, and that reported from other groups. For example 14 new siphoviruses were found that target R078. In addition to the isolation of the first long-tailed myovirus that can target epidemic strains of R001 and hypervirulent strains of R027.

Chapter 5 describes whole genome sequencing of three phage genomes which demonstrated that whilst similar to known phages, they have a number of novel features. For example CDKS8, a siphovirus encodes a GP15 family protein which is a membrane protein with super infection exclusion activity; this has been shown to be present in the bacteriophage HK97 but not in *C. difficile* phages (Cumby et al., 2012). The myovirus CDKM15 is novel because it is the first phage know to encode an active CRISPR array; it has 6 spacers that target other *C. difficile* phage protospacers (Sekulovic et al., 2014). Although a spacer has been found before in a *C. difficile* phage (Goh et al., 2007), that phage is not thought to be active. These observations provide new insights into the co-evolution of phages and the bacteria.

The phylogenetic analysis based on the endolysin gene tree reflected the taxonomic position as siphoviruses or myoviruses. In terms of phage packaging the sequence analysis of the TerL packaging ATPases of these phages suggested that CDKS8 may package their genomes into capsids using either a P22-like head full mechanism whereas the mechanism for CDKM9 and CKM15 remain unknown. The comparative genomic analysis of CDKM9 and CDKM15 with other *C. difficile* phages showed a high degree of mosaicism across the lengths of their genomes, and this can potentially be attributed to the homologous recombination that occurred throughout co-infection. It should be noted that this is natural in the context of the environmental microbial community.

Chapter 5 describes whole genome comparisons of these three new phages and 22 published *C. difficile* phage genomes. The analysis showed that *C. difficile* phages can be classified into 5 taxonomic groups. CDKM9 and CDKM15 are grouped into the phiCD119likevirus genus, and both were clustered in the long tailed myovirus sub-clade within this genus. Furthermore, although as expected CDKS8 grouped with the siphoviruses, it is thus far unique and didn't fall into either the phiCD119likevirus genus or the proposed phiMMP02likevirus genus (Hargreaves and Clokie, 2015). Whole genome analysis and protein cluster analysis and specific gene phylogenetic analysis, all suggest that *C. difficile* phages should be classified into 5 taxonomic groups. This observation is consistent with earlier analysis but expands the current accepted taxonomic framework. I thus intend to propose 3 additional genera to the ICTV so the taxonomy would be: (1) the phicd119virus, (2) phiMMP04virus, and the "phicd38-2virus" containing phiCD38-2, phiCD111, the phiCD146; "phicd211virus'' containing phiCD211 and phiCDIF1296T; and "phicd6356virus".

The final part of this study aimed to determine if *C. difficile* strains from the environment had inherently different properties from those isolated from a clinical environment by comparing their physiology of *C. difficile* strains from the environment little is known about the strains outside clinical sets. The researchers found that the physiology of strains from the estuarine system was different from the clinical strains. One implication of this is that they might produce lower levels of toxins and spores.

In conclusion, by studying *C. difficile* strains in novel areas, new and important facts about the biology of *C. difficile* have been revealed. In this way, the study has

contributed significantly to the extant literature relating to the understanding of the global diversity of *C. difficile* strains and its bacteriophages.

7.1 Future work

Future work will aim to clarify the impact of multiple prophage carriage on the physiology of the hosts, and on how they relate to each other. In addition the therapeutic potential of these phages will be explored. Because of time limitation it was not possible to show if the isolated siphoviruses (CDKS6, CDKS7 and CDKS8) are independent phages and if it is possible to isolate identical phages in nature. This could be done by resequencing the genome of the phages from the original samples and confirming the SNPs from the genomes and to determine differences in the physiology properties of the phages such as host range and kinetics of phage adsorption.

One of the key limitations of this study was the fact that the sample size within each ribotype that was used to determine heat resistant spore production by environmental and clinical strains was small. To determine variation in sporulation accurately between different groups, more than seven *C. difficile* strains are needed in one group. If this experiment is repeated in the future, a larger sample size should be considered within each ribotype.

Chapter 8 Appendices

8.1 Appendix 1.

Table 8-1. A list of all media used for growth *C. difficile* throughout this project.
All media were made up to 500 ml using distilled water and autoclaved to 121°C for 15 minutes.

Media	**Composition of media**
Brain Heart Infusion (BHI) broth	18.5 g BHI broth
BHIS(Brain heart Infusion supplemented) broth 2% (w/v) Sodium chloride	18.5 g of BHI broth 10 g of NaCl
BHI 1% (w/v) agar plates	18.5 g BHI broth 5 g Bacteriological agar
BHI 1% (w/v) agar 7% (v/v) defibrinated horse blood (DHB) plates (blood agar plates)	18.5 g BHI broth 5 g Bacteriological agar 35 ml DHB (added after media has cooled)
BHI 0.4% (w/v) agar (semi-solid medium)	37 g BHI 4 g Bacteriological agar
BHI 0.175% (w/v) agar (motility soft agar)	18.5 g BHI 0.88 g Bacteriological agar
BHIS (brain heart infusion) broth supplemented with 0.1 % (w/v) Sodium taroculate L-cystine	18.5 g BHI 0.1 % Sodium taroculate (0.5g) 0.1% L-cystine (0.5g)
Cycloserine-cefoxitin egg yolk (CCEY) agar plate	28.25 g CCEY supplemented with after has cooled: 25 ml Egg yolk emulsion 5 ml cycloserine (250 mg/l) and cefoxitin (8 mg/l)
Fastidious Anaerobe (FA) Broth	14.85 g of FA
Supplemented FA broth	14.85 g FA 0.1% Sodium Taurocholate (0.5) One vial of 5 ml cycloserine (250 mg/l) and cefoxitin (8 mg/l) added after media has cooled.

8.2 Appendix 2

Table 8-2. List of buffers and solutions that were used throughout this study.
Autoclave was carried out to 121°C for 15 minutes.

Reagent	**Materials**	**purpose**
5 % Chelex 100	0.25 g Chelex 5 ml Ultra-pure (UP) water	For Bacterial DNA extraction
Double salt solution (0.4 M $MgCl_2$, 0.1 M $CaCl_2$)	81.2 g $MgCl_2$ 1.5 g $CaCl_2$ Made up to 100 ml with distilled H_2O. Autoclaved	For phage plaque assay and spot test
SM buffer (phage buffer)	0.1 M NaCl 8 mM $MgSO_4.7H_2O$ 50 ml of 1 M Tris–HCl Made up to 1 L with distilled H_2O. Autoclaved	For phage suspension
20x SSC buffer in 100ml	17.5g Sodium Chloride 8.82 g Sodium Citrate 2 H_2O. Adjust to pH 7 Autoclaved	For Bacterial DNA extraction
0.1x SSC buffer in 10 ml	500 µl 20x SSC 99.95 ultra-pure H_2O	For Bacterial DNA extraction
Bacterial lysis buffer containing 10 mM Tris-Cl (pH 8.0), 1 mM EDTA and 1 % w/v SDS in 100 ml	1 ml 1 M Tris-Cl 20 µl 0.5 M EDTA 10 ml 10 % SDS (w/v) 100 ml made up with ultra-pure H_2O	For Bacterial DNA extraction
Bacterial lysis buffer with 0.2 mg/ml proteinase K in 1 ml	1 ml Bacterial lysis buffer 4 µl 50 mg/ml proteinase K stock solution Proteinase K was added just prior to use.	For Bacterial DNA extraction
1 M Tris-Cl in 1 L	121.1 g Tris base Made up to 1 L with distilled H_2O Adjust to pH 7. Autoclaved	Used for DNA extraction
10mM Tris-Cl	10 ml 1M Tris.Cl 990 ml distilled H_2O	Used for DNA extraction
10 mM Tris-Cl (2.5 mg/ml lysozyme)	100 µl lysozyme (25 mg/ml w/v in UP water 900 µl 10 mM Tris.Cl Lysozyme added just prior to use.	Bacterial DNA extraction
Phage Lysis buffer (50 mM EDTA, 50 mM Tris. Cl (pH 9.0, 1 % SDS (w/v)) in 100 ml	5 ml 1 M Tris. Cl (pH 8) 0.5 M EDTA (pH 8) 10 ml of 10 % SDS (w/v) 10 ml (or 1 g SDS powder) made up to 100 ml with UP H_2O	Phage DNA extraction
Phage lysis buffer with 0.5 mg of proteinase K	9.9 ml Phage lysis buffer 100 µl (50 mg/ml) Proteinase K solution Proteinase K added just prior to use.	Phage DNA extraction

Table 8.2 continued.

0.5 M Ethylenediaminetetra acetic acid EDTA in 500 ml	93.05 g EDTA 500 ml made up with distilled H_2O (PH 8)	For making other solutions
10 % (w/v) SDS (Sodium dodecyl sulfate) in 10 ml	1 g SDS 10 ml distilled ultra-pure H_2O	Phage DNA extraction
10 mg/ml Proteinase K	10 mg Proteinase K powder 1 ml UP H_2O	For DNA extraction
1x TAE buffer	40mM Tris. HCl 1.14 mM Acetate 1 M EDTA	For gel electrophoresis
5x TBE buffer	54 g Tris Base 27.5 g Boric acid 20 ml 0.5 M EDTA pH 8 ultra-pure to 1 L with H_2O	For PFGE
10x TE (100 mM Tris. Cl, 10 mM EDTA) (pH 8) in 100 ml 1M Tris.Cl	10 ml of 1M Tris.Cl 2 ml of 0.5 M EDTA Made up to 100 ml with distilled H_2O. Autoclaved	For PFGE
1 x TE in 100 ml	10 ml 10x TE Made up to 100 ml with UP H_2O	For PFGE
20. 2 % w/v Plug agarose	2 mg Seeplaque® CTG agarose 1 ml 0.5x TBE . Keep warm at 55°C.	For PFGE
1 % w/v PFGE agarose (Pulsed-field certified megabase agarose) in 0.5 x TBE in 200 ml	2 g Pulsed-field certified megabase agarose 200 ml 0.5 x TBE buffer.	For PFGE
22. 1 % w/v RESponse Regular PCR Agarose gel in 1 x TAE in 100 ml	1 g RESponse Regular PCR Agrose 100 ml 1x TAE	For ribotyping
300 µg/ml w/v Mitomycin C	2 mg of Mitomycin C 48 mg of NaCl 6.67 ml of ultra-pure water.	For prophage induction,
300 µg/ml w/v Norfloxacin	10 mg Norfloxacin powder 33 ml Absolute alcohol	For prophage induction,
1 M $MgCl_2$ in 100 ml	2.03 g $MgCl_2$ Made up to 100 ml with ultra-pure H_2O	Phage DNA extraction
3 M Sodium acetate in 10 ml	2.46 g Sodium acetate Made up to 10 ml with ultra-pure H_2O	Phage DNA extraction

8.3 Appendix3

Table 8-3. List of all *C. difficile* strains were used in this study.

Keys: Green cells = environmental strains, blue cells = clinical strains, * = strains used for phage screening, Ref = reference strains, Red = MLVA typed strains.

Strains	Source	Ribotype	Area	Strains	Source	Ribotype	Area
H8	Sediment	001	UK	AML	Human	078	UK
H18	Sediment	001	UK	AMT	Human	078	UK
H4	Sediment	001	UK	Ref/CD630	Human	012	Switzerland
NCTC 11204	Human	001	Ref/UK	K10	Sediment	012	UK
R4001	Human	001	UK	CD10	Human	012	UK
F1	Sediment	001	Kurdistan	AKP	Human	078	UK
F2	Sediment	001	Kurdistan	AKM	Human	078	UK
F5	Sediment	001	Kurdistan	ANJ	Human	078	UK
F6	Sediment	001	Kurdistan	ANL	Human	078	UK
F7	Sediment	001	Kurdistan	AIA	Human	078	UK
F8	Sediment	001	Kurdistan	AHS	Human	078	UK
M7	Sediment	010	Kurdistan	AMM	Human	078	UK
M4	Sediment	010	Kurdistan	R40V0106	Human	106	UK
H15	Sediment	002	UK	H12	Sediment	220	UK
AQV	Human	002	UK	K12	Sediment	220	UK
P	Sediment	002	UK	K	Sediment	220	UK
O	Sediment	002	UK	Q	Sediment	015	UK
S	Sediment	002	UK	Ref/M68	Human	017	Ireland
H5b	Sediment	005	UK	Ref/CF5	Human	017	Belgium
K18	Sediment	005	UK	M	Sediment	031	UK
H5	Sediment	005	UK	H11	Sediment	031	UK
H1b	Sediment	005	UK	R23	Sediment	023	UK
H3*	Sediment	010	UK	AQA	Human	023	UK
K16*	Sediment	010	UK	K9	Sediment	021	UK
I	Sediment	010	UK	Ref\CD106	Human	106	UK
R10*	Infant	010	UK	R6106	Human	106	UK
R	Sediment	010	UK	R39V106	Human	106	UK
H1b	Sediment	010	UK	J6	Sediment	035	Kurdistan
BQT*	Human	010	UK	J10	Sediment	035	Kurdistan
BQR*	Human	010	UK	T6*	Equine	076	UK
K6	Sediment	014	UK	K2	Soil	604	Kurdistan
V	Sediment	014	UK	K5	Soil	604	Kurdistan
ATJ	Human	014/20	UK	K6	Soil	604	Kurdistan
ATT	Human	014/20	UK	J5	Sediment	011	Kurdistan
AIU*	Human	027	UK	F9	Sediment	011	Kurdistan
K15*	Sediment	027	UK	D1	Sediment	091	Kurdistan
HB*	Infant	027	UK	D6	Sediment	091	Kurdistan
Ref/BI-9	Human	027	USA	D9	Sediment	091	Kurdistan
CD0027	Human	027	UK	S2	Soil	091	Kurdistan
Ref/CD196	Human	027	France	D8	Sediment	091	Kurdistan
Ref/R20291	Human	027	UK	D7	Sediment	091	Kurdistan
7L	Human	027	UK	C	Soil	691	Kurdistan
9L	Human	027	UK	B1	Sediment	691	Kurdistan
15L	Human	027	UK	ALE	Human	078	UK
14L	Human	027	UK	R8	Sediment	085	UK
16L	Human	027	UK	Y	Sediment	220	UK
17L	Human	027	UK	K3	Sediment	220	UK
22L	Human	027	UK	K10	Sediment	106	UK
28L	Human	027	UK	ASS*	Human	078	UK
46L	Human	027	UK	ALL*	Human	078	UK
90L	Human	027	UK	AKJ	Human	078	UK
AJV	Human	027	UK	J*	Sediment	078	UK
H5c*	Sediment	078	UK	H17*	Sediment	078	UK

8.4 Appendix 4

Table 8-4. List of *C. difficile* phages used in this study with the accession numbers.
Key: NA = not applicable

C.difficile phages	Accession	*C.difficile* phages	Accession	*C.difficile* phages	Accession
phiC2	NC_009231.1	phiCTP1	NC_014457.1	PhiCP26F	NC_019496
phiCD119	NC_007917.1	phi24R	NC_019523.1	PhiCT9441A	KM983329
phiCDHM1	HG531805	phiCT453A	KM983327.1	PhiCD24-1	LN681534.1
phiCDHM19	LK985322	phiCT19406B	KM983331.1	PhiCDHM3	LN680004
phiCD27	NC_011398.1	phiCT19406C	KM983332.1	PhiCDHM4	LN680005
phiMMP02	NC_019421.1	phiCTC2A	KM983333.1	PhiCDHM5	LN680006
phiCD481-1	LN681538.1	phiCDHM14	LK985321	PhiCDHM6	LN680007
phiCDHM11	HG798901	phiCPV4	NC_018083.1	PhiCDHS1	LN680008
phiCDHM13	HG796225	phiCP9O	JF767210.1	CDKM15	KX228399
phiMMP04	NC_019422.1	phi3626	NC_003524.1	CDKS8	NA
phiCD38-2	NC_015568.1	PhiS63	NC_017978.1	CDKS7	NA
phiCD6356	NC_015262.1	phiCP7R	NC_017980.1	CDKS6	NA
phiCD505	LN681539.1	phiSM101	NC_008265.1		
phiCD111	LN681535.1	phiCTC2B	KM983334.1		
phiMMP03	LN681542.1	phiCT19406A	KM983330.1		
phiCD146	LN681536.1	phiCT453B	KM983328.1		
phiCD211	LN681537.1	phiCDHM2	LN680003		
phiCP34O	NC_019508.1	phiCD506	LN681540.1		
phiMMP01	LN681541.1	phiCP13O	PRJNA181210		
phiCP39_O	NC_011318.1	phi8074-B1	NC_019924		
phiZP2	NC_018084.1	CDKM9	KX228400		

8.5 Appendix 5

Table 8-5. Analysed endolysin sequences and their accession no. Key: NA = not applicable

hage	gene (locus_tag)	product	Description	Taxonomy
CDKM15	NA	NA	N-acetylmuramoyl-L-alanine amidase	Myoviridae
CDKM9	NA	NA	N-acetylmuramoyl-L-alanine amidase	Myoviridae
CDMH1	CDHM1_gp34	CDI66654.1	putative endolysin protein	Myoviridae
phiC2	phiC2p38	ABE99499.1	putative amidase/endolysin	Myoviridae
phiCD111	PHICD111_20024	CEK40298.1	Endolysin	Siphoviridae
phiCD146	PHICD146_20023	CEK40352.1	Endolysin	Siphoviridae
phiCD24-1	PHICD2401_20030	CEK40242.1	N-acetylmuramoyl-L-alanine amidase	Siphoviridae
phiCD27	phiCD27_gp34	ACH91325.1	Endolysin	Myoviridae
phiCD38-2	phiCD38-2_gp23	AEF56898.1	Endolysin	Siphoviridae
phiCD481-1	PHICD48101_20027	CEK40598.1	putative N-acetylmuramoyl-L-alanine amidase	Myoviridae
phiCD505	PHICD505_20034	CEK40659.1	N-acetylmuramoyl-L-alanine amidase	Myoviridae
phiCD506	PHICD506_20027	CEK40728.1	Endolysin	Myoviridae
phiCD6356	phiCD6356_28	ADK37890.1	probable phage N-acetylmuramoyl-L alanine amidase	Siphoviridae
phiCDHM11	phiCDHM11_gp25	CDL68842.1	putative endolysin protein	Myoviridae
phiCDHM13	phiCDHM13_gp25	CDL65308.1	putative endolysin protein	Myoviridae
phiCDHM14	phiCDHM14_gp25	CDU85312.1	putative endolysin protein	Myoviridae
phiCDHM19	phiCDHM19_gp40	CDW17221.1	putative endolysin protein	Myoviridae
phiMMP01	PHIMMP01_20036	CEK40790.1	Endolysin	Myoviridae
phiMMP02	D863_gp34	AFO72095.1	Endolysin	Myoviridae
phiMMP03	PHIMMP03_20039	CEK40874.1	N-acetylmuramoyl-L-alanine amidase	Myoviridae
phiMMP04	D864_gp27	AFO72163.1	Endolysin	Myoviridae
CDKS8	NA	NA	N-acetylmuramoyl-L-alanine amidase	Siphoviridae

8.6 Appendix 6

Table 8-6. Analysed sequences of TerL genes with the accession no.

Organism	Accession number	Description	Taxonomy
Aeromonas phage Aeh1	NP_944105.1	gp17 terminase DNA packaging enzyme large subunit	Myoviridae; Tevenvirinae; T4likevirus
Agrobacterium fabrum str. C58	NP_353972.2	large terminase phage packaging protein	Rhizobiaceae; Rhizobium/ Agrobacterium group; Agrobacterium*
Bacillus phage phi105	NP_690755.1	Un named protein product	Siphoviridae; Lambdalikevirus ; unclassified Lambda-like viruses
Bacillus phage SPO1	YP_002300330.1	gp2.11	Myoviridae; Spounavirinae; Spounalikevirus
Bacillus phage SPP1	NP_690654.1	hypothetical protein SPP1p003	Siphoviridae; Lambdalikevirus ; unclassified Lambda-like viruses
Brucella melitensis	WP_004683294.1	ATP-binding protein	Brucellaceae; Brucella*
Burkholderia phage phiE125	NP_536358.1	putative terminase (large subunit)	Siphoviridae; Phie125likevirus
Carboxydothermus hydrogenoformans	WP_011344575.1	terminase	Thermoanaerobacteraceae; Carboxydothermus*
Caulobacter crescentus CB15	NP_421586.1	hypothetical protein CC_2790	Caulobacteraceae; Caulobacter*
CDMH1	YP_009032144.1	putative phage terminase large subunit	Myoviridae
phiCD3626	NP_612831.1	putative terminase large subunit	Siphoviridae
phiC2	YP_001110720.1	putative terminase large subunit	Myoviridae
phiCD111	CEK40276.1	Terminase large subunit	Siphoviridae
phiCD119	YP_529553.1	putative terminase large subunit	Myoviridae
phiCD146	CEK40331.1	Terminase large subunit	Siphoviridae
phiCD211	CEK40383.1	conserved protein of unknown function	Myoviridae
phiCD24-1	CEK40213.1	Large terminase protein	Siphoviridae
phiCD27	YP_002290878.1	putative terminase B	Myoviridae
phiCD38-2	YP_004508380.1	terminase large subunit	Siphoviridae
phiCD481-1	CEK40573.1	putative phage terminase, large subunit	Myoviridae
phiCD505	CEK40627.1	putative terminase B	Myoviridae
phiCD506	CEK40703.1	putative phage terminase, large subunit	Myoviridae
phiCD6356	YP_004306103.1	putative terminase large subunit	Siphoviridae
phiCDHM11	CDL68818.1	putative terminase large subunit protein	Myoviridae;

Table 8-6 continued.

phiCDHM13	CDL65284.1	putative terminase large subunit protein	Myoviridae
phiCDHM14	CDU85288.1	putative terminase large subunit protein	Myoviridae
phiCDHM19	CDW17183.1	putative terminase large subunit protein	Myoviridae
Clostridium phage phiCP39-O	YP_002265410.1	large terminase	Siphoviridae
phiMMP01	CEK40756.1	putative phage terminase large subunit	Myoviridae
phiMMP02	YP_006990480.1	terminase large subunit	Myoviridae
phiMMP03	CEK40837.1	putative terminase large subunit	Myoviridae
phiMMP04	YP_006990556.1	terminase	Myoviridae
Deinococcus radiodurans R1	NP_285417.1	hypothetical protein DR_A0094	Deinococcaceae; Deinococcus*
Endosymbiont phage APSE-1	NP_050979.1	P18	Podoviridae
Enterobacteria phage 186	AAC34148.1	terminase subunit	Myoviridae; Peduovirinae; P2likevirus
Enterobacteria phage ES18	YP_224140.1	gp2	Siphoviridae
Enterobacteria phage HK022	NP_037663.1	terminase large subunit	Siphoviridae; Lambdalikevirus
Enterobacteria phage lambda	NP_040581.1	DNA packaging protein	Siphoviridae; Lambdalikevirus
Enterobacteria phage LP7	P16938.1	Terminase, large subunit	Podoviridae; P22likevirus
Enterobacteria phage Mu	NP_050632.1	putative portal protein	Myoviridae; Mulikevirus
Enterobacteria phage N15	NP_046897.1	gp2	Siphoviridae; N15likevirus
Enterobacteria phage P2	NP_046758.1	gpP	Myoviridae; Peduovirinae; P2likevirus
Enterobacteria phage P22	YP_063734.1	terminase large subunit	Podoviridae; P22likevirus
Enterobacteria phage phiP27	NP_543088.1	putative terminase	Myoviridae
Enterobacteria phage RB49	NP_891724.1	terminase DNA packaging enzyme large subunit	Myoviridae; Tevenvirinae; T4likevirus
Enterobacteria phage SP6	NP_853601.1	large terminase subunit	Podoviridae; Autographivirinae
Enterobacteria phage T3	NP_523347.1	DNA packaging protein B	Podoviridae; Autographivirinae; T7likevirus
Enterobacteria phage T4	NP_049776.1	gp17 terminase DNA packaging enzyme, large subunit	Myoviridae; Tevenvirinae; T4likevirus

Table 8-6 continued.

Enterobacteria phage T7	NP_042010.1	**DNA maturation protein**	**Podoviridae; Autographivirinae; T7likevirus**
Escherichia coli	WP_000934137.1	DNA packaging protein	Enterobacteriaceae; Escherichia*
Escherichia coli 042	CBG34179.1	phage terminase large subunit	Enterobacteriaceae; Escherichia*
Haemophilus influenzae Rd KW20	NP_439650.1	hypothetical protein HI1500	Pasteurellaceae; Haemophilus*
Haemophilus phage HP1	NP_043485.1	hypothetical protein HP1p21	Myoviridae; Peduovirinae; Hpunalikevirus
Klebsiella phage phiKO2	YP_006582.1	putative large terminase subunit	Siphoviridae
Lactobacillus phage A2	NP_680484.1	putative large terminase subunit	Siphoviridae
Lactobacillus phage LL-H	YP_001285878.1	putative terminase large subunit	Siphoviridae
Lactobacillus phage phiadh	NP_050148.1	hypothetical protein phiadhp40	Siphoviridae
Lactobacillus phage phig1e	NP_695170.1	putative terminase large subunit	Siphoviridae
Lactobacillus prophage Lj965	NP_958579.1	putative terminase large subunit	Siphoviridae
Lactococcus phage BK5-T	NP_116494.1	putative large terminase subunit	Siphoviridae
Lactococcus phage c2	NP_043560.1	terminase	Siphoviridae; C2likevirus
Lactococcus phage SK1	NP_044948.1	terminase large subunit	Siphoviridae; Skunalikevirus
Listeria phage A118	NP_463463.1	putative terminase large subunit	Siphoviridae
Listonella phage phiHSIC	YP_224236.1	putative terminase large subunit	Siphoviridae
Mycobacterium phage D29	NP_046828.1	gp13	Siphoviridae; L5likevirus
CDKM15		terminase large subunit	Myoviridae
CDKM9		terminase large subunit	Myoviridae
phiCDIF1296T	AKP44676.1	putative terminase large subunit	Myoviridae
Pseudomonad phage gh-1	NP_813786.1	DNA packaging protein B	Podoviridae; Autographivirinae; T7likevirus
Pseudomonas phage D3	NP_061498.1	terminase large subunit	Siphoviridae; D3likevirus
Pseudomonas phage phiCTX	NP_490600.1	predicted DNA-dependent ATPase terminase subunit	Myoviridae; Peduovirinae; P2likevirus
Pseudomonas phage phiKMV	NP_877482.1	putative DNA maturase B	Podoviridae; Autographivirinae
Pseudomonas putida KT2440	NP_744442.1	DNA maturase B	Pseudomonadaceae; Pseudomonas*
Rhodobacter capsulatus	AAF13179.1	putative gene transfer agent large terminase	Rhodobacteraceae; Rhodobacter*

Ruminiclostridium thermocellum	WP_020457819.1	terminase	Ruminococcaceae; Ruminiclostridium*
Salmonella enterica	WP_001526457.1	DNA packaging protein	Enterobacteriaceae; Salmonella*
Salmonella enterica subsp. enterica serovar Typhimurium str. LT2	NP_461544.1	terminase-like large protein	Enterobacteriaceae; Salmonella*
Salmonella phage HK620	NP_112076.1	terminase large subunit	Podoviridae; P22likevirus
Shewanella oneidensis MR-1	NP_716299.1	Mu phage large terminase subunit GpE	Shewanellaceae; Shewanella*
Staphylococcus phage phiETA	NP_510934.1	similar to phage O1205 ORF26 (putative large subunit terminase)	Siphoviridae; Phietalikevirus
Staphylococcus phage PVL	NP_058441.1	hypothetical protein PVL_02	Siphoviridae
Staphylococcus phage Twort	YP_238728.1	ORF059	Myoviridae; Spounavirinae; Twortlikevirus
Streptococcus phage MM1	NP_150161.1	putative large terminase subunit	Siphoviridae
Streptococcus phage O1205	NP_695104.1	hypothetical protein O1205p26	Siphoviridae; Sfi1unalikevirus
Streptococcus phage Sfi19	NP_049926.1	putative terminase large subunit	Siphoviridae; Sfi21dtunalikevirus
Streptococcus phage Sfi21	NP_049967.1	putative large subunit terminase	Siphoviridae; Sfi21dtunalikevirus
Streptomyces phage phiC31	NP_047924.1	gp33	Siphoviridae; Phic3unalikevirus
Vibrio phage KVP40	NP_899601.1	terminase DNA packaging enzyme large subunit	Myoviridae; Tevenvirinae; Schizot4likevirus
Wolbachia phage WO	BAA89621.1	orf2	unclassified phages
Yersinia phage phiA1122	NP_848309.1	DNA maturase B	Podoviridae; Autographivirinae; T7likevirus
Yersinia phage phiYeO3-12	NP_052122.1	DNA packaging protein B	Podoviridae; Autographivirinae; T7likevirus
Yersinia phage PY54	NP_892047.1	terminase large subunit	Siphoviridae
CDKS8	NA	Terminase large subunit	Siphoviridae

8.7 Appendix 7

Figure 8-1. Shows presence and absence of virulence genes in *C. difficile* isolates and the reference genome.

8.8 Appendix 8

Table 8-7. Diversity of CRISPR arrays encoded in 11 environmental *C. difficile* strains.

RTs	Number of CRISPR arrays	No. of spacer in each CRISPR array	spacers matches to phage sequences	Spacers matches to Plasmid sequences
RT091	8	3,12,4,11,10,5,28,4	9	
RT091	9	3,12,4,11,10,5,28,4,4	9	
RT001	10	3,14,6,17,7,8,9,6,15,5	3	
RT001	8	7,18,14,5,6,15,8,9	3	
RT035	8	12,11,11,3,11,6,31,4	6	1
RT011	9	11,9,13,3,3,11,21,12,6	6	
RT604	3	9,9,27	3	
RT604	3	27,9,9	3	
RT010	12	6,10,24,4,5,3,9,7,7,3,7,8	5	
RT091	8	3,12,4,11,10,5,28,4	9	
RT035	11	12,11,11,3,11,6,31,6,8,8,5	9	1

8.9 Appendix 9

Table 8-8. *C. difficile* CRISPR spacer match to phage sequences

Spacer ID	start	stop	Score	Strand	Spacer-sequences	Proto-spacer	Proto-spacer hit
CD105KSE1							
spacer 16	15702	15667	36	+	AGCCGUCUCUAAAUCCAUCUAUAACGUUACCAGCAA	TCGGCAGAGATTTAGGTAGATATTGCAATGGTCGTT	phiCD146
Spacer 74	13293	13329	37	-	AUGACUGCAAAUGACACAACUGAAUUUGUUGAUGUGA	TACTGACGTTTACTGTGTTGACTTAAACAACTACACT	phiMMP02
Spacer 9	21284	21248	37	+	AUUAUUACUUAACAUAAUCCUUGCAAUUCUAAGAAUU	TAATAATGAATTGTATTAGGAACGTTAAGATTCTTAA	phiMMP04
Spacer-73	15210	15174	37	+	CUAACUCUGCAAAACUAGAUUUUAAGUUAUCAAAAAC	GATTGAGACGTTTTGATCTAAAATTCAATAGTTTTTG	phiCD27
Spacer-31	3304	3267	38	+	CUAGAAUUAGAACUCAUUAUUAAAACCAUUCUUGCAAG	GATCTTAATCTTGAGTAATAATTTTGGTAAGAACGTTC	phiCD27
Spacer-31	3313	3276	38	+	CUAGAAUUAGAACUCAUUAUUAAAACCAUUCUUGCAAG	GATCTTAATCTTGAGTAATAATTTTGGTAAGAACGTTC	phiCD505
Spacer-31	3312	3275	38	+	CUAGAAUUAGAACUCAUUAUUAAAACCAUUCUUGCAAG	GATCTTAATCTTGAGTAATAATTTTGGTAAGAACGTTC	phiMMP02
_Spacer-36	39945	39981	37	-	GUCAAAGACAAACGUAUGGUCUUAAAUAUACAAGAAU	CAGTTTCTGTTTGCATACCAGAATTTATATGTTCTTA	phiCD111
Spacer-67	41462	41497	36	-	UGGAGAAUGGGCUAGAAACAAUCAAGAAAUUUGUGA	ACCTCTTACCCGATCTTTGTTAGTTCTTTAAACACT	phiCD505
Spacer-54	8992	9027	36	-	UUAGAACAAGAAUAUGCUAAAAAUGAAGAAAAAAUA	AATCTTGTTCTTATACGATTTTTACTTCTTTTTTAT	phiCD6356
Spacer-17	14551	14517	35	+	UUGCUAAGACCGCCGAAAAUCUCUACAGCUUUAUC	AACGATTCTGGCGGCTTTTAGAGATGTCGAAATAG	phiCD38_2

Table 8-8 contined.

CD105KSE2							
Spacer-15	15702	15667	36	+	AGCCGUCUCUAAAUCCAUCUAUAACGUUACCAGCAA	TCGGCAGAGATTTAGGTAGATATTGCAATGGTCGTT	phiCD146
Spacer-78	13293	13329	37	-	AUGACUGCAAAUGACACAACUGAAUUUGUUGAUGUGA	TACTGACGTTTACTGTGTTGACTTAAACAACTACACT	phiMMP02
Spacer-9	21284	21248	37	+	AUUAUUACUUAACAUAAUCCUUGCAAUUCUAAGAAUU	TAATAATGAATTGTATTAGGAACGTTAAGATTCTTAA	phiMMP04
Spacer-77	15210	15174	37	+	CUAACUCUGCAAAACUAGAUUUUAAGUUAUCAAAAAC	GATTGAGACGTTTTGATCTAAAATTCAATAGTTTTTG	phiCD27
Spacer-30	3304	3267	38	+	CUAGAAUUAGAACUCAUUAUUAAAACCAUUCUUGCAAG	GATCTTAATCTTGAGTAATAATTTTGGTAAGAACGTTC	phiCD27
Spacer-30	3313	3276	38	+	CUAGAAUUAGAACUCAUUAUUAAAACCAUUCUUGCAAG	GATCTTAATCTTGAGTAATAATTTTGGTAAGAACGTTC	phiCD505
Spacer-30	3312	3275	38	+	CUAGAAUUAGAACUCAUUAUUAAAACCAUUCUUGCAAG	GATCTTAATCTTGAGTAATAATTTTGGTAAGAACGTTC	phiMMP02
Spacer-35	39945	39981	37	-	GUCAAAGACAAACGUAUGGUCUUAAAUAUACAAGAAU	CAGTTTCTGTTTGCATACCAGAATTTATATGTTCTTA	phiCD111
Spacer-67	41462	41497	36	-	UGGAGAAUGGGCUAGAAACAAUCAAGAAAUUUGUGA	ACCTCTTACCCGATCTTTGTTAGTTCTTTAAACACT	phiCD505
Spacer-53	8992	9027	36	-	UUAGAACAAGAAUAUGCUAAAAAUGAAGAAAAAAUA	AATCTTGTTCTTATACGATTTTTACTTCTTTTTTAT	phiCD6356
spacer-16	14551	14517	35	+	UUGCUAAGACCGCCGAAAAUCUCUACAGCUUUAUC	AACGATTCTGGCGGCTTTTAGAGATGTCGAAATAG	phiCD38_2
CD105KSE3							

Table 8-8 contined.

Spacer-61	3388	3353	36	+	CUCUGUGACUCCAUAUUUUUUUG CAAUAUAUAAUUC	GAGACACTGAGGTATAAAAAA ACGTTATATATTAAG	phiCD146
Spacer-36	14871	14907	37	-	GAAAUAAUAAAUAGCAUUGUCGA AUGGUUUGCAAGCC	CTTTATTATTTATCGTAACAGC TTACCAAACGTTCGG	phiCD146
CD105KSE 4							
Spacer-61	3388	3353	36	+	CUCUGUGACUCCAUAUUUUUUUG CAAUAUAUAAUUC	GAGACACTGAGGTATAAAAAA ACGTTATATATTAAG	phiCD146
Spacer-36	14871	14907	37	-	GAAAUAAUAAAUAGCAUUGUCGA AUGGUUUGCAAGCC	CTTTATTATTTATCGTAACAGC TTACCAAACGTTCGG	phiCD146
CD105KSE 5							
Spacer-6	21284	21248	37	+	AUUAUUACUUAACAUAAUCCUUG CAAUUCUAAGAAUU	TAATAATGAATTGTATTAGGAA CGTTAAGATTCTTAA	phiMMP04
Spacer-41	4207	4172	36	+	GCAGUAACUUAUUAAAAUAUCAU CAUGACAUGAGGA	CGTCATTGAATAATTTTATAGT AGTACTGTACTCCT	pCD630
Spacer-81	42792	42828	37	-	GGAUAAAUGUUGAUAAAGAAUGU GUAUAUGCAAUGUU	CCTATTTACAACTATTTCTTAC ACATATACGTTACAA	phiMMP03
Spacer-81	46329	46365	37	-	GGAUAAAUGUUGAUAAAGAAUGU GUAUAUGCAAUGUU	CCTATTTACAACTATTTCTTAC ACATATACGTTACAA	phiC2
Spacer-44	20225	20260	36	-	GUUUUUAGAGGACCUAACUAUAA GUGGUAUAAAAAG	CAAAAATCTCCTGGATTGATAT TCACCATATTTTTC	phiMMP03
Spacer-44	20178	20213	36	-	GUUUUUAGAGGACCUAACUAUAA GUGGUAUAAAAAG	CAAAAATCTCCTGGATTGATAT TCACCATATTTTTC	phiC2
Spacer-13	4896	4860	37	+	UAAUUAUAUCGUAAGCAUGAUUG CUAUAUGAUUUAUA	ATTAATATAGCATTCGTACTAA CGATATACTAAATAT	phiCD211

Table 8-8 contined.

Spacer-63	8992	9027	36	-	UUAGAACAAGAAUAUGCUAAAAAUGAAGAAAAAAUA	AATCTTGTTCTTATACGATTTTTACTTCTTTTTTAT	phiCD6356
Spacer-82	46601	46636	36	-	UUGAAGAAGUUAUAAAAGAUACUAGAAGUCAAAAAA	AACTTCTTCAATATTTTCTATGATCTTCAGTTTTTT	phiCDMH1
Spacer-82	46279	46314	36	-	UUGAAGAAGUUAUAAAAGAUACUAGAAGUCAAAAAA	AACTTCTTCAATATTTTCTATGATCTTCAGTTTTTT	phiCDHM19
CD105KSE6							
Spacer-15	39709	39745	37	-	AAAUGGAGGAAUAGAAGUGAAAGAGAAAGUUAAGCGC	TTTACCTCCTTATCTTCACTTTCTCTTTCAATTCGCG	phiCD146
Spacer-55	7644	7681	38	-	AAUAACAGCAGAUACAGAUGUUGAAAACUCACCAACUG	TTATTGTCGTCTATGTCTACAACTTTTGAGTGGTTGAC	phiCD6356
Spacer-10	14630	14594	37	+	ACAUCUGCGUCAUAUCUAAAUAAAUUAUCUCCAACUU	TGTAGACGCAGTATAGATTTATTTAATAGAGGTTGAA	phiCD6356
Spacer-89	4595	4560	36	+	CAAUUUAGCAAUCUUACUAAACUCUCUAUACUCUUG	GTTAAATCGTTAGAATGATTTGAGAGATATGAGAAC	phiCD38_2
Spacer-41	5254	5292	39	-	GUAUCCAUGCAAGGCUGAUAUAUUUAGAGAAACAUAUGA	CATAGGTACGTTCCGACTATATAAATCTCTTTGTATACT	phiCD111
Spacer-53	41076	41040	37	+	UGCUGUAGCAUCUCUUCCAUUUUACUUCUCUUGCAA	ACGACATCGTAGAGAAGGTAAAAATGAAGAGAACGTT	phiMMP03
Spacer-53	41301	41265	37	+	UGCUGUAGCAUCUCUUCCAUUUUACUUCUCUUGCAA	ACGACATCGTAGAGAAGGTAAAAATGAAGAGAACGTT	phiCDMH1
Spacer-53	44613	44577	37	+	UGCUGUAGCAUCUCUUCCAUUUUACUUCUCUUGCAA	ACGACATCGTAGAGAAGGTAAAAATGAAGAGAACGTT	phiC2
CD105KSO7							

Table 8-8 contined.

Spacer-39	9240	9204	37	+	AAGCUUAACUAUUUCAUCUUUAUUUUUACUUACAACA	TTCGAATTGATAAAGTAGAAATAAAAATGAATGTTGT	phiMMP02
Spacer-39	9241	9205	37	+	AAGCUUAACUAUUUCAUCUUUAUUUUUACUUACAACA	TTCGAATTGATAAAGTAGAAATAAAAATGAATGTTGT	phiCD505
Spacer-13	25595	25631	37	-	UAGUUUUUGACAUUACUUUAGCUAAUGCUAACUUCUC	ATCAAAAACTGTAATGAAATCGATTACGATTGAAGAG	phiCD27
Spacer-13	25312	25348	37	-	UAGUUUUUGACAUUACUUUAGCUAAUGCUAACUUCUC	ATCAAAAACTGTAATGAAATCGATTACGATTGAAGAG	phiMMP02
CD105KSO8							
Spacer-22	9241	9205	37	+	AAGCUUAACUAUUUCAUCUUUAUUUUUACUUACAACA	TTCGAATTGATAAAGTAGAAATAAAAATGAATGTTGT	phiCD505
Spacer-22	9240	9204	37	+	AAGCUUAACUAUUUCAUCUUUAUUUUUACUUACAACA	TTCGAATTGATAAAGTAGAAATAAAAATGAATGTTGT	phiMMP02
Spacer-40	25595	25631	37	-	UAGUUUUUGACAUUACUUUAGCUAAUGCUAACUUCUC	ATCAAAAACTGTAATGAAATCGATTACGATTGAAGAG	phiCD27
Spacer-40	25312	25348	37	-	UAGUUUUUGACAUUACUUUAGCUAAUGCUAACUUCUC	ATCAAAAACTGTAATGAAATCGATTACGATTGAAGAG	phiMMP02
CD105KSE9							
Spacer-71	121633	121670	38	-	AUGUCUAGCUUAUUUUUAUAUUUAAUUUUAAAGAUUAA	TACAGATCGAATAAAAATATAAATTAAAATTTCTAATT	phiCD211
Spacer-27	30188	30224	37	-	CACAUUGGCGUUGAAGCUGUUAGCAAAGCAUUAAAAG	GTGTAACCGCAACTTCGACAATCGTTTCGTAATTTTC	phiMMP04

Table 8-8 contined.

Spacer-28	19502	19467	36	+	GUAGCCUAUUGUUUGGCUUCCUGACAUUUCGUAAGA	CATCGGATAACAAACCGAAGGACTGTAAAGCATTCT	CDHS1
Spacer-28	19000	18965	36	+	GUAGCCUAUUGUUUGGCUUCCUGACAUUUCGUAAGA	CATCGGATAACAAACCGAAGGACTGTAAAGCATTCT	phiCD146
Spacer-28	18991	18956	36	+	GUAGCCUAUUGUUUGGCUUCCUGACAUUUCGUAAGA	CATCGGATAACAAACCGAAGGACTGTAAAGCATTCT	phiCD38_2
Spacer-65	121670	121633	38	+	UUAAUCUUUAAAAUUAAAUAUAAAAAUAAGCUAGACAU	AATTAGAAATTTTAATTTATATTTTTATTCGATCTGTA	phiCD211
Spacer-30	121005	120971	35	+	UUUAAAUGGUGUGAAUGGUGAAAGGCACACAACAC	AAATTTACCACACTTACCACTTTCCGTGTGTTGTG	phiCD211
CD105KSO10							
Spacer-16	15702	15667	36	+	AGCCGUCUCUAAAUCCAUCUAUAACGUUACCAGCAA	TCGGCAGAGATTTAGGTAGATATTGCAATGGTCGTT	phiCD146
Spacer-75	13293	13329	37	-	AUGACUGCAAAUGACACAACUGAAUUUGUUGAUGUGA	TACTGACGTTTACTGTGTTGACTTAAACAACTACACT	phiMMP02
Spacer-9	21284	21248	37	+	AUUAUUACUUAACAUAAUCCUUGCAAUUCUAAGAAUU	TAATAATGAATTGTATTAGGAACGTTAAGATTCTTAA	phiMMP04
Spacer-74	15210	15174	37	+	CUAACUCUGCAAAACUAGAUUUUAAGUUAUCAAAAAC	GATTGAGACGTTTTGATCTAAAATTCAATAGTTTTTG	phiCD27
Spacer-31	3313	3276	38	+	CUAGAAUUAGAACUCAUUAUUAAAACCAUUCUUGCAAG	GATCTTAATCTTGAGTAATAATTTTGGTAAGAACGTTC	phiCD505
Spacer-31	3312	3275	38	+	CUAGAAUUAGAACUCAUUAUUAAAACCAUUCUUGCAAG	GATCTTAATCTTGAGTAATAATTTTGGTAAGAACGTTC	phiMMP02
Spacer-31	3304	3267	38	+	CUAGAAUUAGAACUCAUUAUUAAAACCAUUCUUGCAAG	GATCTTAATCTTGAGTAATAATTTTGGTAAGAACGTTC	phiCD27

Table 8-8 contined.

Spacer-36	39945	39981	37	-	GUCAAAGACAAACGUAUGGUCUU AAAUAUACAAGAAU	CAGTTTCTGTTTGCATACCAGA ATTTATATGTTCTTA	phiCD111
Spacer-68	41462	41497	36	-	UGGAGAAUGGGCUAGAAACAAUC AAGAAAUUUGUGA	ACCTCTTACCCGATCTTTGTTA GTTCTTTAAACACT	phiCD505
Spacer-54	8992	9027	36	-	UUAGAACAAGAAUAUGCUAAAAA UGAAGAAAAAAUA	AATCTTGTTCTTATACGATTTTT ACTTCTTTTTTAT	phiCD6356
Spacer-17	14551	14517	35	+	UUGCUAAGACCGCCGAAAAUCUC UACAGCUUUAUC	AACGATTCTGGCGGCTTTTAGA GATGTCGAAATAG	phiCD38_2
CD105KSE 11							
Spacer -6	21284	21248	37	+	AUUAUUACUUAACAUAAUCCUUG CAAUUCUAAGAAUU	TAATAATGAATTGTATTAGGAA CGTTAAGATTCTTAA	phiMMP04
Spacer-104	3388	3353	36	+	CUCUGUGACUCCAUAUUUUUUUG CAAUAUAUAAUUC	GAGACACTGAGGTATAAAAAA ACGTTATATATTAAG	phiCD146
Spacer-41	4207	4172	36	+	GCAGUAACUUAUUAAAAUAUCAU CAUGACAUGAGGA	CGTCATTGAATAATTTTATAGT AGTACTGTACTCCT	pCD630,
Spacer-81	42792	42828	37	-	GGAUAAAUGUUGAUAAAGAAUGU GUAUAUGCAAUGUU	CCTATTTACAACTATTTCTTAC ACATATACGTTACAA	phiMMP03
Spacer-81	46329	46365	37	-	GGAUAAAUGUUGAUAAAGAAUGU GUAUAUGCAAUGUU	CCTATTTACAACTATTTCTTAC ACATATACGTTACAA	phiC2
Spacer-90	14907	14871	37	+	GGCUUGCAAACCAUUCGACAAUG CUAUUUAUUAUUUC	CCGAACGTTTGGTAAGCTGTTA CGATAAATAATAAAG	phiCD146
Spacer-44	20225	20260	36	-	GUUUUUAGAGGACCUAACUAUAA GUGGUAUAAAAAG	CAAAAATCTCCTGGATTGATAT TCACCATATTTTTC	phiMMP03
Spacer-44	20178	20213	36	-	GUUUUUAGAGGACCUAACUAUAA GUGGUAUAAAAAG	CAAAAATCTCCTGGATTGATAT TCACCATATTTTTC	phiC2

Table 8-8 contined.

Spacer-13	4896	4860	37	+	UAAUUAUAUCGUAAGCAUGAUUG CUAUAUGAUUUAUA	ATTAATATAGCATTCGTACTAA CGATATACTAAATAT	phiCD211
Spacer-63	8992	9027	36	-	UUAGAACAAGAAUAUGCUAAAAA UGAAGAAAAAAUA	AATCTTGTTCTTATACGATTTTT ACTTCTTTTTTAT	phiCD6356
Spacer-82	46279	46314	36	-	UUGAAGAAGUUAUAAAAGAUACU AGAAGUCAAAAAA	AACTTCTTCAATATTTTCTATG ATCTTCAGTTTTTT	phiCDHM19
Spacer-82	46601	46636	36	-	UUGAAGAAGUUAUAAAAGAUACU AGAAGUCAAAAAA	AACTTCTTCAATATTTTCTATG ATCTTCAGTTTTTT	CDMH1

8.10 Appendix 10

Table 8-9. Host range of phages against environmental (shaded) and clinical strains of *C. difficile*.

Environmental strains included strains from local environment (in red) and from strains from UK environment. Clinical isolates include MLVA typed R027 (in blue). Phages are in columns and isolates in rows, ground according ribotype. Symbol: ○ indicates clear lysis , ◙ indicates turbid zones.

		Phages																	
Ribotype	Isolate	CDKS2	CDKS3	CDKS4	CDKS5	CDKS6	CDKS7	CDKS8	CDKS9	CDKS10	CDKS12	CDK13	CDKS14	CDKS15	DDKS1	CDKM9	CDKM1	CDKM4
001/10	CD001		○	○	○	○	○	○			○			○	○			
	CDNCTC1120				○					○				○		○	○	
	R4001		◙	○	○	○	○	○		◙	○	○	○	○	○			
	H18		○	◙		○	○	○										
	F6	○	○	○	○	○	○	○	○	○	○			○	○			
	F7		○		○				○									
	F8	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
	F5	○	○	○	○	○	○	○	○	○	○	○	○	○				
	F2	◙	◙	○		○	○	○		○								
	F1		◙	○		○	○	◙								◙	◙	○
002	H15																	
	AQV																	
005	AOY	○	○	○	○	○	○	○	○	○	○	○	○	○	○			
	AOO	◙	◙		◙	◙		○		◙	◙	○	◙	◙				
010/7	H3	○	○	○	○	○	○	○	○	○	○	○	○	○	○			
	K16	○	◙	○	◙	○	◙	◙	◙	◙	◙	◙	◙	◙	◙			
	R10								○									
	BRQ																	
	BTQ																	
	M4			○	○	○	○	○	○	◙	○	○	○	○	○			
	M7	○		○	○	○	○	○		○	○			○				
011	F9		4	5	6	6	5	6	5	6	6	5	5	6	4			
	J5																	
012	CD630																	
	K10																	
	CD10																	
014	K6															○	○	
	V															○	○	○
014/20	ATJ															○	○	○
	ATT															○	○	○
015	Q															○		
017	M68																	

Table 8-9 continued

	CF5																	
021	K9																	
023	UK	○		○	○	○	○	○	○	○	○	○	○	○	○			
	AQA																	
031	H11																	
	M																	
027	AIU	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	
	K15	○	○	○	○	○	○	○		○	○	○	○	○	○	○	○	
	HB	○	○	○	○	○	○	○		○	○	○	○	○	○			
	7L	○	○	○	○	○	○	○	○	○	○	○	○	○	○			
	9L	○	○	○	○	○	○	○	○	○	○	○	○	○	○			
	14L	○	○	○	○	○	○	○	○	○	○	○	○	○	○			
	15L	○	○	○	○	○	○	○	○	○	○	○	○	○	○			
	16L	○	○	○	○	○	○	○	○	○	○	○		○				
	17L	○	○	○	○	○	○	○	○	○	○	○		○				
	22L		○	○	○		○	○	○	○	○	○	○	○	○			
	28L	○	○	○	○	○	○	○	○	○	○	○	○	○	○			
	46L	○	○	○	○	○	○	○	○	○	○	○	○	○	○			
	90L	○	○	○	○	○	○	○	○	○	○	○		○	○			
	R2029	○	○			○	○	○	○	○	○	○	○	○				
	B1-9	○	○			○	○	○		○		○	○		○			
	CD196	○	○	○	○	○	○	○	○	○	○		○	○	○			
	AJV	○	○	○	○	○	○	○	○	○	○	○	○	○	○			
	CD007		○	○		○		○					○					
036	J6															○		
	J10															○		
076	T6															○	○	○
078	AIA																	
	ALL	○	◙	○	○	○	○	○	○	○	○	○		○	◙			
	ANL		○	○			○	○		○	○							
	ANJ																	
	AMT																	
	AMM																	
	ASS																	
	ALE																	
	AQA																	
	AHS																	
	AKJ	◙		◙	◙	◙	◙	○	◙	◙	◙	◙		◙	◙			
	AKM																	
	AKP																	
	H17																	
	H5c							○										
	J	○	○	○	○	○	○	○	○	○	○	○		○	○			
091	D1																	
	D6																	

Table 8-9 continued.

	D7																	
	D8																	
	D9																	
	S2																	
085	R8																	
106	CD106															○	○	
	R40V0106																◙	
	R6106																◙	
	R39V106																◙	○
604	K2																	
	K5																	
	K6																	
691	C																	
	B1																	
220	H12																◙	
	K12															○	○	
	Y															○		
	K3																	
	K															○	○	○

8.11 Appendix 11

Table 8-10. Genome annotation for CDKM9.

Symbols indicate sense (+) or antisense (-) strand.

Locus-tag	Start	End	Strand	Description	Top BLASTp hit	top hit organism	E- value
1	1	702	+	Terminase small subunit	Putative terminase small subunit	phiCD27	2E-167
2	695	2104	+	Terminase large subunit	Multispecies: terminase B	*PeptoClostridium difficile*	0
3	2095	3597	+	Portal protein	Multispecies: portal protein	*PeptoClostridium difficile*	0
4	3584	4597	+	Head morphogenesis protein	Putative head morphogenesis protein	*PeptoClostridium difficile* CD8	0
5	4617	5951	+	Uncharacterized protein	Multispecies: hypothetical protein	*PeptoClostridium difficile*	0
6	5967	6350	+	Uncharacterized protein	Multispecies: hypothetical protein	*PeptoClostridium difficile*	2E-86
7	6369	7406	+	Uncharacterized protein	Multispecies: hypothetical protein	*PeptoClostridium difficile*	0
8	7446	7784	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	7E-75
9	7781	8158	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	1E-82
10	8124	8570	+	Uncharacterized protein	Conserved exported hypothetical protein	*PeptoClostridium difficile* T23	2E-104
11	8571	9392	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	0
12	9405	9608	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	2E-36
13	9610	11031	+	Tail sheath	Phage tail sheath family protein	*PeptoClostridium difficile*	0
14	11047	11469	+	XkdM-related protein	Hypothetical protein	*PeptoClostridium difficile*	9E-95

Table8-10 contiuned.

15	11548	12021	+	XkdN-related protein	Phage XkdN-like family protein	*PeptoClostridium difficile*	5E-105
16	12202	16530	+	Tail tape measure	Phage tail tape measure protein	*PeptoClostridium difficile*	0
17	16610	17404	+	Uncharacterized protein	Hypothetical protein QOS_2956	*PeptoClostridium difficile* Y184	0
18	17474	18145	+	LysM family protein	lysM domain protein	*PeptoClostridium difficile* F548	3E-157
19	18142	19104	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	0
20	19097	19459	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	3E-80
21	19437	19895	+	XkdS-related protein	Hypothetical protein	*PeptoClostridium difficile*	1E-105
22	19904	21040	+	Baseplate J family protein	Baseplate J-like family protein	*PeptoClostridium difficile*	0
23	21037	21666	+	Tail protein	Hypothetical protein QC1_0346	*PeptoClostridium difficile* CD21	2E-149
24	21679	22785	+	Tail fiber protein	Phage tail-collar fiber family protein	*PeptoClostridium difficile*	0
25	22799	23437	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	2E-149
26	23447	23659	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	1E-37
27	23659	23841	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	2E-33
28	23876	25111	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	0
29	25108	25407	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	2E-63
30	25409	25546	+	XkdX-related protein	hypothetical protein	*PeptoClostridium difficile*	1E-22
31	25608	26039	+	Holin	Toxin secretion/phage lysis holin family protein	*PeptoClostridium difficile*	2E-96

Table8-10 continued.

32	26039	26878	+	N-acetylmuramoyl-L-alanine amidase	N-acetylmuramoyl-L-alanine amidase	*PeptoClostridium difficile*	0
33	26974	27321	+	Uncharacterized protein	Hypothetical protein QAO_2927	*PeptoClostridium difficile* CD3	2E-74
34	27487	28005	+	Uncharacterized protein	Multispecies: hypothetical protein	*Peptoclostridium*	5E-113
35	28069	29655	-	Uncharacterized protein	Multispecies: hypothetical protein	*Peptoclostridium*	0
36	29631	30152	-	Uncharacterized protein	Multispecies: hypothetical protein	*Peptoclostridium*	5E-124
37	30854	31042	-	Uncharacterized protein	Multispecies: hypothetical protein	*Peptoclostridium*	6E-34
38	31300	31443	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	1E-23
39	31950	32327	+	BlaI/MecI/CopY family transcriptional regulator	Penicillinase repressor family protein	*PeptoClostridium difficile*	2E-79
40	32437	32820	+	BlaI/MecI/CopY family transcriptional regulator	Transcriptional regulator	*PeptoClostridium difficile*	2E-82
41	33131	34345	-	Integrase	Putative integrase	phiC2	0
42	34435	34650	-	Uncharacterized protein	Multispecies: hypothetical protein	*Peptoclostridium*	2E-41
43	34700	34921	-	XRE family transcriptional regulator	Multispecies: hypothetical protein	*Peptoclostridium*	2E-42
44	35379	35774	-	XRE family transcriptional regulator	Repressor	*PeptoClostridium difficile*	4E-85

Table 8.10 continued.

45	35953	36186	+	betR domain DNA-binding protein	BetR domain protein	*PeptoClostridium difficile*	3E-46
46	36186	36974	+	KilAC/pRha domain antirepressor	Antirepressor	*PeptoClostridium difficile*	0
47	37031	37168	+	Uncharacterized protein	Hypothetical protein QK3_0979	*PeptoClostridium difficile DA00145*	7E-21
48	37180	37383	-	Arc-like DNA binding domain protein	hypothetical protein phiC2p53	phiC2	3E-39
49	37470	37646	+	Uncharacterized protein	ribbon-helix-helix, copG family protein	*PeptoClostridium difficile* DA00145	3E-33
50	37639	38277	+	BRO N-terminal domain regulatory protein	antirepressor	*PeptoClostridium difficile*	6E-152
51	38336	38716	+	Uncharacterized protein	hypothetical protein	*PeptoClostridium difficile*	4E-82
52	38706	38858	-	Uncharacterized protein	hypothetical protein phiC2p57	phiC2	1E-23
53	38917	39114	+	XRE family transcriptional regulator	Repressor	*PeptoClostridium difficile*	6E-36
54	39153	39587	+	Uncharacterized protein	hypothetical protein QAQ_3920	*PeptoClostridium difficile CD8*	3E-95
55	39588	39734	+	Uncharacterized protein	Hypothetical protein QAQ_3919	*PeptoClostridium difficile CD8*	2E-25
56	39737	40054	+	Uncharacterized protein	Multispecies: hypothetical protein	*Peptoclostridium*	5E-61
57	40136	40411	+	Uncharacterized protein	hypothetical protein	*PeptoClostridium difficile*	5E-49
58	40404	41306	+	Bet recombination protein	recombination protein Bet	*PeptoClostridium difficile*	0

Table 8.10 continued.

59	41321	42115	+	Uncharacterized protein	hypothetical protein	*PeptoClostridium difficile*	4E-166
60	42178	42351	+	Uncharacterized protein	Phage resolvase/ integrase, partial	*PeptoClostridium difficile*	4E-30
61	42367	42774	+	single-stranded DNA-binding protein	hypothetical protein	*PeptoClostridium difficile*	3E-93
62	42802	43365	+	Uncharacterized protein	hypothetical protein phiCD27_gp58	phiCD27	3E-130
63	43379	43744	+	Uncharacterized protein	Putative regulatory protein	phiCDHM19	7E-82
64	43744	43878	+	Uncharacterized protein	hypothetical protein phiC2p71	phiC2	4E-21
65	43862	44197	+	Uncharacterized protein	hypothetical protein	*PeptoClostridium difficile*	2E-69
66	44306	45379	+	DNA-cytosine methyltransferase	DNA (cytosine-5-)-methyltransferase	*PeptoClostridium difficile*	0
67	45672	45788	+	Uncharacterized protein	putative phage protein	*PeptoClostridium difficile* DA00114	9E-18
68	45742	45861	+	Uncharacterized protein	putative phage protein	*PeptoClostridium difficile* DA00114	8E-19
69	45946	46131	+	Uncharacterized protein	hypothetical protein	*PeptoClostridium difficile*	2E-33
70	46145	46384	+	Uncharacterized protein	phage protein	*PeptoClostridium difficile*	7E-50
71	46403	46516	+	Uncharacterized protein	hypothetical protein	*PeptoClostridium difficile*	5E-15
72	46513	46920	+	RusA Endonuclease	endodeoxyribonuclease RusA family protein	*PeptoClostridium difficile*	7E-89
73	47022	47870	+	KilA-N/ORF6 family anti-repressor	Multispecies: anti-repressor	*Peptoclostridium*	0
74	48013	48501	+	Uncharacterized protein	Sigma factor	*PeptoClostridium difficile*	1E-110
75	49083	49760	+	BRO N-terminal domain regulatory protein	Putative regulatory protein	phiCDMH1	6E-162

8.12 Appendix 12

Table 8-11. Genomic annotation for CDKM15.

Symbols indicate sense (+) or antisense (-).

Locus-tag	Start	End	Strand	Description	Top BLASTp hit	Top hit organism	Evalue
1	1	702	+	Terminase small subunit	ATPase subunit of terminase family protein	*PeptoClostridium difficile*	1E-158
2	695	2104	+	Terminase large subunit	Putative terminase B	phiCD505	0
3	2095	3597	+	Portal protein	Portal protein	*PeptoClostridium difficile*	0
4	3584	4597	+	Head morphogenesis protein	Head morphogenesis protein	phiMMP02	0
5	4617	5951	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	0
6	5967	6350	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	4E-87
7	6368	7405	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	0
8	7445	7783	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	1E-74
9	7780	8157	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	4E-78
10	8159	8569	+	Uncharacterized protein	Multispecies: hypothetical protein	*PeptoClostridium difficile*	2E-91
11	8572	8748	+	Uncharacterized protein	Hypothetical protein QCK_4159	*PeptoClostridium*	7E-32

						difficile CD45	
12	8766	9587	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	0
13	9600	9803	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	3E-37
14	9805	11226	+	Tail sheath protein	Phage tail sheath family protein	*PeptoClostridium difficile*	0
15	11242	11664	+	XkdM-related protein	Hypothetical protein	*PeptoClostridium difficile*	2E-93
16	11741	12214	+	XkdN-related protein	Phage XkdN-like family protein	*PeptoClostridium difficile*	6E-103

Table 8.11 continued.

17	12395	16684	+	Tail tape measure protein	Phage tail tape measure protein	*PeptoClostridium difficile*	0
18	16765	17685	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	0
19	18529	18711	+	Uncharacterized protein	CopG family transcriptional regulator	*PeptoClostridium difficile*	4E-32
20	18832	19716	+	BRO N-terminal domain protein	Hypothetical protein	*PeptoClostridium difficile*	0
21	19779	19982	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	9E-37
22	20158	20316	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	6E-27
23	20679	21350	+	LysM family protein	LysM domain protein	*PeptoClostridium difficile* DA00256	6E-156
24	21347	22309	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	0
25	22302	22664	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	3E-80
26	22642	23100	+	XkdS-related protein	Hypothetical protein	*PeptoClostridium difficile*	1E-105
27	23109	24245	+	Baseplate J family protein	baseplate J-like family protein	*PeptoClostridium difficile*	0
28	24242	24871	+	Tail protein	Hypothetical protein QC1_0346	*PeptoClostridium*	2E-149

						difficile CD21	
29	24884	25990	+	Tail fiber protein	Phage tail-collar fiber family protein	*PeptoClostridium difficile*	0
30	26004	26642	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	2E-149
31	26652	26864	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	1E-37
32	26864	27046	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	2E-33
33	27081	28316	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	0

Table 8.11 continued.

34	28313	28612	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	2E-63
35	28614	28751	+	XkdX-related protein	Hypothetical protein	*PeptoClostridium difficile*	1E-22
36	28813	29244	+	Holin	Toxin secretion/phage lysis holin family protein	*PeptoClostridium difficile*	2E-96
37	29244	30056	+	N-acetylmuramoyl-L-alanine amidase	Endolysin	*PeptoClostridium difficile*	0
38	30306	30797	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	1E-112
39	30799	31365	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	4E-129
40	31775	31963	-	Uncharacterized protein	Multispecies: hypothetical protein	*PeptoClostridium difficile*	6E-34
41	32221	32364	+	Uncharacterized protein	hypothetical protein	*PeptoClostridium difficile*	1E-23
42	32872	33249	+	BlaI/MecI/CopY family transcriptional regulator	Penicillinase repressor family protein	*PeptoClostridium difficile*	1E-78
43	33354	33737	+	BlaI/MecI/CopY family transcriptional regulator	Penicillinase repressor family protein	*PeptoClostridium difficile*	8E-82
44	34048	35214	-	Integrase	Multispecies: phage integrase	*PeptoClostridium difficile*	0
45	35352	35558	-	Uncharacterized protein	Hypothetical protein QAQ_3859	*PeptoClostridium difficile* CD8	1E-39

46	35617	35838	-	Uncharacterized protein	CI repressor	*PeptoClostridium difficile*	1E-42
47	36296	36691	-	Cro/C1-type transcriptional regulator	Repressor	*PeptoClostridium difficile*	3E-84

Table 8.11 continued.

48	36870	37103	+	BetR domain DNA binding protein	BetR domain protein	*PeptoClostridium difficile*	3E-46
49	37103	37855	+	Antirepressor	Phage regulatory, Rha family protein	*PeptoClostridium difficile*	1E-179
50	37900	38154	+	Uncharacterized protein	Hypothetical protein QC7_1110	*PeptoClostridium difficile* CD38	6E-52
51	38191	38625	+	Uncharacterized protein	Hypothetical protein QC9_1023	*PeptoClostridium difficile* CD39	3E-95
52	38626	38742	+	Uncharacterized protein	Hypothetical protein QAW_1262	*PeptoClostridium difficile* CD17	1E-15
53	38753	38941	+	Uncharacterized protein	Phage protein	*PeptoClostridium difficile*	5E-34
54	38973	39116	+	Uncharacterized protein	Hypothetical protein QAW_1264	*PeptoClostridium difficile* CD17	1E-23
55	39076	39255	-	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	3E-32
56	39389	39547	+	Arc-like DNA binding domain protein	Hypothetical protein phiC2p61	Clostridium phage phiC2	2E-28
57	39547	39774	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	4E-45
58	39872	40186	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	8E-64
59	40270	40779	+	Sipho_Gp157 familly protein	Hypothetical protein	*PeptoClostridium difficile*	1E-110

60	40789	41397	+	Essential recombination function family protein	Essential recombination function protein	*PeptoClostridium difficile*	2E-141
61	41407	42288	+	Replication protein DnaD	dnaD domain protein	*PeptoClostridium difficile*	0
62	42349	42522	+	Resolvase	Hypothetical protein	*PeptoClostridium difficile*	2E-31

Table 8.11 continued.

63	42538	42945	+	single-stranded DNA-binding protein	Hypothetical protein	*PeptoClostridium difficile*	9E-94
64	43019	43522	+	Uncharacterized protein	Hypothetical protein QCY_3052	*PeptoClostridium difficile* CD70	5E-116
65	43536	43901	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	9E-82
66	43901	44035	+	Uncharacterized protein	Hypothetical protein phiC2p71	phage phiC2	4E-21
67	44019	44354	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	2E-69
68	44462	45535	+	DNA-cytosine methyltransferase (EC 2.1.1.37) (Putative phage DNA modification methylase)	Multispecies: restriction endonuclease subunit M	*PeptoClostridium difficile*	0
69	45615	45824	+	Uncharacterized protein	Multispecies: hypothetical protein	*PeptoClostridium difficile*	7E-40
70	45836	46840	+	37-kD nucleoid-associated family protein	Hypothetical protein	*PeptoClostridium difficile*	0
71	46849	47016	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	7E-30
72	47101	47373	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	3E-55
73	47396	47626	+	PUA domain protein	ASCH domain protein	*PeptoClostridium difficile*	7E-45
74	47678	48178	+	Uncharacterized protein	Hypothetical protein phiMMP02_gp72	phiMMP02	8E-113
75	48168	48611	+	Holliday junction resolvase	RusA endonuclease	*PeptoClostridium difficile*	2E-102
76	48601	48810	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	8E-40
77	48915	49421	+	Uncharacterized protein	Sigma factor	*PeptoClostridium difficile*	8E-115
78	49700	49936	+	Uncharacterized protein	Hypothetical protein	*PeptoClostridium difficile*	4E-46
79	49933	50571	+	Membrane protein	Hypothetical protein	*PeptoClostridium difficile*	3E-146

8.13 Appendix 13

Table 8-12. Genome annotation for CDKS8.

Symbols indicate sense (+) or antisense (-) strand

Locus-tag	Start	End	Length	Strand	Description	Top BLASTP hit	Top hit organism	E-value
1	26	736	708	+	Terminase small subunit	Terminase small subunit	phiCD111	6.00E-169
2	723	2045	1323	+	Terminase large subunit	Terminase large subunit	phiCD111	0.0
3	2,046	3458	1413	+	Portal protein	Minor capsid protein	phiCD111	0.0
4	3459	4862	1224	+	Minor capsid protein	Minor capsid protein	phiCD111	0.0
5	4864	5010	327	+	Hypothetical protein	Conserved protein of unknown function	phiCD111	3.00E-68
6	5026	5307	282	+	Hypothetical protein	Phage-related protein	phiCD111	4.00E-58
7	5439	6011	573	+	Scaffolding protein	Phage-related protein	phiCD111	1.00E-56
8	6022	7095	1074	+	Coat protein	Phage coat protein	*PeptoClostridium difficile*	0.0
9	7099	7431	333	+	Head-tail connector protein	Phage coat protein	*PeptoClostridium difficile*	0.0
10	7433	7867	435	+	Hypothetical protein	Hypothetical protein	*PeptoClostridium difficile*	3.00E-99

Table 8.12 continued.

11	7874	8221	348	+	Minor capsid protein	Phage capsid family protein	*PeptoClostridium difficile*	8.00E-77
12	8218	8649	432	+	Hypothetical protein	Hypothetical protein	*PeptoClostridium difficile*	1.00E-98
13	8652	9452	801	+	Hypothetical protein	Hypothetical protein	*PeptoClostridium difficile*	0.0
14	9510	9872	636	+	Hypothetical protein	Conserved protein of unknown function	phiCD146	2.00E-73
15	9872	10480	609	+	GP15 family protein	Bacteriophage Gp15 family protein	*PeptoClostridium difficile*	8.00E-130
16	10552	11193	642	+	Hypothetical protein	Conserved protein of unknown function	phiCD111	8.00E-152
17	11299	16602	5304	+	Tail tape measure protein	Tail protein / endopeptidase	phiCD146	0.0
18	16723	17457	735	+	Tail protein	Hypothetical protein	*PeptoClostridium difficile*	2.00E-50
19	17493	19586	2094	+	Tail endopeptidase	Tail fibre protein	phiCD38-2	0.0
20	19613	19855	243	+	Hypothetical protein	Uncharacterized protein	*PeptoClostridium difficile*	2E-50
21	19870	22332	2463	+	Tail fibre protein	Tail component protein	phiCD146	4.00E-176
22	22332	24182	1851	+	Tail fibre protein	Tail protein	*PeptoClostridium difficile*	0.0
23	24268	24531	264	+	Holin	Holin	*PeptoClostridium difficile*	4.00E-50

Table 8.12 continued.

24	24532	25338	807	+	N-acetylmuramoyl -L-alanine amidase	Cell wall hydrolase	*PeptoClostridium difficile*	0.0
25	25387	25599	213		Hypothetical protein	Hypothetical protein	*PeptoClostridium difficile*	4.00E-37
26	25778	26608	831	+	Hypothetical protein	Hypothetical protein	*PeptoClostridium difficile*	0.0
27	26687	27193	507	+	Hypothetical protein	Hypothetical protein	*PeptoClostridium difficile*	1.00E-113
28	27213	27683	471	+	Hypothetical protein	Hypothetical protein	*PeptoClostridium difficile*	5.00E-106
29	27916	28197	282	−	Hypothetical protein	Hypothetical protein	*PeptoClostridium difficile*	4.00E-62
30	28341	28673	333	−	Hypothetical protein	Hypothetical protein	*PeptoClostridium difficile*	3.00E-69
31	28674	29483	810	_	CobQ/CobB/MinD/ParA nucleotide binding domain containing protein	Putative ParA protein	phiCDHM11	0.0
32	29772	29954	183	_	Hypothetical protein	Hypothetical protein	*PeptoClostridium difficile*	4.00E-35
33	30194	30322	129	_	Hypothetical protein	Hypothetical protein phiCD38-2_gp32	phiCD38-2	7.00E-17
34	30775	30927	153	−	Hypothetical protein	Hypothetical protein	*PeptoClostridium difficile*	2E-26
35	31447	33612	2166	+	DNA helicase	Putative DNAse/helicase protein	phiCD146	0.0

Table 8.12 continued.

36	33659	34006	348	+	Hypothetical protein	Conserved hypothetical protein	phiCDHM13	5.00E-26
37	34840	35193	354	_	Cro/C1-type transcriptional regulator	Transcriptional regulator	*PeptoClostridium difficile*	6.00E-77
38	35366	35548	183	+	Hypothetical protein	Hypothetical protein	*PeptoClostridium difficile*	1.00E-32
39	35615	36040	426	+	Hypothetical protein	Hypothetical protein	*PeptoClostridium difficile*	2.00E-72
40	36115	36519	405	+	Replication terminator like protein	Replication terminator protein	phiCD111	4.00E-91
41	36542	37255	714	+	Hypothetical protein	Conserved protein of unknown function	phiCD111	6.00E-84
42	37311	37475	165	+	Hypothetical protein	Conserved protein of unknown function	phiCD146	9.00E-144
43	37472	37783	312	+	Hypothetical protein	Conserved protein of unknown function]	phiCD146	2.00E-54
44	37829	37798	150	+	Hypothetical protein	Hypothetical protein phiCD38-2_gp46	phiCD38-2	2.00E-25
45	38009	38182	174	+	Hypothetical protein	Conserved protein of unknown function	phiCD111	1.00E-28

Table 8.12 continued.

46	38182	38298	117	+	Hypothetical protein	Hypothetical protein	*PeptoClostridium difficile*	3.00E-17
47	38295	38621	327	+	Single-stranded DNA-binding protein	ssDNA binding protein	phiCD38-2	4.00E-71
48	38713	39078	366	+	Hypothetical protein	Hypothetical protein	*PeptoClostridium difficile*	1.00E-82
49	39062	39235	174	+	Hypothetical protein	Hypothetical protein phiCD38-2_gp51	phiCD38-2	4.00E-30
50	39247	39819	573	+	Sigma70/sigma F-like protein	Sigma70/sigma F-like protein	phiCD111	1.00E-128
51	39997	40605	609	+	Integrase	Integrase	*PeptoClostridium difficile*	5.00E-143
52	40796	40921	126	+	Hypothetical protein	Conserved protein of unknown function	phiCD111	6.00E-17
53	41070	41639	571	+	Recombinase/resolvase	Resolvase	*PeptoClostridium difficile*	6.00E-135

8.14 Appendix 14

Statistical Analysis of the Growth comparison for the environmental and clinical strains at 37°C.

Summary of Two-way ANOVA between the strains of R027 at 37°C

Two-way ANOVA Ordinary

P value P < 0.05

Source of Variation	% of total variation	P value	P value summary	Significant?
Interaction	0.9594	0.0866	ns	No
Row Factor	88.80	< 0.0001	****	Yes
Column Factor	0.2358	0.0431	*	Yes

ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
Interaction	0.7548	10	0.07548	F (10, 176) = 1.688	P = 0.0866
Row Factor	69.87	10	6.987	F (10, 176) = 156.3	P < 0.0001
Column Factor	0.1855	1	0.1855	F (1, 176) = 4.150	P = 0.0431
Residual	7.869	176	0.04471		

Number of missing values 0

Summary of Two-way ANOVA between the strains of R010 at 37°C

Two-way ANOVA Ordinary

P value P < 0.05

Source of Variation	% of total variation	P value	P value summary	Significant?
Interaction	0.4615	0.139	ns	No
Row Factor	94.02	< 0.0001	****	Yes
Column Factor	0.1350	0.0371	*	Yes

ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
Interaction	0.3520	10	0.03520	F (10, 176) = 1.508	P = 0.1398
Row Factor	71.72	10	7.172	F (10, 176) = 307.2	P < 0.0001
Column Factor	0.1030	1	0.1030	F (1, 176) = 4.411	P = 0.0371
Residual	4.108	176	0.02334		
Number of missing values	0				

Summary of Two-way ANOVA between the strains of R078 at 37°C

Two-way ANOVAOrdinary

P value P < 0.05

Source of Variation	% of total variation	P value	P value summary	Significant?
Interaction	0.8351	0.0189	*	Yes
Row Factor	92.34	< 0.0001	****	Yes
Column Factor	0.1870	0.0272	*	Yes

ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
Interaction	0.7173	10	0.07173	F (10, 176) = 2.215	P = 0.0189
Row Factor	79.32	10	7.932	F (10, 176) = 244.9	P < 0.0001
Column Factor	0.1607	1	0.1607	F (1, 176) = 4.960	P = 0.0272
Residual	5.701	176	0.03239		
Number of missing values	0				

8.15 Appendix 15

Statistical Analysis of the Growth comparison for the environmental and clinical strains at 23°C

Summary of Two-way ANOVA between the strains of R078 at 23°C

Two-way ANOVA Ordinary

P value P < 0.05

Source of Variation	% of total variation	P value	P value summary	Significant?
Interaction	1.218	0.1531	ns	No
Row Factor	83.84	< 0.0001	****	Yes
Column Factor	0.3771	0.0342	*	Yes

ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
Interaction	0.3902	10	0.03902	F (10, 176) = 1.472	P = 0.1531
Row Factor	26.85	10	2.685	F (10, 176) = 101.3	P < 0.0001
Column Factor	0.1208	1	0.1208	F (1, 176) = 4.557	P = 0.0342
Residual	4.664	176	0.02650		

Number of missing values 0

Summary of Two-way ANOVA between the strains of R010 at 23°C

Two-way ANOVA Ordinary

P value P < 0.05

Source of Variation	% of total variation	P value	P value summary	Significant?
Interaction	0.6510	0.2240	ns	No
Row Factor	90.45	< 0.000	****	Yes
Column Factor	0.2010	0.0452	*	Yes

ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
Interaction	0.2215	10	0.02215	F (10, 176) = 1.318	P = 0.2240

Row Factor	30.77	10	3.077	F (10, 176) = 183.1	P < 0.0001
Column Factor	0.06840	1	0.06840	F (1, 176) = 4.070	P = 0.0452
Residual	2.958	176	0.01680		

Number of missing values 0

Summary of Two-way ANOVA between the strains of R027 at 23°C

Two-way ANOVA Ordinary

P value P < 0.05

Source of Variation	% of total variation	P value	P value summary	Significant?
Interaction	0.6510	0.2240	ns	No
Row Factor	90.45	< 0.0001	****	Yes
Column Factor	0.2010	0.0452	*	Yes

ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
Interaction	0.2215	10	0.02215	F (10, 176) = 1.318	P = 0.2240
Row Factor	30.77	10	3.077	F (10, 176) = 183.1	P < 0.0001
Column Factor	0.06840	1	0.06840	F (1, 176) = 4.070	P = 0.0452
Residual	2.958	176	0.01680		

Number of missing values 0

8.16 Appendix 16

Statistical Analysis of the Growth comparison for the environmental and clinical strains at pH value 8

Summary of Two-way ANOVA between the strains of R010 at pH 8

Two-way ANOVA Ordinary

P value P < 0.05

Source of Variation	% of total variation	P value	P value summary	Significant?
Interaction	0.1824	0.9770	ns	No
Row Factor	89.29	< 0.0001	****	Yes
Column Factor	0.2853	0.0281	*	Yes

ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
Interaction	0.1614	10	0.01614	F (10, 176) = 0.3133	P = 0.9770
Row Factor	79.01	10	7.901	F (10, 176) = 153.4	P < 0.0001
Column Factor	0.2524	1	0.2524	F (1, 176) = 4.900	P = 0.0281
Residual	9.068	176	0.05152		

Number of missing values 0

Summary of Two-way ANOVA between the strains of R078 at pH 8

Two-way ANOVA Ordinary

P value P < 0.05

Source of Variation	% of total variation	P value	P value summary	Significant?
Interaction	0.3159	0.3754	ns	No
Row Factor	94.38	< 0.0001	****	Yes
Column Factor	0.1802	0.0137	*	Yes

ANOVA table	SS	DF	MS	F (DFn, DFd)	P value

Interaction	0.2747	10	0.02747	F (10, 176) = 1.086	P = 0.3754
Row Factor	82.08	10	8.208	F (10, 176) = 324.5	P < 0.0001
Column Factor	0.1567	1	0.1567	F (1, 176) = 6.194	P = 0.0137
Residual	4.452	176	0.02530		
Number of missing values	0				

Summary of Two-way ANOVA between the strains of R027 at pH 8

Two-way ANOVA	Ordinary
P value	P < 0.05

Source of Variation	% of total variation	P value	P value summary	Significant?
Interaction	0.3912	0.6221	ns	No
Row Factor	91.89	< 0.0001	****	Yes
Column Factor	0.2593	0.0220	*	Yes

ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
Interaction	0.2509	10	0.02509	F (10, 154) = 0.8072	P = 0.6221
Row Factor	58.92	10	5.892	F (10, 154) = 189.6	P < 0.0001
Column Factor	0.1663	1	0.1663	F (1, 154) = 5.351	P = 0.0220
Residual	4.786	154	0.03108		
Number of missing values	0				

8.17 Appendix 17

Statistical Analysis of the Growth comparison for the environmental and clinical strains at 2% salinity

Summary of Two-way ANOVA between the strains of R010 at 2% (w/v) salinity.

Two-way ANOVA Ordinary

P value P < 0.05

Source of Variation	% of total variation	P value	P value summary	Significant?
Interaction	5.165	0.0009	***	Yes
Row Factor	80.86	< 0.0001	****	Yes
Column Factor	1.282	0.0049	**	Yes

ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
Interaction	1.087	10	0.1087	F (10, 99) = 3.343	P = 0.0009
Row Factor	17.03	10	1.703	F (10, 99) = 52.35	P < 0.0001
Column Factor	0.2699	1	0.2699	F (1, 99) = 8.298	P = 0.0049
Residual	3.220	99	0.03252		

Number of missing values 11

Summary of Two-way ANOVA between the strains of R078 at 2% salinity

Source of Variation	% of total variation	P value	P value summary	Significant?
Interaction	0.7702	0.8476	ns	No
Row Factor	76.93	< 0.0001	****	Yes
Column Factor	0.9665	0.0091	**	Yes

ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
Interaction	0.1625	10	0.01625	F (10, 154) = 0.5559	P = 0.8476

Row Factor	16.23	10	1.623	F (10, 154) = 55.52	P < 0.0001
Column Factor	0.2039	1	0.2039	F (1, 154) = 6.975	P = 0.0091
Residual	4.501	154	0.02923		
Number of missing values	0				

Summary of Two-way ANOVA between the strains of R027 at 2% (w/v) salinity.

P value summary	Significant?		
Interaction	3.508	0.2051	ns	No
Row Factor	64.47	< 0.0001	****	Yes
Column Factor	2.770	0.0013	**	Yes

ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
Interaction	0.8199	10	0.08199	F (10, 121) = 1.364	P = 0.2051
Row Factor	15.07	10	1.507	F (10, 121) = 25.07	P < 0.0001
Column Factor	0.6475	1	0.6475	F (1, 121) = 10.77	P = 0.0013
Residual	7.274	121	0.06012		
Number of missing values	11				

8.18 Appendix 18

Statistical Analysis of Toxin A and B production within the strains of ribotype 078 over 24 h period

Two-way ANOVA Ordinary

Source of Variation	% of total variation	P value	P value summary	Significant?
Interaction	9.601	< 0.0001	****	Yes
Row Factor	76.72	< 0.0001	****	Yes
Column Factor	7.835	< 0.0001	****	Yes

ANOVA table

SS	DF	MS	F (DFn, DFd)	P value
Interaction	6.202	15	0.4134	F (15, 48) = 5.261	P < 0.0001
Row Factor	49.56	5	9.912	F (5, 48) = 126.1	P < 0.0001
Column Factor	5.061	3	1.687	F (3, 48) = 21.47	P < 0.0001
Residual	3.772	48	0.07859		

Number of missing values 0

Statistical Analysis of Toxin Aand B production within the strains of ribotype 027over 24 h period

Two-way ANOVA Ordinary

Source of Variation	% of total variation	P value summary	Significant?
Interaction	7.595	< 0.0001	****	Yes
Row Factor	89.05	< 0.0001	****	Yes
Column Factor	1.562	0.0001	***	Yes

ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
Interaction	6.780	5	1.356	F (5, 24) = 20.35	P < 0.0001

Row Factor	79.50	5	15.90	F (5, 24) = 238.6	P < 0.0001
Column Factor	1.395	1	1.395	F (1, 24) = 20.93	P = 0.0001
Residual	1.599	24	0.06662		
Number of missing values	0				

8.19 Appendix 19

Statistical Analysis of total spore production of environmental and clinical strains

Paired t test analysis for R027 strains in terms of total spore production

Column B(K15)e

vs.

Column A(AIU)c

Paired t test

P value	0.0489
P value summary	*
Significantly different? (P < 0.05)	Yes
One- or two-tailed P value?	Two-tailed
t, df	t=13.00 df=1
Number of pairs	2

How big is the difference?

Mean of differences	-0.6500
SD of differences	0.07071
SEM of differences	0.05000
95% confidence interval	-1.285 to -0.01469

Statistical Analysis of total spore production of environmental and clinical strains of R010

Ordinary one-way ANOVA summary

F 7.701

P value 0.0388

P value summary *

Are differences among means statistically significant? (P < 0.05) Yes

R square 0.8524

Brown-Forsythe test

F (DFn, DFd) +infinity (3, 4)

P value < 0.0001

P value summary ****

Significantly different standard deviations? (P < 0.05) Yes

Bartlett's test

Bartlett's statistic (corrected)

ANOVA table SS DF MS F (DFn, DFd) P value

Treatment (between columns) 1.444 3 0.4813 F (3, 4) = 7.701 P = 0.0388

Residual (within columns) 0.2500 4 0.06250

Total 1.694 7

Data summary

Number of treatments (columns) 4

Number of values (total) 8

Statistical Analysis of total spore production of environmental and clinical strains of R078

Ordinary one-way ANOVA summary

F 4.766

P value 0.0418

P value summary *

Are differences among means statistically significant? (P < 0.05) Yes R square 0.7988

Brown-Forsythe test

F (DFn, DFd) +infinity (5, 6)

P value < 0.0001

P value summary ****

Significantly different standard deviations? (P < 0.05) Yes

Bartlett's test

Bartlett's statistic (corrected)

P value

P value summary

Significantly different standard deviations? (P < 0.05)

ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
Treatment (between columns)	2.561	5	0.5123	F (5, 6) = 4.766	P = 0.0418
Residal (within columns)	0.6450	6	0.1075		
Total	3.206	11			

Number of treatments (columns) 6

Number of values (total) 12

Chapter 9 Bibiliography

ABEDON, S. T. 2008. *Bacteriophage ecology: population growth, evolution, and impact of bacterial viruses*, Cambridge University Press.

ACKERMANN, H.-W. 2003. Bacteriophage observations and evolution. *Research in Microbiology,* 154**,** 245-251.

ACKERMANN, H.-W. 2006. 5500 Phages examined in the electron microscope. *Archives of Virology,* 152**,** 227-243.

ACKERMANN, H.-W. & KRISCH, H. M. 2014. A catalogue of T4-type bacteriophages. *Archives of Virology,* 142**,** 2329-2345.

ACKERMANN, H. W. 2001. Frequency of morphological phage descriptions in the year 2000. Brief review. *Arch Virol,* 146**,** 843-57.

ACKERMANN, H. W. 2007. 5500 Phages examined in the electron microscope. *Arch Virol,* 152**,** 227-43.

ADAMS, M. H. 1959. Bacteriophages. *Bacteriophages*.

ADHYA, S. & MERRIL, C. 2006. The road to phage therapy. *Nature,* 443**,** 754-755.

ÅGREN, J., SUNDSTRÖM, A., HÅFSTRÖM, T. & SEGERMAN, B. 2012. Gegenees: fragmented alignment of multiple genomes for determining phylogenomic distances and genetic signatures unique for specified target groups. *PloS One,* 7**,** e39107.

AKERLUND, T., PERSSON, I., UNEMO, M., NOREN, T., SVENUNGSSON, B., WULLT, M. & BURMAN, L. G. 2008. Increased sporulation rate of epidemic *Clostridium difficile* Type 027/NAP1. *J Clin Microbiol,* 46**,** 1530-3.

ÅKERLUND, T., SVENUNGSSON, B., LAGERGREN, Å. & BURMAN, L. G. 2006. Correlation of Disease Severity with Fecal Toxin Levels in Patients with *Clostridium difficile*-Associated Diarrhea and Distribution of PCR Ribotypes and Toxin Yields In Vitro of Corresponding Isolates. *Journal of Clinical Microbiology,* 44**,** 353-358.

AKHTER, S., AZIZ, R. K. & EDWARDS, R. A. 2012. PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. *Nucleic Acids Res,* 40**,** e126.

AKTORIES, K. & WILKINS, T. D. 2000. *Clostridium difficile*, Springer Science & Business Media.

AL-TAWFIQ, J. A. & ABED, M. S. 2010. *Clostridium difficile*-associated disease among patients in Dhahran, Saudi Arabia. *Travel Med Infect Dis,* 8.

AL-THANI, A. A., HAMDI, W. S., AL-ANSARI, N. A. & DOIPHODE, S. H. 2014. Polymerase chain reaction ribotyping of *Clostridium difficile* isolates in Qatar: a hospital-based study. *BMC Infect Dis,* 14**,** 502.

AL SAIF, N. & BRAZIER, J. S. 1996. The distribution of *Clostridium difficile* in the environment of South Wales. *J Med Microbiol,* 45**,** 133-7.

ALIKHAN, N. F., PETTY, N. K., BEN ZAKOUR, N. L. & BEATSON, S. A. 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics,* 12**,** 402.

ALONSO, R., GROS, S., PELÁEZ, T., GARCÍA-DE-VIEDMA, D., RODRíGUEZ-CRÉIXEMS, M. & BOUZA, E. 2001. Molecular analysis of relapse vs re-infection in HIV-positive patients suffering from recurrent *Clostridium difficile* associated diarrhoea. *Journal of Hospital Infection,* 48**,** 86-92.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *J Mol Biol,* 215**,** 403-10.

ANANTHAKRISHNAN, A. N. 2011. *Clostridium difficile* infection: epidemiology, risk factors and management. *Nat Rev Gastroenterol Hepatol,* 8**,** 17-26.

ANDREWS, S. 2010. FastQC: A quality control tool for high throughput sequence data. *Reference Source*.

ARROYO, L. G., KRUTH, S. A., WILLEY, B. M., STAEMPFLI, H. R., LOW, D. E. & WEESE, J. S. 2005. PCR ribotyping of *Clostridium difficile* isolates originating from human and animal sources. *J Med Microbiol,* 54**,** 163-6.

ASLAM, S., HAMILL, R. J. & MUSHER, D. M. 2005. Treatment of *Clostridium difficile*-associated disease: old therapies and new strategies. *Lancet Infect Dis,* 5**,** 549-57.

AUBRY, A., HUSSACK, G., CHEN, W., KUOLEE, R., TWINE, S. M., FULTON, K. M., FOOTE, S., CARRILLO, C. D., TANHA, J. & LOGAN, S. M. 2012. Modulation of toxin production by the flagellar regulon in *Clostridium difficile*. *Infection and immunity,* 80**,** 3521-3532.

AVILA, M. B., AVILA, N. P. & DUPONT, A. W. 2016. Recent Advances in the Diagnosis and Treatment of *Clostridium difficile* Infection. *F1000 Research,* 5**,** F1000 Faculty Rev-118.

BABAN, S. T., KUEHNE, S. A., BARKETI-KLAI, A., CARTMAN, S. T., KELLY, M. L., HARDIE, K. R., KANSAU, I., COLLIGNON, A. & MINTON, N. P. 2013. The Role of Flagella in *Clostridium difficile* Pathogenesis: Comparison between a Non-Epidemic and an Epidemic Strain. *PLoS ONE,* 8**,** e73026.

BACON, A. E., FEKETY, R., SCHABERG, D. R. & FAIX, R. G. 1988. Epidemiology of *Clostridium difficile* colonization in newborns: results using a bacteriophage and bacteriocin typing system. *J Infect Dis,* 158**,** 349-54.

BAGDASARIAN, N., RAO, K. & MALANI, P. N. 2015. Diagnosis and treatment of *Clostridium difficile* in adults: a systematic review. *Jama,* 313**,** 398-408.

BAKKER, D., CORVER, J., HARMANUS, C., GOORHUIS, A., KEESSEN, E., FAWLEY, W., WILCOX, M. & KUIJPER, E. 2010. Relatedness of human and animal *Clostridium difficile* PCR ribotype 078 isolates determined on the basis of multilocus variable-number tandem-repeat analysis and tetracycline resistance. *Journal of clinical microbiology,* 48**,** 3744-3749.

BAKRI, M. M., BROWN, D. J., BUTCHER, J. P. & SUTHERLAND, A. D. 2009. *Clostridium difficile* in ready-to-eat salads, Scotland. *Emerg Infect Dis,* 15.

BANKEVICH, A., NURK, S., ANTIPOV, D., GUREVICH, A. A., DVORKIN, M., KULIKOV, A. S., LESIN, V. M., NIKOLENKO, S. I., PHAM, S., PRJIBELSKI, A. D., PYSHKIN, A. V., SIROTKIN, A. V., VYAHHI, N., TESLER, G., ALEKSEYEV, M. A. & PEVZNER, P. A. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol,* 19**,** 455-77.

BARBUT, F., DELMÉE, M., BRAZIER, J. S., PETIT, J. C., POXTON, I. R., RUPNIK, M., LALANDE, V., SCHNEIDER, C., MASTRANTONIO, P., ALONSO, R., KUIPJER, E. & TVEDE, M. 2003. A European survey of diagnostic methods and testing protocols for *Clostridium difficile*. *Clinical Microbiology and Infection,* 9**,** 989-996.

BARBUT, F., MASTRANTONIO, P., DELMEE, M., BRAZIER, J., KUIJPER, E. & POXTON, I. 2007. Prospective study of *Clostridium difficile* infections in Europe with phenotypic and genotypic characterisation of the isolates. *Clin Microbiol Infect,* 13**,** 1048-57.

BARNETT, J. S. 2012. CME *Clostridium difficile* : A new look at an old but increasingly deadly infection. *Journal of the American Academy of Physician Assistants,* 25**,** 32-36.

BARRANGOU, R., FREMAUX, C., DEVEAU, H., RICHARDS, M., BOYAVAL, P., MOINEAU, S., ROMERO, D. A. & HORVATH, P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science,* 315**,** 1709-12.

BARROSO, L., WANG, S.-Z., PHELPS, C., JOHNSON, J. & WILKINS, T. 1990. Nucleotide sequence of *Clostridium difficile* toxin B gene. *Nucleic Acids Research,* 18**,** 4004-4004.

BARTH, H., AKTORIES, K., POPOFF, M. R. & STILES, B. G. 2004. Binary bacterial toxins: biochemistry, biology and applications of common *Clostridium* and *Bacillus* proteins. *Microbiol Mol Biol Revs,* 68.

BARTLETT, J. G., CHANG, T. W., GURWITH, M., GORBACH, S. L. & ONDERDONK, A. B. 1978. Antibiotic-associated pseudomembranous colitis due to toxin-producing clostridia. *New England Journal of Medicine,* 298**,** 531-534.

BAUER, M. P., NOTERMANS, D. W., VAN BENTHEM, B. H., BRAZIER, J. S., WILCOX, M. H., RUPNIK, M., MONNET, D. L., VAN DISSEL, J. T. & KUIJPER, E. J. 2011. *Clostridium difficile* infection in Europe: a hospital-based survey. *Lancet,* 377.

BAVERUD, V., GUSTAFSSON, A., FRANKLIN, A., ASPAN, A. & GUNNARSSON, A. 2003. *Clostridium difficile* : prevalence in horses and environment, and antimicrobial susceptibility. *Equine Vet J,* 35**,** 465-71.

BERGLUND, B. 2015. Environmental dissemination of antibiotic resistance genes and correlation to anthropogenic contamination with antibiotics. *Infection Ecology & Epidemiology,* 5**,** 10.3402/iee.v5.28564.

BESEMER, J. & BORODOVSKY, M. 2005. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Res,* 33**,** W451-4.

BEST, E. L., FAWLEY, W. N., PARNELL, P. & WILCOX, M. H. 2010. The potential for airborne dispersal of *Clostridium difficile* from symptomatic patients. *Clin Infect Dis,* 50**,** 1450-7.

BEVERIDGE, T. J. 1990. Mechanism of gram variability in select bacteria. *J Bacteriol,* 172**,** 1609-20.

BHAYA, D., DAVISON, M. & BARRANGOU, R. 2011. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu Rev Genet,* 45**,** 273-97.

BIDET, P., BARBUT, F., LALANDE, V., BURGHOFFER, B. & PETIT, J. 1999. Development of a new PCR-ribotyping method for *Clostridium difficile* based on ribosomal RNA gene sequencing. *FEMS Microbiol Lett,* 175.

BIDET, P., LALANDE, V., SALAUZE, B., BURGHOFFER, B., AVESANI, V., DELMEE, M., ROSSIER, A., BARBUT, F. & PETIT, J. C. 2000. Comparison of PCR-ribotyping, arbitrarily primed PCR, and pulsed-field gel electrophoresis for typing *Clostridium difficile*. *J Clin Microbiol,* 38**,** 2484-7.

BIGNARDI, G. E. 1998. Risk factors for *Clostridium difficile* infection. *Journal of Hospital Infection,* 40**,** 1-15.

BISWAS, A., GAGNON, J. N., BROUNS, S. J. J., FINERAN, P. C. & BROWN, C. M. 2013. CRISPRTarget. *RNA Biology,* 10**,** 817-827.

BLOSSOM, D. B. & MCDONALD, L. C. 2007. The challenges posed by reemerging *Clostridium difficile* infection. *Clin Infect Dis,* 45**,** 222-7.

BOLOTIN, A., QUINQUIS, B., SOROKIN, A. & EHRLICH, S. D. 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology,* 151**,** 2551-61.

BORDALO, A. A., ONRASSAMI, R. & DECHSAKULWATANA, C. 2002. Survival of faecal indicator bacteria in tropical estuarine waters (Bangpakong River, Thailand). *J Appl Microbiol,* 93**,** 864-71.

BORRIELLO, S. 1979. *Clostridium difficile* and its toxin in the gastrointestinal tract in health and disease. *Res Clin Forums,* 1**,** 33-35.

BORRIELLO, S. 1998. Pathogenesis of *Clostridium difficile* infection. *Journal of antimicrobial chemotherapy,* 41**,** 13-19.

BORRIELLO, S., DAVIES, H. A. & BARCLAY, F. E. 1988. Detection of fimbriae amongst strains of *Clostridium difficile*. *FEMS microbiology letters,* 49**,** 65-67.

BORRIELLO, S. P., HONOUR, P., TURNER, T. & BARCLAY, F. 1983. Household pets as a potential reservoir for *Clostridium difficile* infection. *J Clin Pathol,* 36**,** 84-7.

BOUDRY, P., SEMENOVA, E., MONOT, M., DATSENKO, K. A., LOPATINA, A., SEKULOVIC, O., OSPINA-BEDOYA, M., FORTIER, L. C., SEVERINOV, K., DUPUY, B. & SOUTOURINA, O. 2015. Function of the CRISPR-Cas System of the Human Pathogen *Clostridium difficile*. *MBio,* 6**,** e01112-15.

BOUVET, P. J. & POPOFF, M. R. 2008. Genetic relatedness of *Clostridium difficile* isolates from various origins determined by triple-locus sequence analysis based on toxin regulatory genes tcdC, tcdR, and cdtR. *J Clin Microbiol,* 46**,** 3703-13.

BOUZA, E. 2012. Consequences of *Clostridium difficile infection* : understanding the healthcare burden. *Clin Microbiol Infect,* 18 Suppl 6**,** 5-12.

BRAUN, V., HUNDSBERGER, T., LEUKEL, P., SAUERBORN, M. & VON EICHEL-STREIBER, C. 1996. Definition of the single integration site of the pathogenicity locus in *Clostridium difficile*. *Gene,* 181**,** 29-38.

BRAZIER, J. S. 1998. The epidemiology and typing of *Clostridium difficile*. *J Antimicrob Chemother,* 41 Suppl C**,** 47-57.

BRAZIER, J. S. 2001. Typing of *Clostridium difficile*. *Clinical Microbiology and Infection,* 7**,** 428-431.

BRAZIER, J. S. & BORRIELLO, S. P. 2000. Microbiology, epidemiology and diagnosis of *Clostridium difficile* infection. *Curr Top Microbiol Immunol,* 250**,** 1-33.

BREITBART, M., SALAMON, P., ANDRESEN, B., MAHAFFY, J. M., SEGALL, A. M., MEAD, D., AZAM, F. & ROHWER, F. 2002. Genomic analysis of uncultured marine viral communities. *Proc Natl Acad Sci U S A,* 99**,** 14250-5.

BROUWER, M. S., ALLAN, E., MULLANY, P. & ROBERTS, A. P. 2012. Draft genome sequence of the nontoxigenic *Clostridium difficile* strain CD37. *J Bacteriol,* 194**,** 2125-6.

BROWN, K., VALENTA, K., FISMAN, D., SIMOR, A. & DANEMAN, N. 2015. Hospital ward antibiotic prescribing and the risks of *Clostridium difficile* infection. *JAMA Intern Med,* 175**,** 626-33.

BRÜSSOW, H., CANCHAYA, C. & HARDT, W.-D. 2004. Phages and the Evolution of Bacterial Pathogens: from Genomic Rearrangements to Lysogenic Conversion. *Microbiology and Molecular Biology Reviews,* 68**,** 560-602.

BUFFIE, C. G., JARCHUM, I., EQUINDA, M., LIPUMA, L., GOBOURNE, A., VIALE, A., UBEDA, C., XAVIER, J. & PAMER, E. G. 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.

BUGGY, B. P., HAWKINS, C. C. & FEKETY, R. 1985. Effect of adding sodium taurocholate to selective media on the recovery of *Clostridium difficile* from environmental surfaces. *J Clin Microbiol,* 21**,** 636-7.

BUGGY, B. P., WILSON, K. H. & FEKETY, R. 1983. Comparison of methods for recovery of *Clostridium difficile* from an environmental surface. *J Clin Microbiol,* 18**,** 348-52.

BURKHARDT III, W., CALCI, K. R., WATKINS, W. D., RIPPEY, S. R. & CHIRTEL, S. J. 2000. Inactivation of indicator microorganisms in estuarine waters. *Water Research,* 34**,** 2207-2214.

BURNS, D. A., HEAP, J. T. & MINTON, N. P. 2010. The diverse sporulation characteristics of *Clostridium difficile* clinical isolates are not associated with type. *Anaerobe,* 16**,** 618-22.

BURNS, D. A., HEEG, D., CARTMAN, S. T. & MINTON, N. P. 2011. Reconsidering the sporulation characteristics of hypervirulent *Clostridium difficile* BI/NAP1/027. *PLoS One,* 6**,** e24894.

BURNS, D. A. & MINTON, N. P. 2011. Sporulation studies in *Clostridium difficile*. *Journal of Microbiological Methods,* 87**,** 133-138.

BURROWES, B., HARPER, D. R., ANDERSON, J., MCCONVILLE, M. & ENRIGHT, M. C. 2011. Bacteriophage therapy: potential uses in the control of antibiotic-resistant pathogens. *Expert Review of Anti-infective Therapy,* 9**,** 775+.

CAIRNS, M. D., STABLER, R. A., SHETTY, N. & WREN, B. W. 2012. The continually evolving *Clostridium difficile* species. *Future Microbiol,* 7**,** 945-57.

CALABI, E., WARD, S., WREN, B., PAXTON, T., PANICO, M., MORRIS, H., DELL, A., DOUGAN, G. & FAIRWEATHER, N. 2001. Molecular characterization of the surface layer proteins from *Clostridium difficile*. *Mol Microbiol,* 40**,** 1187-99.

CAMPBELL, A. 2003. The future of bacteriophage biology. *Nat Rev Genet,* 4**,** 471-7.

CANN, A. J. 2005. *Principles of Molecular Virology (Standard Edition),* Burlington, GB, Academic Press.

CAPELLA-GUTIÉRREZ, S., SILLA-MARTÍNEZ, J. M. & GABALDÓN, T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics,* 25**,** 1972-1973.

CARTER, G. P., CHAKRAVORTY, A., PHAM NGUYEN, T. A., MILETO, S., SCHREIBER, F., LI, L., HOWARTH, P., CLARE, S., CUNNINGHAM, B., SAMBOL, S. P., CHEKNIS, A., FIGUEROA, I., JOHNSON, S., GERDING, D., ROOD, J. I., DOUGAN, G., LAWLEY, T. D. & LYRAS, D. 2015. Defining the Roles of TcdA and TcdB in Localized Gastrointestinal Disease, Systemic Organ Damage, and the Host Response during *Clostridium difficile* Infections. *MBio,* 6**,** e00551.

CARTER, G. P., LYRAS, D., ALLEN, D. L., MACKIN, K. E., HOWARTH, P. M., O'CONNOR, J. R. & ROOD, J. I. 2007. Binary Toxin Production in *Clostridium difficile* Is Regulated by CdtR, a LytTR Family Response Regulator. *Journal of Bacteriology,* 189**,** 7290-7301.

CARTWRIGHT, C. P., STOCK, F., BEEKMANN, S. E., WILLIAMS, E. C. & GILL, V. J. 1995. PCR amplification of rRNA intergenic spacer regions as a method for epidemiologic typing of *Clostridium difficile*. *J Clin Microbiol,* 33**,** 184-7.

CARVER, T., HARRIS, S. R., BERRIMAN, M., PARKHILL, J. & MCQUILLAN, J. A. 2012. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics,* 28**,** 464-9.

CASJENS, S. R. 2005. Comparative genomics and evolution of the tailed-bacteriophages. *Curr Opin Microbiol,* 8**,** 451-8.

CASJENS, S. R. & GILCREASE, E. B. 2009. Determining DNA packaging strategy by analysis of the termini of the chromosomes in tailed-bacteriophage virions. *Methods Mol Biol,* 502**,** 91-111.

CASJENS, S. R., GILCREASE, E. B., WINN-STAPLEY, D. A., SCHICKLMAIER, P., SCHMIEGER, H., PEDULLA, M. L., FORD, M. E., HOUTZ, J. M., HATFULL, G. F. & HENDRIX, R. W. 2005. The generalized transducing *Salmonella* bacteriophage ES18: complete genome sequence and DNA packaging strategy. *J Bacteriol,* 187**,** 1091-104.

CERQUETTI, M., MOLINARI, A., SEBASTIANELLI, A., DIOCIAIUTI, M., PETRUZZELLI, R., CAPO, C. & MASTRANTONIO, P. 2000. Characterization of surface layer proteins from different *Clostridium difficile* clinical isolates. *Microbial pathogenesis,* 28**,** 363-372.

CERQUETTI, M., SERAFINO, A., SEBASTIANELLI, A. & MASTRANTONIO, P. 2002. Binding of *Clostridium difficile* to Caco-2 epithelial cell line and to extracellular matrix proteins. *FEMS Immunology & Medical Microbiology,* 32**,** 211-218.

CHAKRABORTY, S., MUKHOPADHYAY, A. K., BHADRA, R. K., GHOSH, A. N., MITRA, R., SHIMADA, T., YAMASAKI, S., FARUQUE, S. M., TAKEDA, Y., COLWELL, R. R. & NAIR, G. B. 2000. Virulence Genes in Environmental Strains of *Vibrio cholerae*. *Applied and Environmental Microbiology,* 66**,** 4022-4028.

CHEN, J., CARPENA, N., QUILES-PUCHALT, N., RAM, G., NOVICK, R. P. & PENADES, J. R. 2015. Intra- and inter-generic transfer of pathogenicity island-encoded virulence genes by cos phages. *ISME J,* 9**,** 1260-1263.

CHOI, J., KOTAY, S. M. & GOEL, R. 2010. Various physico-chemical stress factors cause prophage induction in Nitrosospira multiformis 25196-an ammonia oxidizing bacteria. *Water research,* 44**,** 4550-4558.

CLEMENTS, A. C., MAGALHAES, R. J., TATEM, A. J., PATERSON, D. L. & RILEY, T. V. 2010. *Clostridium difficile* PCR ribotype 027: assessing the risks of further worldwide spread. *Lancet Infect Dis,* 10**,** 395-404.

CLOKIE, M. R. & KROPINSKI, A. M. 2009. *Bacteriophages*, Springer.

CLOKIE, M. R. J., MILLARD, A. D., LETAROV, A. V. & HEAPHY, S. 2011. Phages in nature. *Bacteriophage,* 1**,** 31-45.

COETZEE, J. N. & DE KLERK, H. C. 1962. Lysogeny in the genus Lactobacillus. *Nature,* 194**,** 505.

COLLINS, M. D., LAWSON, P. A., WILLEMS, A., CORDOBA, J. J., FERNANDEZ-GARAYZABAL, J., GARCIA, P., CAI, J., HIPPE, H. & FARROW, J. A. 1994. The phylogeny of the genus *Clostridium*: proposal of five new genera and eleven new species combinations. *Int J Syst Bacteriol,* 44**,** 812-26.

CORNAX, R., MORIÑIGO, M. A., PAEZ, I. G., MUÑOZ, M. A. & BORREGO, J. J. 1990. Application of direct plaque assay for detection and enumeration of bacteriophages of Bacteroides fragilis from contaminated-water samples. *Applied and environmental microbiology,* 56**,** 3170-3173.

COSTA, M., REID-SMITH, R., GOW, S., HANNON, S., BOOKER, C., ROUSSEAU, J., BENEDICT, K., MORLEY, P. & WEESE, J. 2012. Prevalence and molecular characterization of *Clostridium difficile* isolated from feedlot beef cattle upon arrival and mid-feeding period. *BMC Veterinary Research,* 8**,** 1-7.

COWLES, P. B. 1934. A Bacteriophage for *Cl. tetani*. *Journal of Bacteriology,* 27**,** 163-164.

CUMBY, N., EDWARDS, A. M., DAVIDSON, A. R. & MAXWELL, K. L. 2012. The bacteriophage HK97 gp15 moron element encodes a novel superinfection exclusion protein. *J Bacteriol,* 194**,** 5012-9.

D'COSTA, V. M., MCGRANN, K. M., HUGHES, D. W. & WRIGHT, G. D. 2006. Sampling the antibiotic resistome. *Science,* 311**,** 374-7.

DARLING, A. E., MAU, B. & PERNA, N. T. 2010. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLoS ONE,* 5**,** e11147.

DARLING, A. E., WORDEN, P., CHAPMAN, T. A., CHOWDHURY, P. R., CHARLES, I. G. & DJORDJEVIC, S. P. 2014. The genome of *Clostridium difficile* 5.3. *Gut Pathogens,* 6**,** 1-4.

DAVIES, J. 1994. Inactivation of antibiotics and the dissemination of resistance genes. *Science,* 264**,** 375-82.

DAVIES, K., ASHWIN, H., LONGSHAW, C., BURNS, D., DAVIS, G. & WILCOX, M. 2015. Diversity of *Clostridium difficile* PCR-ribotypes in Europe: results from the EUCLID study. *Eurosurveillance*.

DE SORDI, L., BUTT, M. A., PYE, H., KOHOUTOVA, D., MOSSE, C. A., YAHIOGLU, G., STAMATI, I., DEONARAIN, M., BATTAH, S., READY, D., ALLAN, E., MULLANY, P. & LOVAT, L. B. 2015. Development of Photodynamic Antimicrobial Chemotherapy (PACT) for *Clostridium difficile*. *PLoS One,* 10**,** e0135039.

DEI, R. 1989. Observations on phage-typing of *Clostridium difficile* : preliminary evaluation of a phage panel. *Eur J Epidemiol,* 5**,** 351-4.

DELCHER, A. L., BRATKE, K. A., POWERS, E. C. & SALZBERG, S. L. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics,* 23**,** 673-9.

DENÈVE, C., JANOIR, C., POILANE, I., FANTINATO, C. & COLLIGNON, A. 2009. New trends in *Clostridium difficile* virulence and pathogenesis. *International Journal of Antimicrobial Agents,* 33**,** S24-S28.

DEPESTEL, D. D. & ARONOFF, D. M. 2013. Epidemiology of *Clostridium difficile* Infection. *J Pharm Pract,* 26**,** 464-75.

DIAB, S. S., UZAL, F. A. & SONGER, J. G. 2016. Diseases produced by *Clostridium difficile*. *Clostridial Diseases of Animals***,** 177.

DIDELOT, X., BOWDEN, R., WILSON, D. J., PETO, T. E. & CROOK, D. W. 2012. Transforming clinical microbiology with bacterial genome sequencing. *Nat Rev Genet,* 13**,** 601-12.

DINGLE, K. E., ELLIOTT, B., ROBINSON, E., GRIFFITHS, D., EYRE, D. W., STOESSER, N., VAUGHAN, A., GOLUBCHIK, T., FAWLEY, W. N., WILCOX, M. H., PETO, T. E., WALKER, A. S., RILEY, T. V., CROOK, D. W. & DIDELOT, X. 2014. Evolutionary history of the *Clostridium difficile* pathogenicity locus. *Genome Biol Evol,* 6**,** 36-52.

DINGLE, K. E., GRIFFITHS, D., DIDELOT, X., EVANS, J., VAUGHAN, A., KACHRIMANIDOU, M., STOESSER, N., JOLLEY, K. A., GOLUBCHIK, T. & HARDING, R. M. 2011. Clinical *Clostridium difficile*: clonality and pathogenicity locus diversity. *PLoS One,* 6**,** e19993.

DOLMAN, C. & CHANG, E. 1972. Bacteriophages of *Clostridium botulinum*. *Canadian journal of microbiology,* 18**,** 67-76.

DONTA, S. T., SULLIVAN, N. & WILKINS, T. D. 1982. Differential Effects of *Clostridium difficile* Toxins on Tissue-Cultured Cells. *Journal of Clinical Microbiology,* 15**,** 1157-1158.

DUBBERKE, E. R. & OLSEN, M. A. 2012. Burden of *Clostridium difficile* on the healthcare system. *Clin Infect Dis,* 55 Suppl 2**,** S88-92.

DUBBERKE, E. R. & WERTHEIMER, A. I. 2009. Review of current literature on the economic burden of *Clostridium difficile* infection. *Infection Control & Hospital Epidemiology,* 30**,** 57-66.

DUPUY, B., GOVIND, R., ANTUNES, A. & MATAMOUROS, S. 2008. *Clostridium difficile* toxin synthesis is negatively regulated by TcdC. *J Med Microbiol,* 57**,** 685-9.

ECKERT, C., EMIRIAN, A., LE MONNIER, A., CATHALA, L., DE MONTCLOS, H., GORET, J., BERGER, P., PETIT, A., DE CHEVIGNY, A., JEAN-PIERRE, H., NEBBAD, B., CAMIADE, S., MECKENSTOCK, R., LALANDE, V., MARCHANDIN, H. & BARBUT, F. 2015. Prevalence and pathogenicity of binary toxin–positive *Clostridium difficile* strains that do not produce toxins A and B. *New Microbes and New Infections,* 3**,** 12-17.

EDGAR, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res,* 32**,** 1792-7.

EDGAR, R. C. 2007. PILER-CR: fast and accurate identification of CRISPR repeats. *BMC Bioinformatics,* 8**,** 18.

EDWARDS, U., ROGALL, T., BLOCKER, H., EMDE, M. & BOTTGER, E. C. 1989. Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. *Nucleic Acids Res,* 17**,** 7843-53.

EKSETH, O. K., KUIPER, M. & MIRONOV, V. 2014. orthAgogue: an agile tool for the rapid prediction of orthology relations. *Bioinformatics,* 30**,** 734-6.

ESFANDIARI, Z., JALALI, M., EZZATPANAH, H., WEESE, J. S. & CHAMANI, M. 2014. Prevalence and Characterization of *Clostridium difficile* in Beef and Mutton Meats of Isfahan Region, Iran. *Jundishapur J Microbiol,* 7**,** e16771.

EVEILLARD, M., FOUREL, V., BARC, M. C., KERNEIS, S., COCONNIER, M. H., KARJALAINEN, T., BOURLIOUX, P. & SERVIN, A. L. 1993. Identification and characterization of adhesive factors of *Clostridium difficile* involved in adhesion to human colonic enterocyte-like Caco-2 and mucus-secreting HT29 cells in culture. *Mol Microbiol,* 7**,** 371-81.

EYRE, D. W., CULE, M. L., WILSON, D. J., GRIFFITHS, D., VAUGHAN, A., O'CONNOR, L., IP, C. L., GOLUBCHIK, T., BATTY, E. M., FINNEY, J. M., WYLLIE, D. H., DIDELOT, X., PIAZZA, P., BOWDEN, R., DINGLE, K. E., HARDING, R. M., CROOK, D. W., WILCOX, M. H., PETO, T. E. & WALKER, A. S. 2013. Diverse sources of *C. difficile* infection identified on whole-genome sequencing. *N Engl J Med,* 369**,** 1195-205.

EYRE, D. W., GOLUBCHIK, T., GORDON, N. C., BOWDEN, R., PIAZZA, P., BATTY, E. M., IP, C. L., WILSON, D. J., DIDELOT, X., O'CONNOR, L., LAY, R., BUCK, D., KEARNS, A. M., SHAW, A., PAUL, J., WILCOX, M. H., DONNELLY, P. J., PETO, T. E., WALKER, A. S. & CROOK, D. W. 2012. A pilot study of rapid benchtop sequencing of *Staphylococcus aureu*s and *Clostridium difficile* for outbreak detection and surveillance. *BMJ Open,* 2.

FAGAN, R. P., JANOIR, C., COLLIGNON, A., MASTRANTONIO, P., POXTON, I. R. & FAIRWEATHER, N. F. 2011. A proposed nomenclature for cell wall proteins of *Clostridium difficile*. *J Med Microbiol,* 60**,** 1225-8.

FORTIER, L.-C. & MOINEAU, S. 2007. Morphological and genetic diversity of temperate phages in *Clostridium difficile*. *Applied and environmental microbiology,* 73**,** 7358-7366.

FRAZIER, K. S., HERRON, A. J., HINES, M. E., 2ND, GASKIN, J. M. & ALTMAN, N. H. 1993. Diagnosis of enteritis and enterotoxemia due to *Clostridium difficile* in captive ostriches (Struthio camelus). *J Vet Diagn Invest,* 5**,** 623-5.

FREEMAN, J., BAINES, S. D., SAXTON, K. & WILCOX, M. H. 2007. Effect of metronidazole on growth and toxin production by epidemic *Clostridium difficile* PCR ribotypes 001 and 027 in a human gut model. *J Antimicrob Chemother,* 60**,** 83-91.

FREEMAN, J., BAUER, M. P., BAINES, S. D., CORVER, J., FAWLEY, W. N., GOORHUIS, B., KUIJPER, E. J. & WILCOX, M. H. 2010. The changing epidemiology of *Clostridium difficile* infections. *Clin Microbiol Rev,* 23**,** 529-49.

GAL, M., NORTHEY, G. & BRAZIER, J. S. 2005. A modified pulsed-field gel electrophoresis (PFGE) protocol for subtyping previously non-PFGE typeable isolates of *Clostridium difficile* polymerase chain reaction ribotype 001. *J Hosp Infect,* 61**,** 231-6.

GAULTON, T., MISRA, R., ROSE, G., BAYBAYAN, P., HALL, R., FREEMAN, J., TURTON, J., PICTON, S., KORLACH, J., GHARBIA, S. & SHAH, H. 2015. Complete Genome Sequence of the Hypervirulent Bacterium *Clostridium difficile* Strain G46, Ribotype 027. *Genome Announc,* 3.

GEBHART, D., WILLIAMS, S. R., BISHOP-LILLY, K. A., GOVONI, G. R., WILLNER, K. M., BUTANI, A., SOZHAMANNAN, S., MARTIN, D., FORTIER, L. C. & SCHOLL, D. 2012. Novel high-molecular-weight, R-type bacteriocins of *Clostridium difficile*. *J Bacteriol,* 194**,** 6240-7.

GERDING, D. N. 2004. Clindamycin, cephalosporins, fluoroquinolones, and *Clostridium difficile*-associated diarrhea: this is an antimicrobial resistance problem. *Clin Infect Dis,* 38**,** 646-8.

GERDING, D. N., JOHNSON, S., RUPNIK, M. & AKTORIES, K. 2014. *Clostridium difficile* binary toxin CDT: mechanism, epidemiology, and potential clinical importance. *Gut Microbes,* 5**,** 15-27.

GERDING, D. N., MUTO, C. A. & OWENS, R. C., JR. 2008. Measures to control and prevent *Clostridium difficile* infection. *Clin Infect Dis,* 46 Suppl 1**,** S43-9.

GERIC, B., CARMAN, R. J., RUPNIK, M., GENHEIMER, C. W., SAMBOL, S. P., LYERLY, D. M., GERDING, D. N. & JOHNSON, S. 2006. Binary toxin-producing, large clostridial toxin-negative *Clostridium difficile* strains are enterotoxic but do not cause disease in hamsters. *J Infect Dis,* 193**,** 1143-50.

GERIC, B., RUPNIK, M., GERDING, D. N., GRABNAR, M. & JOHNSON, S. 2004. Distribution of *Clostridium difficile* variant toxinotypes and strains with binary toxin genes among clinical isolates in an American hospital. *Journal of medical microbiology,* 53**,** 887-894.

GHOSE, C. 2013. *Clostridium difficile* infection in the twenty-first century. *Emerg Microbes Infect,* 2**,** e62.

GIACOMONI, P. U. 1982. Induction by mitomycin C of recA protein synthesis in bacteria and spheroplasts. *Journal of Biological Chemistry,* 257**,** 14932-14936.

GOH, S., CHANG, B. J. & RILEY, T. V. 2005a. Effect of phage infection on toxin production by *Clostridium difficile*. *J Med Microbiol,* 54**,** 129-35.

GOH, S., ONG, P. F., SONG, K. P., RILEY, T. V. & CHANG, B. J. 2007. The complete genome sequence of *Clostridium difficile* phage phiC2 and comparisons to phiCD119 and inducible prophages of CD630. *Microbiology,* 153**,** 676-85.

GOH, S., RILEY, T. V. & CHANG, B. J. 2005b. Isolation and characterization of temperate bacteriophages of *Clostridium difficile*. *Appl Environ Microbiol,* 71**,** 1079-83.

GOMEZ-TREVINO, M., BOUREAU, H., KARJALAINEN, T. & BOURLIOUX, P. 1996. *Clostridium difficile* adherence to mucus: results of an in vivo and ex vivo assay. *Microbial ecology in health and disease,* 9**,** 329-334.

GOODARZI, H., ALBOUYEH, M., AZIMI RAD, M., ZALI, M. R. & ASLAN, M. M. 2012. Molecular typing of *Clostridium difficile* isolated from hospitalized patients by PCR ribotyping. *Pejouhesh,* 36**,** 68-75.

GOORHUIS, A., BAKKER, D., CORVER, J., DEBAST, S. B., HARMANUS, C., NOTERMANS, D. W., BERGWERFF, A. A., DEKKER, F. W. & KUIJPER, E. J. 2008a. Emergence of *Clostridium difficile* infection due to a new hypervirulent strain, polymerase chain reaction ribotype 078. *Clin Infect Dis,* 47.

GOORHUIS, A., DEBAST, S. B., VAN LEENGOED, L. A., HARMANUS, C., NOTERMANS, D. W., BERGWERFF, A. A. & KUIJPER, E. J. 2008b. *Clostridium difficile* PCR ribotype 078: an emerging strain in humans and in pigs? *J Clin Microbiol,* 46**,** 1157; author reply 1158.

GOORHUIS, A., VAN DER KOOI, T., VAESSEN, N., DEKKER, F., VAN DEN BERG, R., HARMANUS, C., VAN DEN HOF, S., NOTERMANS, D. & KUIJPER, E. 2007. Spread and epidemiology of

Clostridium difficile polymerase chain reaction ribotype 027/toxinotype III in The Netherlands. *Clinical Infectious Diseases,* 45**,** 695-703.

GOUDARZI, M., GOUDARZI, H., ALEBOUYEH, M., AZIMI RAD, M., SHAYEGAN MEHR, F. S., ZALI, M. R. & ASLANI, M. M. 2013. Antimicrobial susceptibility of *Clostridium difficile* clinical isolates in iran. *Iran Red Crescent Med J,* 15**,** 704-11.

GOUDARZI, M., SEYEDJAVADI, S. S., GOUDARZI, H., MEHDIZADEH AGHDAM, E. & NAZERI, S. 2014. *Clostridium difficile* Infection: Epidemiology, Pathogenesis, Risk Factors, and Therapeutic Options. *Scientifica,* 2014**,** 9.

GOUJON, M., MCWILLIAM, H., LI, W., VALENTIN, F., SQUIZZATO, S., PAERN, J. & LOPEZ, R. 2010. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res,* 38**,** W695-9.

GOULD, L. H. & LIMBAGO, B. 2010. *Clostridium difficile* in food and domestic animals: a new foodborne pathogen? *Clin Infect Dis,* 51**,** 577-82.

GOVIND R, F. J., ROLFE R. 2011. In vivo lysogenization of a *Clostridium difficile* bacteriophage ΦCD119. *Anaerobe,* 17**,** 125-129.

GOVIND, R., FRALICK, J. A. & ROLFE, R. D. 2006. Genomic organization and molecular characterization of *Clostridium difficile* bacteriophage PhiCD119. *J Bacteriol,* 188**,** 2568-77.

GOVIND, R., VEDIYAPPAN, G., ROLFE, R. D., DUPUY, B. & FRALICK, J. A. 2009. Bacteriophage-mediated toxin gene regulation in *Clostridium difficile*. *J Virol,* 83**,** 12037-45.

GRAD, Y. H., LIPSITCH, M., FELDGARDEN, M., ARACHCHI, H. M., CERQUEIRA, G. C., FITZGERALD, M., GODFREY, P., HAAS, B. J., MURPHY, C. I., RUSS, C., SYKES, S., WALKER, B. J., WORTMAN, J. R., YOUNG, S., ZENG, Q., ABOUELLEIL, A., BOCHICCHIO, J., CHAUVIN, S., DESMET, T., GUJJA, S., MCCOWAN, C., MONTMAYEUR, A., STEELMAN, S., FRIMODT-MOLLER, J., PETERSEN, A. M., STRUVE, C., KROGFELT, K. A., BINGEN, E., WEILL, F. X., LANDER, E. S., NUSBAUM, C., BIRREN, B. W., HUNG, D. T. & HANAGE, W. P. 2012. Genomic epidemiology of the *Escherichia coli* O104:H4 outbreaks in Europe, 2011. *Proc Natl Acad Sci U S A,* 109**,** 3065-70.

GRIFFITHS, D., FAWLEY, W., KACHRIMANIDOU, M., BOWDEN, R., CROOK, D. W., FUNG, R., GOLUBCHIK, T., HARDING, R. M., JEFFERY, K. J., JOLLEY, K. A., KIRTON, R., PETO, T. E., REES, G., STOESSER, N., VAUGHAN, A., WALKER, A. S., YOUNG, B. C., WILCOX, M. & DINGLE, K. E. 2010. Multilocus sequence typing of *Clostridium difficile*. *J Clin Microbiol,* 48**,** 770-8.

GRISSA, I., VERGNAUD, G. & POURCEL, C. 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res,* 35**,** W52-7.

GUPTA, A. & KHANNA, S. 2014. Community-acquired *Clostridium difficile* infection: an increasing public health threat. *Infection and Drug Resistance,* 7**,** 63-72.

GURTLER, V. 1993. Typing of *Clostridium difficile* strains by PCR-amplification of variable length 16S-23S rDNA spacer regions. *J Gen Microbiol,* 139**,** 3089-97.

HAFIZ, S. 1974. *Clostridium difficile and its toxins.* (thesis, Ph.D.) (thesis, Ph.D.), University of Leeds

HAFIZ, S. & OAKLEY, C. L. 1976. *Clostridium difficile* : isolation and characteristics. *J Med Microbiol,* 9**,** 129-36.

HALL, I. C. & 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.

HAMMOND, G. A. & JOHNSON, J. L. 1995. The toxigenic element of *Clostridium difficile* strain VPI 10463. *Microb Pathog,* 19**,** 203-13.

HANLON, G. W. 2007. Bacteriophages: an appraisal of their role in the treatment of bacterial infections. *International journal of antimicrobial agents,* 30**,** 118-128.

HARGREAVES, K. R. 2012. *ISOLATION AND CHARACTERISATION OF BACTERIOPHAGES INFECTING ENVIRONMENTAL STRAINS OF CLOSTRIDIUM DIFFICILE.* Doctor of Philosophy, Leicester.

HARGREAVES, K. R. & CLOKIE, M. R. J. 2014. *Clostridium difficile* phages: still difficult? *Frontiers in Microbiology,* 5.

HARGREAVES, K. R. & CLOKIE, M. R. J. 2015. A Taxonomic Review of *Clostridium difficile* Phages and Proposal of a Novel Genus, "Phimmp04likevirus". *Viruses,* 7**,** 2534-2541.

HARGREAVES, K. R., COLVIN, H. V., PATEL, K. V., CLOKIE, J. J. & CLOKIE, M. R. 2013. Genetically diverse *Clostridium difficile* strains harboring abundant prophages in an estuarine environment. *Appl Environ Microbiol,* 79**,** 6236-43.

HARGREAVES, K. R., FLORES, C. O., LAWLEY, T. D. & CLOKIE, M. R. 2014a. 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.

HARGREAVES, K. R., KROPINSKI, A. M. & CLOKIE, M. R. J. 2014b. What Does the Talking?: Quorum Sensing Signalling Genes Discovered in a Bacteriophage Genome. *PLoS ONE,* 9**,** e85131.

HARGREAVES, K. R., OTIENO, J. R., THANKI, A., BLADES, M. J., MILLARD, A. D., BROWNE, H. P., LAWLEY, T. D. & CLOKIE, M. R. 2015. As Clear as Mud? Determining the Diversity and Prevalence of Prophages in the Draft Genomes of Estuarine Isolates of *Clostridium difficile*. *Genome Biol Evol,* 7**,** 1842-55.

HARVEY, R. B., NORMAN, K. N., ANDREWS, K., NORBY, B., HUME, M. E., SCANLAN, C. M., HARDIN, M. D. & SCOTT, H. M. 2011. *Clostridium difficile* in retail meat and processing plants in Texas. *J Vet Diagn Invest,* 23**,** 807-11.

HATFULL, G. F. 2008. Bacteriophage genomics. *Curr Opin Microbiol,* 11**,** 447-53.

HAWKINS, C. C., BUGGY, B. P., FEKETY, R. & SCHABERG, D. R. 1984. Epidemiology of colitis induced by *Clostridium difficile* in hamsters: application of a bacteriophage and bacteriocin typing system. *J Infect Dis,* 149**,** 775-80.

HE, M. 2012. *Genomic variation and evolution of Clostridium difficile.* Doctor of Philosophy, University of Cambridge.

HE, M., MIYAJIMA, F., ROBERTS, P., ELLISON, L., PICKARD, D. J., MARTIN, M. J., CONNOR, T. R., HARRIS, S. R., FAIRLEY, D. & BAMFORD, K. B. 2013. Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*. *Nature genetics,* 45**,** 109-113.

HE, M., SEBAIHIA, M., LAWLEY, T. D., STABLER, R. A., DAWSON, L. F., MARTIN, M. J., HOLT, K. E., SETH-SMITH, H. M., QUAIL, M. A., RANCE, R., BROOKS, K., CHURCHER, C., HARRIS, D., BENTLEY, S. D., BURROWS, C., CLARK, L., CORTON, C., MURRAY, V., ROSE, G., THURSTON, S., VAN TONDER, A., WALKER, D., WREN, B. W., DOUGAN, G. & PARKHILL, J. 2010. Evolutionary dynamics of *Clostridium difficile* over short and long time scales. *Proc Natl Acad Sci U S A,* 107**,** 7527-32.

HEALTH PROTECTION AGENCY, H. 2012. *Clostridium difficile*: updated guidance on diagnosis and reporting. *Department of Health*.

HEALTH PROTECTION AGENCY, H. 2015. Health protection report. 9.

HEGARTY, J. P., SANGSTER, W., ASHLEY, R. E., MYERS, R., HAFENSTEIN, S. & STEWART, D. B. 2016. Induction and Purification of *C. difficile* Phage Tail-Like Particles. *In:* ROBERTS, P. A. & MULLANY, P. (eds.) *Clostridium difficile: Methods and Protocols.* New York, NY: Springer New York.

HENDRIX, R. W. 2002. Bacteriophages: evolution of the majority. *Theor Popul Biol,* 61**,** 471-80.

HENDRIX, R. W. 2003. Bacteriophage genomics. *Current Opinion in Microbiology,* 6**,** 506-511.

HENDRIX, R. W., LAWRENCE, J. G., HATFULL, G. F. & CASJENS, S. 2000. The origins and ongoing evolution of viruses. *Trends in microbiology,* 8**,** 504-508.

HENSGENS, M. P. M., KEESSEN, E. C., SQUIRE, M. M., RILEY, T. V., KOENE, M. G. J., DE BOER, E., LIPMAN, L. J. A. & KUIJPER, E. J. 2012. *Clostridium difficile* infection in the community: a zoonotic disease? *Clinical Microbiology and Infection,* 18**,** 635-645.

HIGAZI, T. B., MOHANNAD, A., BURKETT, M. & PUSOK, R. 2011. PCR Detection of *Clostridium difficile* and its Toxigenic Strains in Public Places in Southeast Ohio. *Intl. J,* 2**,** 105-111.

HONDA, H. & DUBBERKE, E. R. 2014. The changing epidemiology of *Clostridium difficile* infection. *Curr Opin Gastroenterol,* 30**,** 54-62.

HORGAN, M., O'SULLIVAN, O., COFFEY, A., FITZGERALD, G. F., VAN SINDEREN, D., MCAULIFFE, O. & ROSS, R. P. 2010. Genome analysis of the *Clostridium difficile* phage ΦCD6356, a temperate phage of the Siphoviridae family. *Gene,* 462**,** 34-43.

HORVATH, P., COUTE-MONVOISIN, A. C., ROMERO, D. A., BOYAVAL, P., FREMAUX, C. & BARRANGOU, R. 2009. Comparative analysis of CRISPR loci in lactic acid bacteria genomes. *Int J Food Microbiol,* 131**,** 62-70.

HUBER, C. A., FOSTER, N. F., RILEY, T. V. & PATERSON, D. L. 2013. Challenges for Standardization of *Clostridium difficile* Typing Methods. *Journal of Clinical Microbiology,* 51**,** 2810-2814.

HUBERT, B., LOO, V. G., BOURGAULT, A. M., POIRIER, L., DASCAL, A., FORTIN, E., DIONNE, M. & LORANGE, M. 2007. A portrait of the geographic dissemination of the *Clostridium difficile* North American pulsed-field type 1 strain and the epidemiology of *C. difficile*-associated disease in Quebec. *Clin Infect Dis,* 44**,** 238-44.

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.

HUSON, D. H. & BRYANT, D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol,* 23**,** 254-67.

HYATT, D., CHEN, G. L., LOCASCIO, P. F., LAND, M. L., LARIMER, F. W. & HAUSER, L. J. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics,* 11**,** 119.

HYMAN, P. & ABEDON, S. T. 2010. Bacteriophage host range and bacterial resistance. *Advances in applied microbiology,* 70**,** 217-248.

INDRA, A., HUHULESCU, S., SCHNEEWEIS, M., HASENBERGER, P., KERNBICHLER, S., FIEDLER, A., WEWALKA, G., ALLERBERGER, F. & KUIJPER, E. J. 2008. Characterization of *Clostridium difficile* isolates using capillary gel electrophoresis-based PCR ribotyping. *J Med Microbiol,* 57.

IVARSSON, M. E. 2014. *Toxin Inhibitors for the Treatment of Clostridium difficile Infection.* Diss., Eidgenössische Technische Hochschule ETH Zürich, Nr. 22210.

IVARSSON, M. E., LEROUX, J.-C. & CASTAGNER, B. 2015. Investigational new treatments for *Clostridium difficile* infection. *Drug Discovery Today,* 20**,** 602-608.

JALALI, M., KHORVASH, F., WARRINER, K. & WEESE, J. S. 2012. *Clostridium difficile* infection in an Iranian hospital. *BMC Research Notes,* 5**,** 1-5.

JANEZIC, S., OCEPEK, M., ZIDARIC, V. & RUPNIK, M. 2012. *Clostridium difficile* genotypes other than ribotype 078 that are prevalent among human, animal and environmental isolates. *BMC Microbiol,* 12.

JANEZIC, S. & RUPNIK, M. 2010. Molecular typing methods for *Clostridium difficile*: pulsed-field gel electrophoresis and PCR ribotyping. *Methods Mol Biol,* 646.

JANEZIC, S. & RUPNIK, M. 2015. Genomic diversity of *Clostridium difficile* strains. *Res Microbiol,* 166**,** 353-60.

JANEZIC, S., ZIDARIC, V., PARDON, B., INDRA, A., KOKOTOVIC, B., BLANCO, J. L., SEYBOLDT, C., DIAZ, C. R., POXTON, I. R., PERRETEN, V., DRIGO, I., JIRASKOVA, A., OCEPEK, M., WEESE, J. S., SONGER, J. G., WILCOX, M. H. & RUPNIK, M. 2014. International *Clostridium difficile* animal strain collection and large diversity of animal associated strains. *BMC Microbiology,* 14**,** 1-10.

JANOIR, C. 2016. Virulence factors of *Clostridium difficile* and their role during infection. *Anaerobe,* 37**,** 13-24.

JANOIR, C., PÉCHINÉ, S., GROSDIDIER, C. & COLLIGNON, A. 2007. Cwp84, a surface-associated protein of *Clostridium difficile*, is a cysteine protease with degrading activity on extracellular matrix proteins. *Journal of bacteriology,* 189**,** 7174-7180.

JIANG, S. C. & PAUL, J. H. 1998. Significance of Lysogeny in the Marine Environment: Studies with Isolates and a Model of Lysogenic Phage Production. *Microb Ecol,* 35**,** 235-43.

JOHNSON, S. 2014. Editorial commentary: changing epidemiology of *Clostridium difficile* and emergence of new virulent strains. *Clinical Infectious Diseases,* 58**,** 1731-1733.

JOHNSON , S., SAMORE , M. H., FARROW , K. A., KILLGORE , G. E., TENOVER , F. C., LYRAS , D., ROOD , J. I., DEGIROLAMI , P., BALTCH , A. L., RAFFERTY , M. E., PEAR , S. M. & GERDING , D. N. 1999. Epidemics of Diarrhea Caused by a Clindamycin-Resistant Strain of *Clostridium difficile* in Four Hospitals. *New England Journal of Medicine,* 341**,** 1645-1651.

JONES, P., BINNS, D., CHANG, H. Y., FRASER, M., LI, W., MCANULLA, C., MCWILLIAM, H., MASLEN, J., MITCHELL, A., NUKA, G., PESSEAT, S., QUINN, A. F., SANGRADOR-VEGAS, A., SCHEREMETJEW, M., YONG, S. Y., LOPEZ, R. & HUNTER, S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics,* 30**,** 1236-40.

JOSE, S. & MADAN, R. 2016. Neutrophil-mediated inflammation in the pathogenesis of *Clostridium difficile* infections. *Anaerobe*.

JOSHI NA, F. J. 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files

https://github.com/najoshi/sickle.

.

JUHALA, R. J., FORD, M. E., DUDA, R. L., YOULTON, A., HATFULL, G. F. & HENDRIX, R. W. 2000. Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages1. *Journal of Molecular Biology,* 299**,** 27-51.

JUMP, R. L. P., PULTZ, M. J. & DONSKEY, C. J. 2007. Vegetative *Clostridium difficile* Survives in Room Air on Moist Surfaces and in Gastric Contents with Reduced Acidity: a Potential Mechanism To Explain the Association between Proton Pump Inhibitors and *C. difficile*-Associated Diarrhea? *Antimicrobial Agents and Chemotherapy,* 51**,** 2883-2887.

JUST, I. & GERHARD, R. 2004. Large clostridial cytotoxins. *Rev Physiol Biochem Pharmacol,* 152**,** 23-47.

JUST, I., SELZER, J., WILM, M., VON EICHEL-STREIBER, C., MANN, M. & AKTORIES, K. 1995. Glucosylation of Rho proteins by *Clostridium difficile* toxin B. *Nature,* 375**,** 500-3.

KALCHAYANAND, N., ARTHUR, T. M., BOSILEVAC, J. M., BRICHTA-HARHAY, D. M., SHACKELFORD, S. D., WELLS, J. E., WHEELER, T. L. & KOOHMARAIE, M. 2013. Isolation and characterization of *Clostridium difficile* associated with beef cattle and commercially produced ground beef. *J Food Prot,* 76**,** 256-64.

KALIA, V. C., MUKHERJEE, T., BHUSHAN, A., JOSHI, J., SHANKAR, P. & HUMA, N. 2011. Analysis of the unexplored features of rrs (16S rDNA) of the Genus *Clostridium*. *BMC Genomics,* 12**,** 1-35.

KAMIYA, S., OGURA, H., MENG, X. Q. & NAKAMURA, S. 1992. Correlation between cytotoxin production and sporulation in *Clostridium difficile*. *Journal of Medical Microbiology,* 37**,** 206.

KARJALAINEN, T., BARC, M. C., COLLIGNON, A., TROLLÉ, S., BOUREAU, H., COTTE-LAFFITTE, J. & BOURLIOUX, P. 1994. Cloning of a genetic determinant from *Clostridium difficile* involved in adherence to tissue culture cells and mucus. *Infection and Immunity,* 62**,** 4347-4355.

KARJALAINEN, T., SAUMIER, N., BARC, M. C., DELMEE, M. & COLLIGNON, A. 2002. *Clostridium difficile* genotyping based on slpA variable region in S-layer gene sequence: an alternative to serotyping. *J Clin Microbiol,* 40**,** 2452-8.

KARLSSON, S., DUPUY, B., MUKHERJEE, K., NORIN, E., BURMAN, L. G. & AKERLUND, T. 2003. Expression of *Clostridium difficile* toxins A and B and their sigma factor TcdD is controlled by temperature. *Infect Immun,* 71**,** 1784-93.

KATO, H., KATO, N., WATANABE, K., IWAI, N., NAKAMURA, H., YAMAMOTO, T., SUZUKI, K., KIM, S., CHONG, Y. & WASITO, E. 1998. Identification of toxin A-negative, toxin B-positive *Clostridium difficile* by PCR. *J Clin Microbiol,* 36.

KATO, H., KATO, N., WATANABE, K., UENO, K., USHIJIMA, H., HASHIRA, S. & ABE, T. 1994. Application of typing by pulsed-field gel electrophoresis to the study of *Clostridium difficile* in a neonatal intensive care unit. *J Clin Microbiol,* 32**,** 2067-70.

KATO, N., OU, C. Y., KATO, H., BARTLEY, S. L., BROWN, V. K., DOWELL, V. R. & UENO, K. 1991. Identification of toxigenic *Clostridium difficile* by the polymerase chain reaction. *Journal of Clinical Microbiology,* 29**,** 33-37.

KATOH, K., MISAWA, K., KUMA, K. & MIYATA, T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res,* 30**,** 3059-66.

KATSURA, I. 1987. Determination of bacteriophage lambda tail length by a protein ruler. *Nature,* 327**,** 73-5.

KEARSE, M., MOIR, R., WILSON, A., STONES-HAVAS, S., CHEUNG, M., STURROCK, S., BUXTON, S., COOPER, A., MARKOWITZ, S., DURAN, C., THIERER, T., ASHTON, B., MEINTJES, P. & DRUMMOND, A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics,* 28**,** 1647-9.

KEEL, K., BRAZIER, J. S., POST, K. W., WEESE, S. & SONGER, J. G. 2007. Prevalence of PCR ribotypes among *Clostridium difficile* isolates from pigs, calves, and other species. *J Clin Microbiol,* 45**,** 1963-4.

KEIM, P., PRICE, L. B., KLEVYTSKA, A. M., SMITH, K. L., SCHUPP, J. M., OKINAKA, R., JACKSON, P. J. & HUGH-JONES, M. E. 2000. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. *J Bacteriol,* 182**,** 2928-36.

KELLY, C. P., POTHOULAKIS, C. & LAMONT, J. T. 1994. *Clostridium difficile* colitis. *N Engl J Med,* 330**,** 257-62.

KHANNA, S. & PARDI, D. S. 2012. IBD: Poor outcomes after *Clostridium difficile* infection in IBD. *Nat Rev Gastroenterol Hepatol,* 9**,** 307-8.

KHOSHDEL, A., HABIBIAN, R., PARVIN, N., DOOSTI, A., FAMOURI, F., ESHRAGHI, A. & HAFIZI, M. 2015. Molecular characterization of nosocomial *Clostridium difficile* infection in pediatric ward in Iran. *SpringerPlus,* 4**,** 1-6.

KILLGORE, G. E. & KATO, H. 1994. Use of arbitrary primer PCR to type *Clostridium difficile* and comparison of results with those by immunoblot typing. *J Clin Microbiol,* 32**,** 1591-3.

KIM, K. H., FEKETY, R., BATTS, D. H., BROWN, D., CUDMORE, M., SILVA, J., JR. & WATERS, D. 1981. Isolation of *Clostridium difficile* from the environment and contacts of patients with antibiotic-associated colitis. *J Infect Dis,* 143**,** 42-50.

KIM, Y. K. & MCCARTER, L. L. 2000. Analysis of the polar flagellar gene system of Vibrio parahaemolyticus. *J Bacteriol,* 182**,** 3693-704.

KLINGLER, P. J., METZGER, P. P., SEELIG, M. H., PETTIT, P. D., KNUDSEN, J. M. & ALVAREZ, S. A. 2000. *Clostridium difficile* infection: risk factors, medical and surgical management. *Dig Dis,* 18**,** 147-60.

KNETSCH, C., CONNOR, T. R., MUTREJA, A., VAN DORP, S., SANDERS, I., BROWNE, H., HARRIS, D., LIPMAN, L., KEESSEN, E. & CORVER, J. 2014. Whole genome sequencing reveals potential spread of *Clostridium difficile* between humans and farm animals in the Netherlands, 2002 to 2011. *Eurosurveillance,* 19**,** 30-41.

KNETSCH, C. W., LAWLEY, T. D., HENSGENS, M. P., CORVER, J., WILCOX, M. W. & KUIJPER, E. J. 2013. Current application and future perspectives of molecular typing methods to study *Clostridium difficile* infections. *Euro Surveill,* 18**,** 20381.

KNIGHT, D. R., ELLIOTT, B., CHANG, B. J., PERKINS, T. T. & RILEY, T. V. 2015. Diversity and Evolution in the Genome of *Clostridium difficile*. *Clin Microbiol Rev,* 28**,** 721-741.

KOENE, M. G., MEVIUS, D., WAGENAAR, J. A., HARMANUS, C., HENSGENS, M. P., MEETSMA, A. M., PUTIRULAN, F. F., VAN BERGEN, M. A. & KUIJPER, E. J. 2012. *Clostridium difficile* in Dutch animals: their presence, characteristics and similarities with human isolates. *Clin Microbiol Infect,* 18**,** 778-84.

KOSER, C. U., HOLDEN, M. T., ELLINGTON, M. J., CARTWRIGHT, E. J., BROWN, N. M., OGILVY-STUART, A. L., HSU, L. Y., CHEWAPREECHA, C., CROUCHER, N. J., HARRIS, S. R., SANDERS, M., ENRIGHT, M. C., DOUGAN, G., BENTLEY, S. D., PARKHILL, J., FRASER, L. J., BETLEY, J. R., SCHULZ-TRIEGLAFF, O. B., SMITH, G. P. & PEACOCK, S. J. 2012. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *N Engl J Med,* 366**,** 2267-75.

KOSKELLA, B. & MEADEN, S. 2013. Understanding Bacteriophage Specificity in Natural Microbial Communities. *Viruses,* 5**,** 806-823.

KOTILA, S. M., PITKANEN, T., BRAZIER, J., EEROLA, E., JALAVA, J., KUUSI, M., KONONEN, E., LAINE, J., MIETTINEN, I. T., VUENTO, R. & VIROLAINEN, A. 2013. *Clostridium difficile* contamination of public tap water distribution system during a waterborne outbreak in Finland. *Scand J Public Health,* 41**,** 541-5.

KREIMEYER, I., EULER, F., MARCKSCHEFFEL, A., TATGE, H., PICH, A., OLLING, A., SCHWARZ, J., JUST, I. & GERHARD, R. 2011. Autoproteolytic cleavage mediates cytotoxicity of *Clostridium difficile* toxin A. *Naunyn Schmiedebergs Arch Pharmacol,* 383**,** 253-62.

KRISHNA, M. M., POWELL, N. B. & BORRIELLO, S. P. 1996. Cell surface properties of *Clostridium difficile* : haemagglutination, relative hydrophobicity and charge. *J Med Microbiol,* 44**,** 115-23.

KROPINSKI, A. M., SULAKVELIDZE, A., KONCZY, P. & POPPE, C. 2007. *Salmonella* phages and prophages--genomics and practical aspects. *Methods Mol Biol,* 394**,** 133-75.

KRUPOVIC, M. & BAMFORD, D. H. 2008. Holin of bacteriophage lambda: structural insights into a membrane lesion. *Mol Microbiol,* 69**,** 781-3.

KUEHNE, S. A., COLLERY, M. M., KELLY, M. L., CARTMAN, S. T., COCKAYNE, A. & MINTON, N. P. 2014. Importance of toxin A, toxin B, and CDT in virulence of an epidemic *Clostridium difficile* strain. *J Infect Dis,* 209**,** 83-6.

KUIJPER, E., BARBUT, F., BRAZIER, J., KLEINKAUF, N., ECKMANNS, T., LAMBERT, M., DRUDY, D., FITZPATRICK, F., WIUFF, C. & BROWN, D. 2008. Update of *Clostridium difficile* infection due to PCR ribotype 027 in Europe, 2008. *Euro surveillance: bulletin européen sur les maladies transmissibles= European communicable disease bulletin*.

KUIJPER, E., COIGNARD, B. & TÜLL, P. 2006. Emergence of *Clostridium difficile*-associated disease in North America and Europe. *Clinical Microbiology and Infection,* 12**,** 2-18.

KUIJPER, E. J., DE WEERDT, J., KATO, H., KATO, N., VAN DAM, A. P., VAN DER VORM, E. R., WEEL, J., VAN RHEENEN, C. & DANKERT, J. 2001. Nosocomial outbreak of *Clostridium difficile*-associated diarrhoea due to a clindamycin-resistant enterotoxin A-negative strain. *Eur J Clin Microbiol Infect Dis,* 20**,** 528-34.

KUIJPER, E. J., VAN DEN BERG, R. J. & BRAZIER, J. S. 2009. Comparison of molecular typing methods applied to *Clostridium difficile*. *Methods Mol Biol,* 551**,** 159-71.

KUMAR, S., STECHER, G. & TAMURA, K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol*.

KURTI, Z., LOVASZ, B. D., MANDEL, M. D., CSIMA, Z., GOLOVICS, P. A., CSAKO, B. D., MOHAS, A., GÖNCZI, L., GECSE, K. B., KISS, L. S., SZATHMARI, M. & LAKATOS, P. L. 2015. Burden of *Clostridium difficile* infection between 2010 and 2013: Trends and outcomes from an

academic center in Eastern Europe. *World Journal of Gastroenterology : WJG,* 21**,** 6728-6735.

KUTTER, E. 2009. Phage host range and efficiency of plating. *Methods Mol Biol,* 501**,** 141-9.

KUTTER, E. & SULAKVELIDZE, A. 2004. *Bacteriophages: biology and applications*, CRC Press.

LABRIE, S. J., SAMSON, J. E. & MOINEAU, S. 2010. Bacteriophage resistance mechanisms. *Nat Rev Microbiol,* 8**,** 317-27.

LANE, D. J., PACE, B., OLSEN, G. J., STAHL, D. A., SOGIN, M. L. & PACE, N. R. 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc Natl Acad Sci U S A,* 82**,** 6955-9.

LARKIN, M. A., BLACKSHIELDS, G., BROWN, N. P., CHENNA, R., MCGETTIGAN, P. A., MCWILLIAM, H., VALENTIN, F., WALLACE, I. M., WILM, A. & LOPEZ, R. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics,* 23.

LAVIGNE, R., DARIUS, P., SUMMER, E. J., SETO, D., MAHADEVAN, P., NILSSON, A. S., ACKERMANN, H. W. & KROPINSKI, A. M. 2009. Classification of Myoviridae bacteriophages using protein sequence similarity. *BMC microbiology,* 9**,** 224.

LAWLEY, T. D., CROUCHER, N. J., YU, L., CLARE, S., SEBAIHIA, M., GOULDING, D., PICKARD, D. J., PARKHILL, J., CHOUDHARY, J. & DOUGAN, G. 2009. Proteomic and genomic characterization of highly infectious *Clostridium difficile* 630 spores. *J Bacteriol,* 191**,** 5377-86.

LEMEE, L., DHALLUIN, A., PESTEL-CARON, M., LEMELAND, J. F. & PONS, J. L. 2004. Multilocus sequence typing analysis of human and animal *Clostridium difficile* isolates of various toxigenic types. *J Clin Microbiol,* 42**,** 2609-17.

LESSA, F. C., GOULD, C. V. & MCDONALD, L. C. 2012. Current status of *Clostridium difficile* infection epidemiology. *Clin Infect Dis,* 55 Suppl 2**,** S65-70.

LESSA , F. C., MU , Y., BAMBERG , W. M., BELDAVS , Z. G., DUMYATI , G. K., DUNN , J. R., FARLEY , M. M., HOLZBAUER , S. M., MEEK , J. I., PHIPPS , E. C., WILSON , L. E., WINSTON , L. G., COHEN , J. A., LIMBAGO , B. M., FRIDKIN , S. K., GERDING , D. N. & MCDONALD , L. C. 2015. Burden of *Clostridium difficile* Infection in the United States. *New England Journal of Medicine,* 372**,** 825-834.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G. & DURBIN, R. 2009. The Sequence Alignment Map format and SAMtools. *Bioinformatics,* 25**,** 2078-9.

LI, R., LI, Y., KRISTIANSEN, K. & WANG, J. 2008. SOAP: short oligonucleotide alignment program. *Bioinformatics,* 24**,** 713-4.

LI, R., ZHU, H., RUAN, J., QIAN, W., FANG, X., SHI, Z., LI, Y., LI, S., SHAN, G., KRISTIANSEN, K., LI, S., YANG, H., WANG, J. & WANG, J. 2010. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res,* 20**,** 265-72.

LILLEHAUG, D. 1997. An improved plaque assay for poor plaque-producing temperate *lactococcal* bacteriophages. *Journal of Applied Microbiology,* 83**,** 85-90.

LOO, V. G., POIRIER, L., MILLER, M. A., OUGHTON, M., LIBMAN, M. D., MICHAUD, S., BOURGAULT, A.-M., NGUYEN, T., FRENETTE, C. & KELLY, M. 2005. A predominantly clonal multi-institutional outbreak of *Clostridium difficile*–associated diarrhea with high morbidity and mortality. *New England Journal of Medicine,* 353**,** 2442-2449.

LUND, B. M. & PECK, M. W. 2015. A Possible Route for Foodborne Transmission of *Clostridium difficile*? *Foodborne Pathogens and Disease,* 12**,** 177-182.

LYDERSEN, B. K. & POLLARD, E. C. 1975. The induction of lambda-prophage by ionizing radiation. *Radiat Res,* 64**,** 237-44.

LYERLY, D. M., KRIVAN, H. C. & WILKINS, T. D. 1988. *Clostridium difficile* : its disease and toxins. *Clinical Microbiology Reviews,* 1**,** 1-18.

LYNCH, M., TAUXE, R. & HEDBERG, C. 2009. The growing burden of foodborne outbreaks due to contaminated fresh produce: risks and opportunities. *Epidemiology and infection,* 137**,** 307-315.

LYRAS, D., O'CONNOR, J. R., HOWARTH, P. M., SAMBOL, S. P., CARTER, G. P., PHUMOONNA, T., POON, R., ADAMS, V., VEDANTAM, G., JOHNSON, S., GERDING, D. N. & ROOD, J. I. 2009. Toxin B is essential for virulence of *Clostridium difficile*. *Nature,* 458**,** 1176-9.

MAGILL, S. S., EDWARDS, J. R., BAMBERG, W., BELDAVS, Z. G., DUMYATI, G., KAINER, M. A., LYNFIELD, R., MALONEY, M., MCALLISTER-HOLLOD, L. & NADLE, J. 2014. Multistate point-prevalence survey of health care–associated infections. *New England Journal of Medicine,* 370**,** 1198-1208.

MAHONY, D. E., BELL, P. D. & EASTERBROOK, K. B. 1985. Two bacteriophages of *Clostridium difficile*. *J Clin Microbiol,* 21**,** 251-4.

MAIDEN, M. C. 2006. Multilocus sequence typing of bacteria. *Annu Rev Microbiol,* 60**,** 561-88.

MAIDEN, M. C., BYGRAVES, J. A., FEIL, E., MORELLI, G., RUSSELL, J. E., URWIN, R., ZHANG, Q., ZHOU, J., ZURTH, K., CAUGANT, D. A., FEAVERS, I. M., ACHTMAN, M. & SPRATT, B. G. 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.

MALHAM, S. K., RAJKO-NENOW, P., HOWLETT, E., TUSON, K. E., PERKINS, T. L., PALLETT, D. W., WANG, H., JAGO, C. F., JONES, D. L. & MCDONALD, J. E. 2014. The interaction of human microbial pathogens, particulate material and nutrients in estuarine environments and their impacts on recreational and shellfish waters. *Environmental Science: Processes & Impacts,* 16**,** 2145-2155.

MANI, N., LYRAS, D., BARROSO, L., HOWARTH, P., WILKINS, T., ROOD, J. I., SONENSHEIN, A. L. & DUPUY, B. 2002. Environmental response and autoregulation of *Clostridium difficile* TxeR, a sigma factor for toxin gene expression. *J Bacteriol,* 184**,** 5971-8.

MARCHLER-BAUER, A. & BRYANT, S. H. 2004. CD-Search: protein domain annotations on the fly. *Nucleic Acids Res,* 32**,** W327-31.

MARRA, F. & NG, K. 2015. Controversies Around Epidemiology, Diagnosis and Treatment of *Clostridium difficile* Infection. *Drugs,* 75**,** 1095-118.

MARSH, J. W., O'LEARY, M. M., SHUTT, K. A., PASCULLE, A. W., JOHNSON, S., GERDING, D. N., MUTO, C. A. & HARRISON, L. H. 2006. Multilocus variable-number tandem-repeat analysis for investigation of *Clostridium difficile* transmission in Hospitals. *J Clin Microbiol,* 44**,** 2558-66.

MARSH, J. W., O'LEARY, M. M., SHUTT, K. A., SAMBOL, S. P., JOHNSON, S., GERDING, D. N. & HARRISON, L. H. 2010. Multilocus variable-number tandem-repeat analysis and multilocus sequence typing reveal genetic relationships among *Clostridium difficile* isolates genotyped by restriction endonuclease analysis. *Journal of clinical microbiology,* 48**,** 412-418.

MARTIN, J. S. H., MONAGHAN, T. M. & WILCOX, M. H. 2016. *Clostridium difficile* infection: epidemiology, diagnosis and understanding transmission. *Nature Reviews Gastroenterology & Hepatology,* 13**,** 206-216.

MARTIROSIAN, G., SZCZESNY, A. & SILVA, J., JR. 2005. *Clostridium difficile* in emergency room. *Anaerobe,* 11**,** 258-61.

MATAMOUROS, S., ENGLAND, P. & DUPUY, B. 2007. *Clostridium difficile* toxin expression is inhibited by the novel regulator TcdC. *Mol Microbiol,* 64**,** 1274-88.

MAYER, M. J., GAREFALAKI, V., SPOERL, R., NARBAD, A. & MEIJERS, R. 2011. Structure-Based Modification of a *Clostridium difficile*-Targeting Endolysin Affects Activity and Host Range. *Journal of Bacteriology,* 193**,** 5477-5486.

MAYER, M. J., NARBAD, A. & GASSON, M. J. 2008. Molecular characterization of a *Clostridium difficile* bacteriophage and its cloned biologically active endolysin. *J Bacteriol,* 190**,** 6734-40.

MCCOUBREY, J. & POXTON, I. R. 2001. Variation in the surface layer proteins of *Clostridium difficile*. *FEMS Immunology & Medical Microbiology,* 31**,** 131-135.

MCDONALD, E. G., MILLIGAN, J., FRENETTE, C. & LEE, T. C. 2015. Continuous Proton Pump Inhibitor Therapy and the Associated Risk of Recurrent *Clostridium difficile* Infection. *JAMA Intern Med,* 175**,** 784-91.

MCDONALD, L. C., KILLGORE, G. E., THOMPSON, A., OWENS JR, R. C., KAZAKOVA, S. V., SAMBOL, S. P., JOHNSON, S. & GERDING, D. N. 2005. An epidemic, toxin gene–variant strain of *Clostridium difficile*. *New England Journal of Medicine,* 353**,** 2433-2441.

MCDONALD, L. C., OWINGS, M. & JERNIGAN, D. B. 2006. *Clostridium difficile* infection in patients discharged from US short-stay hospitals, 1996–2003. *Emerg Infect Dis,* 12**,** 409-415.

MCFARLAND, L. V., SURAWICZ, C. M., RUBIN, M., FEKETY, R., ELMER, G. W. & GREENBERG, R. N. 1999. Recurrent *Clostridium difficile* Disease: Epidemiology and Clinical Characteristics. *Infection Control and Hospital Epidemiology,* 20**,** 43-50.

MCMULLEN, K. M., MAYFIELD, J. L. & DUBBERKE, E. R. 2006. Blue Ribbon Abstract Award. *American Journal of Infection Control,* 34**,** E155-E156.

MEADER, E., MAYER, M. J., GASSON, M. J., STEVERDING, D., CARDING, S. R. & NARBAD, A. 2010. Bacteriophage treatment significantly reduces viable *Clostridium difficile* and prevents toxin production in an in vitro model system. *Anaerobe,* 16**,** 549-54.

MEESSEN-PINARD, M., SEKULOVIC, O. & FORTIER, L.-C. 2012. Evidence of in vivo prophage induction during *Clostridium difficile* infection. *Applied and environmental microbiology,* 78**,** 7662-7670.

MENDE, D. R., SUNAGAWA, S., ZELLER, G. & BORK, P. 2013. Accurate and universal delineation of prokaryotic species. *Nat Meth,* 10**,** 881-884.

MERRIGAN, M., VENUGOPAL, A., MALLOZZI, M., ROXAS, B., VISWANATHAN, V. K., JOHNSON, S., GERDING, D. N. & VEDANTAM, G. 2010. Human hypervirulent *Clostridium difficile* strains exhibit increased sporulation as well as robust toxin production. *J Bacteriol,* 192**,** 4904-11.

MERRIL, C. R., SCHOLL, D. & ADHYA, S. L. 2003. The prospect for bacteriophage therapy in Western medicine. *Nat Rev Drug Discov,* 2**,** 489-97.

METCALF, D., AVERY, B. P., JANECKO, N., MATIC, N., REID-SMITH, R. & WEESE, J. S. 2011. *Clostridium difficile* in seafood and fish. *Anaerobe,* 17**,** 85-6.

METCALF, D. S., COSTA, M. C., DEW, W. M. & WEESE, J. S. 2010. *Clostridium difficile* in vegetables, Canada. *Lett Appl Microbiol,* 51**,** 600-2.

MINOT, S., BRYSON, A., CHEHOUD, C., WU, G. D., LEWIS, J. D. & BUSHMAN, F. D. 2013. Rapid evolution of the human gut virome. *Proceedings of the National Academy of Sciences of the United States of America,* 110**,** 12450-12455.

MIZUSAWA, M., DORON, S. & GORBACH, S. 2015. *Clostridium difficile* Diarrhea in the Elderly: Current Issues and Management Options. *Drugs Aging,* 32**,** 639-47.

MONCRIEF, J. S., BARROSO, L. A. & WILKINS, T. D. 1997. Positive regulation of *Clostridium difficile* toxins. *Infect Immun,* 65**,** 1105-8.

MONOT, M., BOURSAUX-EUDE, C., THIBONNIER, M., VALLENET, D., MOSZER, I., MEDIGUE, C., MARTIN-VERSTRAETE, I. & DUPUY, B. 2011. Reannotation of the genome sequence of *Clostridium difficile* strain 630. *Journal of Medical Microbiology,* 60**,** 1193-1199.

MONOT, M., ECKERT, C., LEMIRE, A., HAMIOT, A., DUBOIS, T., TESSIER, C., DUMOULARD, B., HAMEL, B., PETIT, A., LALANDE, V., MA, L., BOUCHIER, C., BARBUT, F. & DUPUY, B. 2015. *Clostridium difficile*: New Insights into the Evolution of the Pathogenicity Locus. *Scientific Reports,* 5**,** 15023.

MORGAN, O. W., RODRIGUES, B., ELSTON, T., VERLANDER, N. Q., BROWN, D. F. J., BRAZIER, J. & REACHER, M. 2008. Clinical Severity of *Clostridium difficile* PCR Ribotype 027: A Case-Case Study. *PLoS ONE,* 3**,** e1812.

MULLANY, P., ALLAN, E. & ROBERTS, A. P. 2015. Mobile genetic elements in *Clostridium difficile* and their role in genome function. *Res Microbiol,* 166**,** 361-7.

MUTO, C. A., POKRYWKA, M., SHUTT, K., MENDELSOHN, A. B., NOURI, K., POSEY, K., ROBERTS, T., CROYLE, K., KRYSTOFIAK, S., PATEL-BROWN, S., PASCULLE, A. W., PATERSON, D. L., SAUL, M. & HARRISON, L. H. 2005. A large outbreak of *Clostridium difficile*-associated disease with an unexpected proportion of deaths and colectomies at a teaching hospital following increased fluoroquinolone use. *Infect Control Hosp Epidemiol,* 26**,** 273-80.

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.

NALE, J. Y. 2013. *Isolation and Characterisation of Temperate Bacteriophages of the Hypervirulent Clostridium difficile 027 Strains.* University of Leicester.

NALE, J. Y., SHAN, J., HICKENBOTHAM, P. T., FAWLEY, W. N., WILCOX, M. H. & CLOKIE, M. R. 2012. Diverse temperate bacteriophage carriage in *Clostridium difficile* 027 strains. *PloS one,* 7**,** e37263.

NALE, J. Y., SPENCER, J., HARGREAVES, K. R., BUCKLEY, A. M., TRZEPINSKI, P., DOUCE, G. R. & CLOKIE, M. R. 2015a. Bacteriophage Combinations Significantly Reduce *Clostridium difficile* Growth In Vitro and Proliferation In Vivo. *Antimicrob Agents Chemother,* 60**,** 968-81.

NALE, J. Y., SPENCER, J., HARGREAVES, K. R., BUCKLEY, A. M., TRZEPINSKI, P., DOUCE, G. R. & CLOKIE, M. R. 2015b. Better together: bacteriophage combinations significantly reduce *Clostridium difficile* growth in vitro and proliferation in vivo. *Antimicrob Agents Chemother*.

NANWA, N., KENDZERSKA, T., KRAHN, M., KWONG, J. C., DANEMAN, N., WITTEMAN, W., MITTMANN, N., CADARETTE, S. M., ROSELLA, L. & SANDER, B. 2015. The Economic Impact of *Clostridium difficile* Infection: A Systematic Review. *Am J Gastroenterol,* 110**,** 511-519.

NAWROCKI, E. P. & EDDY, S. R. 2013. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics,* 29**,** 2933-5.

NICHOLSON, M. R., THOMSEN, I. P., SLAUGHTER, J. C., CREECH, C. B. & EDWARDS, K. M. 2015. Novel risk factors for recurrent *Clostridium difficile* infection in children. *J Pediatr Gastroenterol Nutr,* 60**,** 18-22.

NICKLIN, J. 1999. *Instant notes in microbiology*, Bios Scientific Publishers; Springer.

NIEVES, B. M., GIL, F. & CASTILLO, F. J. 1981. Growth inhibition activity and bacteriophage and bacteriocinlike particles associated with different species of *Clostridium*. *Canadian journal of microbiology,* 27**,** 216-225.

NIKAEEN, M., DEHNAVI, H. A., HSSANZADEH, A. & JALALI, M. 2015. Occurrence of *Clostridium difficile* in two types of wastewater treatment plants. *Journal of the Formosan Medical Association,* 114**,** 663-665.

NILSSON, H., CARDOSO-PALACIOS, C., HAGGARD-LJUNGQUIST, E. & NILSSON, A. S. 2011. Phylogenetic structure and evolution of regulatory genes and integrases of P2-like phages. *Bacteriophage,* 1**,** 207-218.

NOBREGA, F. L., COSTA, A. R., KLUSKENS, L. D. & AZEREDO, J. 2015. Revisiting phage therapy: new applications for old resources. *Trends in Microbiology,* 23**,** 185-191.

NOLLER, A. C., MCELLISTREM, M. C., PACHECO, A. G., BOXRUD, D. J. & HARRISON, L. H. 2003. Multilocus variable-number tandem repeat analysis distinguishes outbreak and sporadic *Escherichia coli* O157:H7 isolates. *J Clin Microbiol,* 41**,** 5389-97.

NORÉN, T. 2010. *Clostridium difficile* and the Disease It Causes. *In:* MULLANY, P. & ROBERTS, P. A. (eds.) *Clostridium difficile: Methods and Protocols.* Totowa, NJ: Humana Press.

NORMAN, K. N., HARVEY, R. B., SCOTT, H. M., HUME, M. E., ANDREWS, K. & BRAWLEY, A. D. 2009. Varied prevalence of *Clostridium difficile* in an integrated swine operation. *Anaerobe,* 15**,** 256-60.

O'NEIL, H. S. & MARQUIS, H. 2006. *Listeria monocytogenes* flagella are used for motility, not as adhesins, to increase host cell invasion. *Infect Immun,* 74**,** 6675-81.

O'NEILL, G., ADAMS, J. E., BOWMAN, R. A. & RILEY, T. V. 1993. A molecular characterization of *Clostridium difficile* isolates from humans, animals and their environments. *Epidemiology & Infection,* 111**,** 257-264.

O'NEILL, G., OGUNSOLA, F., BRAZIER, J. & DUERDEN, B. 1996. Modification of a PCR Ribotyping Method for Application as a Routine Typing Scheme for *Clostridium difficile*. *Anaerobe,* 2**,** 205-209.

OKONECHNIKOV, K., CONESA, A. & GARCIA-ALCALDE, F. 2015. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*.

OLIVEIRA, H., MELO, L. D., SANTOS, S. B., NOBREGA, F. L., FERREIRA, E. C., CERCA, N., AZEREDO, J. & KLUSKENS, L. D. 2013. Molecular aspects and comparative genomics of bacteriophage endolysins. *J Virol,* 87**,** 4558-70.

OTIENO, J. R. 2013. *ARE NATURAL ENVIRONMENTS BREEDING GROUNDS FOR THE EMERGENCE OF NEW PATHOGENIC STRAINS OF THE HOSPITAL ACQUIRED BACTERIAL PATHOGEN CLOSTRIDIUM DIFFICILE?*

. MSc, University of Leicester.

OVERBEEK, R., OLSON, R., PUSCH, G. D., OLSEN, G. J., DAVIS, J. J., DISZ, T., EDWARDS, R. A., GERDES, S., PARRELLO, B., SHUKLA, M., VONSTEIN, V., WATTAM, A. R., XIA, F. & STEVENS, R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research,* 42**,** D206-D214.

OWENS, R. C., DONSKEY, C. J., GAYNES, R. P., LOO, V. G. & MUTO, C. A. 2008. Antimicrobial-associated risk factors for *Clostridium difficile* infection. *Clinical Infectious Diseases,* 46**,** S19-S31.

PAREDES-SABJA, D., BOND, C., CARMAN, R. J., SETLOW, P. & SARKER, M. R. 2008. Germination of spores of *Clostridium difficile* strains, including isolates from a hospital outbreak of *Clostridium difficile*-associated disease (CDAD). *Microbiology,* 154**,** 2241-50.

PAREDES-SABJA, D., SHEN, A. & SORG, J. A. 2014. *Clostridium difficile* spore biology: sporulation, germination, and spore structural proteins. *Trends in Microbiology,* 22**,** 406-416.

PARIJA, S. C. 2014. *Textbook of microbiology & immunology*, Elsevier Health Sciences.

PARKINSON, J. S. 1993. Signal transduction schemes of bacteria. *Cell,* 73**,** 857-71.

PASQUALE, V., ROMANO, V., RUPNIK, M., CAPUANO, F., BOVE, D., ALIBERTI, F., KROVACEK, K. & DUMONTET, S. 2012. Occurrence of toxigenic *Clostridium difficile* in edible bivalve molluscs. *Food Microbiol,* 31**,** 309-12.

PASQUALE, V., ROMANO, V. J., RUPNIK, M., DUMONTET, S., CIZNAR, I., ALIBERTI, F., MAURI, F., SAGGIOMO, V. & KROVACEK, K. 2011. Isolation and characterization of *Clostridium difficile* from shellfish and marine environments. *Folia Microbiol (Praha),* 56**,** 431-7.

PATEL, K. 2013. *Investigation into the dominant strains of Clostridium difficile within hospitals and strategic cleaning regimes.* © Krusha Patel.

PECHINE, S., JANOIR, C. & COLLIGNON, A. 2005. Variability of *Clostridium difficile* surface proteins and specific serum antibody response in patients with *Clostridium difficile*-associated disease. *J Clin Microbiol,* 43**,** 5018-25.

PEPIN, J. 2008. Vancomycin for the treatment of *Clostridium difficile* Infection: for whom is this expensive bullet really magic? *Clin Infect Dis,* 46**,** 1493-8.

PEPIN, J., ALARY, M. E., VALIQUETTE, L., RAICHE, E., RUEL, J., FULOP, K., GODIN, D. & BOURASSA, C. 2005. Increasing risk of relapse after treatment of *Clostridium difficile* colitis in Quebec, Canada. *Clin Infect Dis,* 40**,** 1591-7.

PEPIN, J., VALIQUETTE, L., ALARY, M. E., VILLEMURE, P., PELLETIER, A., FORGET, K., PEPIN, K. & CHOUINARD, D. 2004. *Clostridium difficile*-associated diarrhea in a region of Quebec from 1991 to 2003: a changing pattern of disease severity. *CMAJ,* 171**,** 466-72.

PERELLE, S., GIBERT, M., BOURLIOUX, P., CORTHIER, G. & POPOFF, M. R. 1997. Production of a complete binary toxin (actin-specific ADP-ribosyltransferase) by *Clostridium difficile* CD196. *Infect Immun,* 65**,** 1402-7.

PERKINS, S. E., FOX, J. G., TAYLOR, N. S., GREEN, D. L. & LIPMAN, N. S. 1995. Detection of *Clostridium difficile* toxins from the small intestine and cecum of rabbits with naturally acquired enterotoxemia. *Lab Anim Sci,* 45**,** 379-84.

PERRIÈRE, G. & GOUY, M. 1996. WWW-query: An on-line retrieval system for biological sequence banks. *Biochimie,* 78**,** 364-369.

PIRS, T., OCEPEK, M. & RUPNIK, M. 2008. Isolation of *Clostridium difficile* from food animals in Slovenia. *J Med Microbiol,* 57**,** 790-2.

PITUCH, H., OBUCH-WOSZCZATYNSKI, P., VAN DEN BRAAK, N., VAN BELKUM, A., KUJAWA, M., LUCZAK, M. & MEISEL-MIKOLAJCZYK, F. 2002. Variable flagella expression among clonal toxin A-/B+*Clostridium difficile* strains with highly homogeneous flagellin genes. *Clin Microbiol Infect,* 8**,** 187-8.

POILANE, I., KARJALAINEN, T., BARC, M.-C., BOURLIOUX, P. & COLLIGNON, A. 1998. Protease activity of *Clostridium difficile* strains. *Canadian journal of microbiology,* 44**,** 157-161.

POLI, A., DI MATTEO, S., BRUNO, G. M., FORNAI, E., VALENTINO, M. C. & COLOMBO, G. L. 2015. Economic burden of *Clostridium difficile* in five hospitals of the Florence health care system in Italy. *Risk Management and Healthcare Policy,* 8**,** 207-213.

POPE, W. H., JACOBS-SERA, D., RUSSELL, D. A., PEEBLES, C. L., AL-ATRACHE, Z., ALCOSER, T. A., ALEXANDER, L. M., ALFANO, M. B., ALFORD, S. T., AMY, N. E., ANDERSON, M. D., ANDERSON, A. G., ANG, A. A., ARES, M., JR., BARBER, A. J., BARKER, L. P., BARRETT, J. M., BARSHOP, W. D., BAUERLE, C. M., BAYLES, I. M., BELFIELD, K. L., BEST, A. A., BORJON, A., JR., BOWMAN, C. A., BOYER, C. A., BRADLEY, K. W., BRADLEY, V. A., BROADWAY, L. N., BUDWAL, K., BUSBY, K. N., CAMPBELL, I. W., CAMPBELL, A. M., CAREY, A., CARUSO, S. M., CHEW, R. D., COCKBURN, C. L., COHEN, L. B., CORAJOD, J. M., CRESAWN, S. G., DAVIS, K. R., DENG, L., DENVER, D. R., DIXON, B. R., EKRAM, S., ELGIN, S. C., ENGELSEN, A. E., ENGLISH, B. E., ERB, M. L., ESTRADA, C., FILLIGER, L. Z., FINDLEY, A. M., FORBES, L., FORSYTH, M. H., FOX, T. M., FRITZ, M. J., GARCIA, R., GEORGE, Z. D., GEORGES, A. E., GISSENDANNER, C. R., GOFF, S., GOLDSTEIN, R., GORDON, K. C., GREEN, R. D., GUERRA, S. L., GUINEY-OLSEN, K. R., GUIZA, B. G., HAGHIGHAT, L., HAGOPIAN, G. V., HARMON, C. J., HARMSON, J. S., HARTZOG, G. A., HARVEY, S. E., HE, S., HE, K. J., HEALY, K. E., HIGINBOTHAM, E. R., HILDEBRANDT, E. N., HO, J. H., HOGAN, G. M., HOHENSTEIN, V. G., HOLZ, N. A., HUANG, V. J., HUFFORD, E. L., HYNES, P. M., JACKSON, A. S., JANSEN, E. C., JARVIK, J., JASINTO, P. G., JORDAN, T. C., KASZA, T., KATELYN, M. A., KELSEY, J. S., KERRIGAN, L. A., KHAW, D., KIM, J., KNUTTER, J. Z., KO, C. C., LARKIN, G. V., LAROCHE, J. R., LATIF, A., et al. 2011. Expanding the diversity of mycobacteriophages: insights into genome architecture and evolution. *PLoS One,* 6**,** e16329.

POSTNOVA, T., GOMEZ-DUARTE, O. G. & RICHARDSON, K. 1996. Motility mutants of Vibrio cholerae O1 have reduced adherence in vitro to human small intestinal epithelial cells as demonstrated by ELISA. *Microbiology,* 142 (Pt 10)**,** 2767-76.

POTHOULAKIS, C., CASTAGLIUOLO, I., KELLY, C. P. & LAMONT, J. T. 1993. *Clostridium difficile*-associated diarrhea and colitis: pathogenesis and therapy. *Int J Antimicrob Agents,* 3**,** 17-32.

POXTON, I. R., MCCOUBREY, J. & BLAIR, G. 2001. The pathogenicity of *Clostridium difficile*. *Clin Microbiol Infect,* 7**,** 421-7.

PRUITT, R. N. & LACY, D. B. 2012. Toward a structural understanding of *Clostridium difficile* toxins A and B. *Frontiers in Cellular and Infection Microbiology,* 2**,** 28.

QUESADA-GOMEZ, C., MULVEY, M. R., VARGAS, P., GAMBOA-CORONADO MDEL, M., RODRIGUEZ, C. & RODRIGUEZ-CAVILLINI, E. 2013. Isolation of a toxigenic and clinical genotype of *Clostridium difficile* in retail meats in Costa Rica. *J Food Prot,* 76**,** 348-51.

RAHIMI, E., AFZALI, Z. S. & BAGHBADORANI, Z. T. 2015. *Clostridium difficile* in ready-to-eat foods in Isfahan and Shahrekord, Iran. *Asian Pacific Journal of Tropical Biomedicine,* 5**,** 128-131.

RAKHUBA, D., KOLOMIETS, E., DEY, E. S. & NOVIK, G. 2010. Bacteriophage receptors, mechanisms of phage adsorption and penetration into host cell. *Pol J Microbiol,* 59**,** 145-155.

RAMESH, V., FRALICK, J. A. & ROLFE, R. D. 1999. Prevention of *Clostridium difficile*-induced ileocecitis with bacteriophage. *Anaerobe,* 5**,** 69-78.

RAMOS, H. C., RUMBO, M. & SIRARD, J.-C. 2004. Bacterial flagellins: mediators of pathogenicity and host immune responses in mucosa. *Trends in microbiology,* 12**,** 509-517.

RAO, V. B. & FEISS, M. 2008. The bacteriophage DNA packaging motor. *Annual review of genetics,* 42**,** 647-681.

RAO, V. B. & FEISS, M. 2015. Mechanisms of DNA Packaging by Large Double-Stranded DNA Viruses. *Annu Rev Virol,* 2**,** 351-78.

RASHID, S. R. J., CLOKIE, M. R. J. & MILLARD, A. D. 2016. Draft Genome Sequences of Three Novel *Clostridium* Isolates from Northern Iraq. *Genome Announcements,* 4**,** e00033-16.

RENTAS, F. J. & RAO, V. B. 2003. Defining the bacteriophage T4 DNA packaging machine: evidence for a C-terminal DNA cleavage domain in the large terminase/packaging protein gp17. *J Mol Biol,* 334**,** 37-52.

REPORT, H. P. 2015. Health protection report 9.

RIEGLER, M., SEDIVY, R., POTHOULAKIS, C., HAMILTON, G., ZACHERL, J., BISCHOF, G., COSENTINI, E., FEIL, W., SCHIESSEL, R. & LAMONT, J. T. 1995. *Clostridium difficile* toxin B is more potent than toxin A in damaging human colonic epithelium in vitro. *Journal of Clinical Investigation,* 95**,** 2004-2011.

RINTTILA, T., KASSINEN, A., MALINEN, E., KROGIUS, L. & PALVA, A. 2004. Development of an extensive set of 16S rDNA-targeted primers for quantification of pathogenic and indigenous bacteria in faecal samples by real-time PCR. *J Appl Microbiol,* 97**,** 1166-77.

ROMANO, V., PASQUALE, V., KROVACEK, K., MAURI, F., DEMARTA, A. & DUMONTET, S. 2012. Toxigenic *Clostridium difficile* PCR ribotypes from wastewater treatment plants in southern Switzerland. *Appl Environ Microbiol,* 78**,** 6643-6.

ROSEMAN, D. & RICHARDSON, R. L. 1969. Isolation of bacteriophage for *Clostridium tetani*. *Journal of virology,* 3**,** 350-350.

ROSEY, E. L., KENNEDY, M. J. & YANCEY, R. J., JR. 1996. Dual flaA1 flaB1 mutant of Serpulina hyodysenteriae expressing periplasmic flagella is severely attenuated in a murine model of swine dysentery. *Infect Immun,* 64**,** 4154-62.

ROTIMI, V. O., JAMAL, W. Y., MOKADDAS, E. M., BRAZIER, J. S., JOHNY, M. & DUERDEN, B. I. 2003. Prevalent PCR ribotypes of clinical and environmental strains of *Clostridium difficile* isolated from intensive-therapy unit patients in Kuwait. *J Med Microbiol,* 52**,** 705-9.

ROZEN, S. & SKALETSKY, H. 1999. Primer3 on the WWW for general users and for biologist programmers. *Bioinformatics methods and protocols.* Springer.

RUPNIK, M. 2001. How to detect *Clostridium difficile* variant strains in a routine laboratory. *Clin Microbiol Infect,* 7**,** 417-20.

RUPNIK, M. 2010. *Clostridium difficile* toxinotyping. *Methods Mol Biol,* 646**,** 67-76.

RUPNIK, M., AVESANI, V., JANC, M., VON EICHEL-STREIBER, C. & DELMEE, M. 1998. A novel toxinotyping scheme and correlation of toxinotypes with serogroups of *Clostridium difficile* isolates. *J Clin Microbiol,* 36.

RUPNIK, M., BRAUN, V., SOEHN, F., JANC, M., HOFSTETTER, M., LAUFENBERG-FELDMANN, R. & VON EICHEL-STREIBER, C. 1997. Characterization of polymorphisms in the toxin A and B genes of *Clostridium difficile*. *FEMS Microbiol Lett,* 148**,** 197-202.

RUPNIK, M. & JANEZIC, S. 2016. An Update on *Clostridium difficile* Toxinotyping. *J Clin Microbiol,* 54**,** 13-8.

RUPNIK, M. & SONGER, J. G. 2010a. Chapter 3 - *Clostridium difficile*: Its Potential as a Source of Foodborne Disease. *In:* STEVE, L. T. (ed.) *Advances in Food and Nutrition Research.* Academic Press.

RUPNIK, M. & SONGER, J. G. 2010b. *Clostridium difficile* Its Potential as a Source of Foodborne Disease. *Adv Food Nutr Res,* 60.

RUPNIK, M., WIDMER, A., ZIMMERMANN, O., ECKERT, C. & BARBUT, F. 2008. *Clostridium difficile* toxinotype V, ribotype 078, in animals and humans. *J Clin Microbiol,* 46**,** 2146.

RUPNIK, M., WILCOX, M. H. & GERDING, D. N. 2009. *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. *Nat Rev Micro,* 7**,** 526-536.

RYAN, A., LYNCH, M., SMITH, S. M., AMU, S., NEL, H. J., MCCOY, C. E., DOWLING, J. K., DRAPER, E., O'REILLY, V., MCCARTHY, C., O'BRIEN, J., NÍ EIDHIN, D., O'CONNELL, M. J., KEOGH, B., MORTON, C. O., ROGERS, T. R., FALLON, P. G., O'NEILL, L. A., KELLEHER, D. & LOSCHER, C. E. 2011. A Role for TLR4 in *Clostridium difficile* Infection and the Recognition of Surface Layer Proteins. *PLoS Pathogens,* 7**,** e1002076.

SABRINA, B., KÅRE, M., MARIANNE, K. K. & KATHARINA, E. P. O. 2011. Binary Toxin and Death after *Clostridium difficile* Infection. *Emerging Infectious Disease journal,* 17**,** 976.

SADEGHIFARD, N., GURTLER, V., BEER, M. & SEVIOUR, R. J. 2006. The mosaic nature of intergenic 16S-23S rRNA spacer regions suggests rRNA operon copy number variation in *Clostridium difficile* strains. *Appl Environ Microbiol,* 72**,** 7311-23.

SADEGHIFARD, N., SALARI, M. H., GHASSEMI, M. R., ESHRAGHI, S. & AMIN HARATI, F. 2010a. The incidence of nosocomial toxigenic *Clostridium difficile* associated diarrhea in Tehran tertiary medical centers. *Acta Med Iran,* 48.

SADEGHIFARD, N., SALARI, M. H., RANJBAR, R., GHAFOURYAN, S., RAFTARI, M., ABDULAMIR, A. S., FATIMAH, A. B. & KAZEMI, B. 2010b. The clinical and environmental spread and diversity of toxigenic *Clostridium difficile* diarrheic in the region of the Middle East. *Rev Infection,* 1.

SAMBOL, S. P., MERRIGAN, M. M., LYERLY, D., GERDING, D. N. & JOHNSON, S. 2000. Toxin gene analysis of a variant strain of *Clostridium difficile* that causes human clinical disease. *Infect Immun,* 68**,** 5480-7.

SAMBROOK, J. & RUSSELL, D. W. 2001. *Molecular cloning: a laboratory manual,* New York, CSHL press, Cold Spring Harbor Laboratory Press.

SANGSTER, W., HEGARTY, J. P. & STEWART, D. B., SR. 2015. Phage tail-like particles kill *Clostridium difficile* and represent an alternative to conventional antibiotics. *Surgery,* 157**,** 96-103.

SCHALLEHN, G. 1985. *Clostridium sordellii* bacteriophage active on *Clostridium difficile*. *Letters in applied microbiology,* 1**,** 71-74.

SCHATTNER, P., BROOKS, A. N. & LOWE, T. M. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Research,* 33.

SCHWAN, C., STECHER, B., TZIVELEKIDIS, T., VAN HAM, M., ROHDE, M., HARDT, W. D., WEHLAND, J. & AKTORIES, K. 2009. *Clostridium difficile* toxin CDT induces formation of microtubule-based protrusions and increases adherence of bacteria. *PLoS Pathog,* 5**,** e1000626.

SEBAIHIA, M., WREN, B. W., MULLANY, P., FAIRWEATHER, N. F., MINTON, N., STABLER, R., THOMSON, N. R., ROBERTS, A. P., CERDENO-TARRAGA, A. M., WANG, H., HOLDEN, M. T., WRIGHT, A., CHURCHER, C., QUAIL, M. A., 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.

SEDDON, S. V., HEMINGWAY, I. & BORRIELLO, S. P. 1990. Hydrolytic enzyme production by *Clostridium difficile* and its relationship to toxin production and virulence in the hamster model. *Journal of Medical Microbiology,* 31**,** 169-174.

SEED, K. D., LAZINSKI, D. W., CALDERWOOD, S. B. & CAMILLI, A. 2013. A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature,* 494**,** 489-91.

SEEMANN, T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics,* 30**,** 2068-9.

SEKULOVIC, O. & FORTIER, L. C. 2015. Global transcriptional response of *Clostridium difficile* carrying the CD38 prophage. *Appl Environ Microbiol,* 81**,** 1364-74.

SEKULOVIC, O., GARNEAU, J. R., NERON, A. & FORTIER, L. C. 2014. Characterization of temperate phages infecting *Clostridium difficile* isolates of human and animal origins. *Appl Environ Microbiol,* 80**,** 2555-63.

SEKULOVIC, O., MEESSEN-PINARD, M. & FORTIER, L. C. 2011. Prophage-stimulated toxin production in *Clostridium difficile* NAP1/027 lysogens. *J Bacteriol,* 193**,** 2726-34.

SEKULOVIC, O., OSPINA BEDOYA, M., FIVIAN-HUGHES, A. S., FAIRWEATHER, N. F. & FORTIER, L. C. 2015. The *Clostridium difficile* cell wall protein CwpV confers phase-variable phage resistance. *Mol Microbiol*.

SELL, T. L., SCHABERG, D. R. & FEKETY, F. R. 1983. Bacteriophage and bacteriocin typing scheme for *Clostridium difficile*. *Journal of Clinical Microbiology,* 17**,** 1148-1152.

SHAN, J., PATEL, K. V., HICKENBOTHAM, P. T., NALE, J. Y., HARGREAVES, K. R. & CLOKIE, M. R. 2012. Prophage carriage and diversity within clinically relevant strains of *Clostridium difficile*. *Appl Environ Microbiol,* 78**,** 6027-34.

SHERIDAN, P. P., MITEVA, V. I. & BRENCHLEY, J. E. 2003. Phylogenetic analysis of anaerobic psychrophilic enrichment cultures obtained from a Greenland glacier ice core. *Applied and Environmental Microbiology,* 69**,** 2153-2160.

SHIMAMURA, T., NAKAMURA, S., HAYASE, M. & NISHIDA, S. 1974. Mitomycin-induced lysis of *Clostridium sordellii*. *J Med Microbiol,* 7**,** 277-83.

SIANI, H. 2014. *The identification of therapeutic targets and virulence factors of Clostridium difficile.* Master of Philosophiae, Cardiff University.

SIEFERT, J. 2009. Defining the Mobilome. *In:* GOGARTEN, M., GOGARTEN, J. & OLENDZENSKI, L. (eds.) *Horizontal Gene Transfer.* Humana Press.

SILVA, R. O. S., SANTOS, R. L. R., PIRES, P. S., PEREIRA, L. C., PEREIRA, S. T., DUARTE, M. C., DE ASSIS, R. A. & LOBATO, F. C. F. 2013. Detection of toxins A/B and isolation of *Clostridium difficile* and *Clostridium perfringens* from dogs in Minas Gerais, Brazil. *Brazilian Journal of Microbiology,* 44**,** 133-137.

SIMANGO, C. 2006. Prevalence of *Clostridium difficile* in the environment in a rural community in Zimbabwe. *Trans R Soc Trop Med Hyg,* 100**,** 1146-50.

SIMANGO, C. & MWAKURUDZA, S. 2008. *Clostridium difficile* in broiler chickens sold at market places in Zimbabwe and their antimicrobial susceptibility. *Int J Food Microbiol,* 124**,** 268-70.

SIMON, M. N., DAVIS, R. W. & DAVIDSON, N. 1971. Heteroduplexes of DNA molecules of lambdoid phages: physical mapping of their base sequence relationships by electron microscopy. *Cold Spring Harbor Monograph Archive,* 2**,** 313-328.

SMITH, A. 2005. Outbreak of *Clostridium difficile* infection in an English hospital linked to hypertoxin-producing strains in Canada and the US. *Euro Surveill,* 10**,** E050630 2.

SMITS, W. K., LYRAS, D., LACY, D. B., WILCOX, M. H. & KUIJPER, E. J. 2016. *Clostridium difficile* infection. *Nat Rev Dis Primers,* 2**,** 16020.

SONGER, J. G., TRINH, H. T., KILLGORE, G. E., THOMPSON, A. D., MCDONALD, L. C. & LIMBAGO, B. M. 2009. *Clostridium difficile* in retail meat products, USA, 2007. *Emerg Infect Dis,* 15**,** 819-21.

SORG, J. A. & DINEEN, S. S. 2009. Laboratory maintenance of *Clostridium difficile*. *Curr Protoc Microbiol,* Chapter 9**,** Unit9A 1.

SOUTOURINA, O. A., MONOT, M., BOUDRY, P., SAUJET, L., PICHON, C., SISMEIRO, O., SEMENOVA, E., SEVERINOV, K., LE BOUGUENEC, C., COPPEE, J. Y., DUPUY, B. & MARTIN-VERSTRAETE, I. 2013. Genome-wide identification of regulatory RNAs in the human pathogen *Clostridium difficile*. *PLoS Genet,* 9**,** e1003493.

STABLER, R., GERDING, D., SONGER, J., DRUDY, D., BRAZIER, J., TRINH, H., WITNEY, A., HINDS, J. & WREN, B. 2006. Comparative phylogenomics of *Clostridium difficile* reveals clade specificity and microevolution of hypervirulent strains. *Journal of bacteriology,* 188**,** 7297-7305.

STABLER, R. A., DAWSON, L. F., VALIENTE, E., CAIRNS, M. D., MARTIN, M. J., DONAHUE, E. H., RILEY, T. V., SONGER, J. G., KUIJPER, E. J., DINGLE, K. E. & WREN, B. W. 2012. Macro and micro diversity of *Clostridium difficile* isolates from diverse sources and geographical locations. *PLoS One,* 7**,** e31559.

STABLER, R. A., HE, M., DAWSON, L., MARTIN, M., VALIENTE, E., CORTON, C., LAWLEY, T. D., SEBAIHIA, M., QUAIL, M. A., ROSE, G., GERDING, D. N., GIBERT, M., POPOFF, M. R., PARKHILL, J., DOUGAN, G. & WREN, B. W. 2009. Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the evolution of a hypervirulent bacterium. *Genome Biol,* 10**,** R102.

STAMATAKIS, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics,* 30**,** 1312-3.

STARE, B. G., DELMEE, M. & RUPNIK, M. 2007. Variant forms of the binary toxin CDT locus and tcdC gene in *Clostridium difficile* strains. *J Med Microbiol,* 56**,** 329-35.

STEFFEN, E. K. & HENTGES, D. J. 1981. Hydrolytic enzymes of anaerobic bacteria isolated from human infections. *J Clin Microbiol,* 14**,** 153-6.

STERN, A., MICK, E., TIROSH, I., SAGY, O. & SOREK, R. 2012. CRISPR targeting reveals a reservoir of common phages associated with the human gut microbiome. *Genome Res,* 22**,** 1985-94.

STEVENSON, E., MINTON, N. P. & KUEHNE, S. A. 2015. The role of flagella in *Clostridium difficile* pathogenicity. *Trends in Microbiology,* 23**,** 275-282.

STUBBS, S., RUPNIK, M., GIBERT, M., BRAZIER, J., DUERDEN, B. & POPOFF, M. 2000. Production of actin-specific ADP-ribosyltransferase (binary toxin) by strains of *Clostridium difficile*. *FEMS Microbiol Lett,* 186**,** 307-12.

STUBBS, S. L. J., BRAZIER, J. S., O'NEILL, G. L. & DUERDEN, B. I. 1999. PCR Targeted to the *16S-23S rRNA* Gene Intergenic Spacer Region of *Clostridium difficile* and Construction of a Library Consisting of 116 Different PCR Ribotypes. *Journal of Clinical Microbiology,* 37**,** 461-463.

SULAKVELIDZE, A., ALAVIDZE, Z. & MORRIS, J. G., JR. 2001. Bacteriophage therapy. *Antimicrob Agents Chemother,* 45**,** 649-59.

SULLIVAN, N. M., PELLETT, S. & WILKINS, T. D. 1982. Purification and characterization of toxins A and B of *Clostridium difficile*. *Infect Immun,* 35**,** 1032-40.

SURAWICZ, C. M., BRANDT, L. J., BINION, D. G., ANANTHAKRISHNAN, A. N., CURRY, S. R., GILLIGAN, P. H., MCFARLAND, L. V., MELLOW, M. & ZUCKERBRAUN, B. S. 2013. Guidelines for Diagnosis, Treatment, and Prevention of *Clostridium difficile* Infections. *Am J Gastroenterol,* 108**,** 478-498.

SUTTLE, C. A. 2007. Marine viruses--major players in the global ecosystem. *Nat Rev Microbiol,* 5**,** 801-12.

SUTTON, P. A., LI, S., WEBB, J., SOLOMON, K., BRAZIER, J. & MAHIDA, Y. R. 2008. Essential role of toxin A in *C. difficile* 027 and reference strain supernatant-mediated disruption of Caco-2 intestinal epithelial barrier function. *Clin Exp Immunol,* 153**,** 439-47.

SVENUNGSSON, B., BURMAN, L. G., JALAKAS-PORNULL, K., LAGERGREN, A., STRUWE, J. & AKERLUND, T. 2003. Epidemiology and molecular characterization of *Clostridium difficile* strains from patients with diarrhea: low disease incidence and evidence of limited cross-infection in a Swedish teaching hospital. *J Clin Microbiol,* 41**,** 4031-7.

TAKUMI, K., KINOUCHI, T. & KAWATA, T. 1980. Isolation of two inducible bacteriophages from *Clostridium botulinum* type A 190L. *FEMS Microbiology Letters,* 9**,** 23-27.

TAN, K. S., WEE, B. Y. & SONG, K. P. 2001. Evidence for holin function of tcdE gene in the pathogenicity of *Clostridium difficile*. *J Med Microbiol,* 50**,** 613-9.

TASTEYRE, A., BARC, M. C., COLLIGNON, A., BOUREAU, H. & KARJALAINEN, T. 2001. Role of FliC and FliD flagellar proteins of *Clostridium difficile* in adherence and gut colonization. *Infect Immun,* 69**,** 7937-40.

TAYLOR, N. S., THORNE, G. M. & BARTLETT, J. G. 1981. Comparison of two toxins produced by *Clostridium difficile*. *Infect Immun,* 34**,** 1036-43.

TEASLEY, D. G., GERDING, D. N., OLSON, M. M., PETERSON, L. R., GEBHARD, R. L., SCHWARTZ, M. J. & LEE, J. T., JR. 1983. Prospective randomised trial of metronidazole versus vancomycin for *Clostridium-difficile*-associated diarrhoea and colitis. *Lancet,* 2**,** 1043-6.

TEDESCO, F. 1976. Clindamycin-associated colitis. Review of the clinical spectrum of 47 cases. *The American journal of digestive diseases,* 21**,** 26-32.

TENOVER, F. C., ÅKERLUND, T., GERDING, D. N., GOERING, R. V., BOSTRÖM, T., JONSSON, A.-M., WONG, E., WORTMAN, A. T. & PERSING, D. H. 2011. Comparison of Strain Typing Results for *Clostridium difficile* Isolates from North America. *Journal of Clinical Microbiology,* 49**,** 1831-1837.

TERHES, G., BRAZIER, J. S., URBAN, E., SOKI, J. & NAGY, E. 2006. Distribution of *Clostridium difficile* PCR ribotypes in regions of Hungary. *J Med Microbiol,* 55**,** 279-82.

THANKI, A. M. 2016. *Development of a phage-based diagnostic test for the identification of Clostridium difficile.* Loughborough University.

TICKLER, I. A., GOERING, R. V., WHITMORE, J. D., LYNN, A. N., PERSING, D. H. & TENOVER, F. C. 2014. Strain types and antimicrobial resistance patterns of *Clostridium difficile* isolates from the United States, 2011 to 2013. *Antimicrob Agents Chemother,* 58**,** 4214-8.

VAISHNAVI, C. 2010. Clinical spectrum & pathogenesis of *Clostridium difficile* associated diseases. *Indian J Med Res,* 131**,** 487-99.

VARDAKAS, K. Z., POLYZOS, K. A., PATOUNI, K., RAFAILIDIS, P. I., SAMONIS, G. & FALAGAS, M. E. 2012. Treatment failure and recurrence of *Clostridium difficile* infection following treatment with vancomycin or metronidazole: a systematic review of the evidence. *Int J Antimicrob Agents,* 40**,** 1-8.

VECCHIO, A. L. & ZACUR, G. M. 2012. *Clostridium difficile* infection: an update on epidemiology, risk factors, and therapeutic options. *Current opinion in gastroenterology,* 28**,** 1-9.

VEDANTAM, G., CLARK, A., CHU, M., MCQUADE, R., MALLOZZI, M. & VISWANATHAN, V. K. 2012. *Clostridium difficile* infection: Toxins and non-toxin virulence factors, and their contributions to disease establishment and host response. *Gut Microbes,* 3**,** 121-134.

VISWANATHAN, V. K., MALLOZZI, M. J. & VEDANTAM, G. 2010. *Clostridium difficile* infection: An overview of the disease and its pathogenesis, epidemiology and interventions. *Gut Microbes,* 1**,** 234-242.

VOELKER, R. 2012. Study: Vast majority of *C. difficile* infections occur in medical settings. *JAMA,* 307**,** 1356.

VOHRA, P. & POXTON, I. R. 2011. Comparison of toxin and spore production in clinically relevant strains of *Clostridium difficile*. *Microbiology,* 157**,** 1343-1353.

VON EICHEL-STREIBER, C. & SAUERBORN, M. 1990. *Clostridium difficile* toxin A carries a C-terminal repetitive structure homologous to the carbohydrate binding region of streptococcal glycosyltransferases. *Gene,* 96**,** 107-13.

VOTH, D. E. & BALLARD, J. D. 2005. *Clostridium difficile* Toxins: Mechanism of Action and Role in Disease. *Clinical Microbiology Reviews,* 18**,** 247-263.

WAKITA, Y., SHIMIZU, A., HAJEK, V., KAWANO, J. & YAMASHITA, K. 2002. Characterization of *Staphylococcus* intermedius from pigeons, dogs, foxes, mink, and horses by pulsed-field gel electrophoresis. *J Vet Med Sci,* 64**,** 237-43.

WALIGORA, A.-J., BARC, M.-C., BOURLIOUX, P., COLLIGNON, A. & KARJALAINEN, T. 1999. *Clostridium difficile* Cell Attachment Is Modified by Environmental Factors. *Applied and Environmental Microbiology,* 65**,** 4234-4238.

WALIGORA, A.-J., HENNEQUIN, C., MULLANY, P., BOURLIOUX, P., COLLIGNON, A. & KARJALAINEN, T. 2001. Characterization of a Cell Surface Protein of*Clostridium difficile* with Adhesive Properties. *Infection and immunity,* 69**,** 2144-2153.

WALK, S. T., MICIC, D., JAIN, R., LO, E. S., TRIVEDI, I., LIU, E. W., ALMASSALHA, L. M., EWING, S. A., RING, C., GALECKI, A. T., ROGERS, M. A., WASHER, L., NEWTON, D. W., MALANI, P. N., YOUNG, V. B. & ARONOFF, D. M. 2012. *Clostridium difficile* ribotype does not predict severe infection. *Clin Infect Dis,* 55.

WANG, I. N., SMITH, D. L. & YOUNG, R. 2000. Holins: the protein clocks of bacteriophage infections. *Annu Rev Microbiol,* 54**,** 799-825.

WARNY, M., PEPIN, J., FANG, A., KILLGORE, G., THOMPSON, A., BRAZIER, J., FROST, E. & MCDONALD, L. C. 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.

WASLAWSKI, S., LO, E. S., EWING, S. A., YOUNG, V. B., ARONOFF, D. M., SHARP, S. E., NOVAK-WEEKLEY, S. M., CRIST, A. E., DUNNE, W. M., HOPPE-BAUER, J., JOHNSON, M., BRECHER, S. M., NEWTON, D. W. & WALK, S. T. 2013. *Clostridium difficile* ribotype diversity at six health care institutions in the United States. *J Clin Microbiol,* 51.

WEESE, J. S. 2010. *Clostridium difficile* in food–-innocent bystander or serious threat? *Clinical Microbiology and Infection,* 16**,** 3-10.

WHELAN, S. & GOLDMAN, N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol,* 18**,** 691-9.

WIECZORKIEWICZ, J. T., LOPANSRI, B. K., CHEKNIS, A., OSMOLSKI, J. R., HECHT, D. W., GERDING, D. N. & JOHNSON, S. 2016. Fluoroquinolone and Macrolide Exposure Predict *Clostridium difficile* Infection with the Highly Fluoroquinolone- and Macrolide-Resistant Epidemic *C. difficile* Strain BI/NAP1/027. *Antimicrob Agents Chemother,* 60**,** 418-23.

WIEGAND, P. N., NATHWANI, D., WILCOX, M. H., STEPHENS, J., SHELBAYA, A. & HAIDER, S. 2012. Clinical and economic burden of *Clostridium difficile* infection in Europe: a systematic review of healthcare-facility-acquired infection. *Journal of Hospital Infection,* 81**,** 1-14.

WILCOX, M. H. & FAWLEY, W. N. 2000. Hospital disinfectants and spore formation by *Clostridium difficile*. *The Lancet,* 356**,** 1324.

WILCOX, M. H., SHETTY, N., FAWLEY, W. N., SHEMKO, M., COEN, P., BIRTLES, A., CAIRNS, M., CURRAN, M. D., DODGSON, K. J., GREEN, S. M., HARDY, K. J., HAWKEY, P. M., MAGEE, J. G., SAILS, A. D. & WREN, M. W. 2012. Changing epidemiology of *Clostridium difficile* infection following the introduction of a national ribotyping-based surveillance scheme in England. *Clin Infect Dis,* 55**,** 1056-63.

WILKS, M. & TABAQCHALI, S. 1994. Typing of *Clostridium difficile* by polymerase chain reaction with an arbitrary primer. *J Hosp Infect,* 28**,** 231-4.

WILLIAMS, C. 2001. Occurrence and significance of gastric colonization during acid-inhibitory therapy. *Best Pract Res Clin Gastroenterol,* 15**,** 511-21.

WILSON, K. H., KENNEDY, M. J. & FEKETY, F. R. 1982. Use of sodium taurocholate to enhance spore recovery on a medium selective for *Clostridium difficile*. *J Clin Microbiol,* 15**,** 443-6.

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 Announcements,* 3**,** e00839-15.

WORMSER, G. P., WROE, A. & SCHNEIDERS, T. 2009. *Clostridia*: Molecular Biology in the Post-Genomic Era Edited by Holger Brüggemann and Gerhard Gottschalk Norfolk, United Kingdom: Caister Academic Press, 2009. 230 pp. *Clinical Infectious Diseases,* 49**,** 486-486.

WORTH, L. J., SPELMAN, T., BULL, A. L., BRETT, J. A. & RICHARDS, M. J. 2016. Epidemiology of *Clostridium difficile* infections in Australia: enhanced surveillance to evaluate time trends and severity of illness in Victoria, 2010-2014. *J Hosp Infect*.

WULLT, M., BURMAN, L. G., LAURELL, M. H. & AKERLUND, T. 2003. Comparison of AP-PCR typing and PCR-ribotyping for estimation of nosocomial transmission of *Clostridium difficile*. *J Hosp Infect,* 55**,** 124-30.

YAMAMOTO, K. R., ALBERTS, B. M., BENZINGER, R., LAWHORNE, L. & TREIBER, G. 1970. Rapid bacteriophage sedimentation in the presence of polyethylene glycol and its application to large-scale virus purification. *Virology,* 40**,** 734-44.

YAMOUDY, M., MIRLOHI, M., ISFAHANI, B. N., JALALI, M., ESFANDIARI, Z. & HOSSEINI, N. S. 2015. Isolation of toxigenic *Clostridium difficile* from ready-to-eat salads by multiplex polymerase chain reaction in Isfahan, Iran. *Advanced Biomedical Research,* 4**,** 87.

YUTIN, N. & GALPERIN, M. Y. 2013. A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced *clostridia*. *Environ Microbiol,* 15**,** 2631-41.

ZARB, P., COIGNARD, B., GRISKEVICIENE, J., MULLER, A., VANKERCKHOVEN, V., WEIST, K., GOOSSENS, M., VAERENBERG, S., HOPKINS, S. & CATRY, B. 2012. The European Centre for Disease Prevention and Control (ECDC) pilot point prevalence survey of healthcare-associated infections and antimicrobial use. *Euro Surveill,* 17**,** 20316.

ZHOU, Y., BURNHAM, C. A., HINK, T., CHEN, L., SHAIKH, N., WOLLAM, A., SODERGREN, E., WEINSTOCK, G. M., TARR, P. I. & DUBBERKE, E. R. 2014. Phenotypic and genotypic analysis of *Clostridium difficile* isolates: a single-center study. *J Clin Microbiol,* 52**,** 4260-6.

ZIDARIC, V., BEIGOT, S., LAPAJNE, S. & RUPNIK, M. 2010. The occurrence and high diversity of *Clostridium difficile* genotypes in rivers. *Anaerobe,* 16**,** 371-375.