

UNIVERSITÉ DE SHERBROOKE

**ÉTUDE DE L'IMPACT DES PROPHAGES SUR LA BIOLOGIE DE
*CLOSTRIDIUM DIFFICILE***

Par
Ognjen SEKULOVIC

Département de Microbiologie et Infectiologie
Laboratoire de Louis-Charles Fortier

Mémoire présenté à la Faculté de médecine et des sciences de la santé en vue de
l'obtention du grade de Maître ès sciences (M.Sc) en microbiologie

Mai 2010

Évaluateurs (Université de Sherbrooke) :

Dr Louis-Charles Fortier, Microbiologie et Infectiologie

Dr Raymund Wellinger, Microbiologie et Infectiologie

Dr Vincent Burrus, Biologie

Directeur de recherche

Membre de jury interne

Membre de jury externe



Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file *Votre référence*
ISBN: 978-0-494-79771-6
Our file *Notre référence*
ISBN: 978-0-494-79771-6

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

TABLE DES MATIÈRES

TABLE DES MATIÈRES	ii
LISTE DES FIGURES.....	v
INTRODUCTION	v
ARTICLE	vi
LISTE DES TABLEAUX	vii
ARTICLE	vii
LISTE DES ABREVIATIONS.....	viii
RÉSUMÉ.....	xi
1. INTRODUCTION.....	1
1.1 <i>Clostridium difficile</i>	1
1.2 Facteurs de virulence de <i>C. difficile</i>	3
1.2.1 Mode d'action des toxines A et B	5
1.3 Facteurs de virulence secondaires chez <i>C. difficile</i>	6
1.4 Souche hypervirulente BI/NAP1/027	10
1.5 Caractéristiques de la souche hypervirulente BI/NAP1/027	11
1.6 Les bactériophages.....	14
1.7 Impact des bactériophages sur la biologie bactérienne.....	17
1.7.1 Facteurs de virulence encodés par les bactériophages	18
1.7.2 Inactivation des gènes bactériens suite à l'intégration des prophages ..	22
1.7.3 Immunité de l'hôte face aux phages du même groupe	23
1.7.4 Réarrangements chromosomiques via la recombinaison homologue ...	25
1.8 Phages de <i>C. difficile</i>	27
1.9 Les objectifs de la présente étude	31
2. ARTICLE.....	33
2.1 Avant propos.....	33
2.2 Résumé de l'article	35

Prophage-stimulated toxin production in <i>Clostridium difficile</i> NAP1/027 lysogens	36
ABSTRACT.....	37
INTRODUCTION	38
MATERIALS AND METHODS.....	40
Bacteria and growth conditions.....	40
Bacterial DNA extraction and PCR ribotyping.....	40
Prophage induction and phage propagation	41
Transmission electron microscopy (TEM).....	41
Host range determination and one step growth curve assays.....	42
Analysis of structural proteins by SDS-PAGE and mass spectrometry.....	42
Phage DNA purification and restriction analysis.....	43
Phage genome sequencing and bioinformatics analysis	43
Isolation of lysogens	43
Detection of toxins A and B in culture supernatants.....	44
RNA extraction and gene expression analysis	45
Nucleotide sequence accession number	46
RESULTS	46
Phage isolation	46
Host range and lytic growth cycle.....	47
Genome sequence.....	48
DNA homology and other similar prophages	48
Gene products and annotation.....	49
Structural proteins	55
Lysis module	56
DNA replication, transcription and gene regulation	59
Lysogeny control.....	60
Putative lysogenic conversion genes.....	61
Prophage-stimulated toxin production in ϕ CD38-2 lysogens.....	63
DISCUSSION.....	67
First genome sequence of a <i>pac</i> -type <i>Siphoviridae</i> phage infecting <i>C. difficile</i> ..	69

Putative lysogenic conversion genes.....	70
Prophage-stimulated toxin production in NAP/027 lysogens.....	72
Conclusion.....	74
ACKNOWLEDGMENTS	74
REFERENCES	75
3. DISCUSSION	80
3.1 Rappel des objectifs du projet.....	80
3.2 Isolement de phages tempérés de <i>Clostridium difficile</i>	81
3.3 Caractérisation microbiologique et moléculaire du phage ϕ CD38-2	82
3.3.1 Spectre d'hôte et dynamique d'infection du phage ϕ CD38-2.....	83
3.3.2 Séquençage et comparaisons génomiques du phage ϕ CD38-2	86
3.3.3 Analyse globale du génome par bioinformatique.....	88
3.3.4 Analyse détaillée des différents modules	89
3.4 Caractérisation de l'impact du phage ϕ CD38-2 sur la biologie bactérienne	94
3.4.1 Impact du phage sur la croissance bactérienne.....	95
3.4.2 Impact du phage sur le locus de pathogénicité.....	95
4. REMERCIEMENTS	101
5. BIBLIOGRAPHIE.....	102

LISTE DES FIGURES

INTRODUCTION

Figure 1	Représentation schématique du locus de pathogénicité de <i>C. difficile</i> (~19 kb)	4
Figure 2	Émergence de la souche hypervirulente	11
Figure 3	Cycle lysogénique	15
Figure 4	La polylysogénie parmi les espèces bactériennes	16
Figure 5	Cycle lytique	25
Figure 6	Représentation schématique d'une réorganisation chromosomique	26

LISTE DES FIGURES**ARTICLE**

Figure 1	Genetic organization of the complete ϕ CD38-2 genome (41,090 bp)	51
Figure 2	Analysis of ϕ CD38-2 structural proteins	56
Figure 3	Multiple protein alignment of phage endolysins	58
Figure 4	Relative G+C content of ϕ CD38-2	60
Figure 5	Growth and toxin production in the wild-type CD274 (open circles) and a CD274/ ϕ CD38-2 lysogen (black squares)	64
Figure 6	Immuno-dot blot detection of toxins in cleared supernatants from 24 h cultures of wild-type CD274 and a CD274/ ϕ CD38-2 lysogen	65
Figure 7	Relative levels of <i>tcdA</i> (gray bars) and <i>tcdB</i> (black bars) mRNA in ϕ CD38-2 lysogens versus their respective wild- type parental strain, as determined by qRT-PCR analysis	66

LISTE DES TABLEAUX**ARTICLE**

Table S1	Primers used for real-time qRT-PCR assays	46
Table 1	Host range analysis of ϕ CD38-2	47
Table S2	Putative coding sequences from ϕ CD38-2 and bioinformatics analyses of the deduced proteins	52
Table 2	Putative sigma70-like promoters identified in the genome of ϕ CD38-2	62

LISTE DES ABREVIATIONS

ADN	Acide désoxyribonucléique
ARN	Acide ribonucléique
ARNm	Acide ribonucléique messenger
ARNr	Acide ribonucléique ribosomal
ARNt	Acide ribonucléique de transfert
BHI	<i>Brain Heart Infusion</i>
BLAST	<i>Basic Local Alignment Search Tool</i>
CD	<i>Clostridium difficile</i>
CDAD	<i>Clostridium difficile associated disease</i>
cDNA	<i>Complementary deoxyribonucleic acid</i>
CDT	<i>Clostridium difficile Toxin</i>
CT	<i>Cholera Toxin</i>
dNTP	Déoxyribonucléotides
DO ₆₀₀	Densité Optique à la longueur d'onde 600 nanomètres
DT	<i>Diphtheria Toxin</i>
EDTA	Acide éthylène-diamine-tetraacétique
ELISA	<i>Enzyme-Linked ImmunoSorbent Assay</i>
HRP	<i>Horseradish Peroxydase</i>
ICD	Infections à <i>Clostridium difficile</i>
kDa	Kilodaltons
kV	Kilovolts
LC-MS/MS	<i>Liquid Chromatography-tandem Mass Spectrometry</i>

M	Molaire
mM	Millimolaires
MCP	<i>Major Capsid Protein</i>
MOI	<i>Multiplicity of Infection</i>
MSP	<i>Major Structural Protein</i>
MTP	<i>Major Tail Protein</i>
NAP	<i>North America Pulsed-field</i>
NCBI	<i>National Center for Biotechnology Information</i>
ng	Nanogrammes
ORF	<i>Open reading frame</i>
PaLoc	<i>Pathogenicity locus</i>
pb	Paires de base
PBS	<i>Phosphate-Buffered Saline</i>
PBST	<i>Phosphate-Buffered Saline + 1% Tween</i>
PCR	<i>Polymerase Chain Reaction</i>
PFGE	<i>Pulsed-field gel electrophoresis</i>
PFU/ml	<i>Plaque-forming units per millilitre</i>
pH	Potentiel hydrogen
qRT-PCR	<i>Quantitative Reverse-transcriptase Polymerase Chain Reaction</i>
RBP	<i>Receptor binding protein</i>
RPM	Rotation Par Minute
RT-PCR	<i>Reverse-transcriptase Polymerase Chain Reaction</i>
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SLP	<i>Surface Layer Protein</i>

SSB	<i>Single Strand Binding protein</i>
TCP	<i>Toxin Coregulated Pilus</i>
TAE	Tris-acétate EDTA
TEM	<i>Transmission Electron Microscopy</i>
TY	<i>Tryptose-Yeast extract</i>
UA	Uranyl-acétate
ufp/ml	Unité formant des plages par millilitre
w/v	Rapport poids/volume

RÉSUMÉ

Étude de l'impact des prophages sur la biologie de *Clostridium difficile*, Mémoire de maîtrise présenté par Ognjen Sekulovic, Département de microbiologie et infectiologie, Université de Sherbrooke, Mai 2010

La bactérie *Clostridium difficile* est maintenant considérée comme un pathogène majeur responsable d'infections nosocomiales en Amérique du Nord et en Europe. De plus, l'émergence de souches hypervirulentes, telle la souche NAP1/027 responsable de récentes épidémies, est un phénomène inquiétant. Un enjeu crucial au cours des prochaines années sera de mieux comprendre les mécanismes de virulence et d'évolution de *C. difficile*.

Les bactériophages (c.-à-d. des virus bactériens, ou phages) sont des joueurs clés dans l'évolution de la plupart des bactéries, pathogènes ou non. Les données concernant l'impact des phages sur *C. difficile* sont très limitées. Par contre, deux études récentes démontrent que les phages semblent influencer la virulence de *C. difficile* en altérant la production de toxines TcdA et TcdB. L'objectif de mes travaux de recherche est donc d'étudier au niveau microbiologique et moléculaire les phages de *C. difficile* et de démontrer leur impact sur l'évolution et la virulence de ce pathogène.

Dans ce contexte, plusieurs phages ont été induits à partir d'isolats cliniques de *C. difficile*. Dans cette collection, un phage en particulier, le ϕ CD38-2, a été choisi pour la caractérisation subséquente dû à sa divergence génomique par rapport aux autres phages et à sa capacité d'infecter la grande majorité des souches ayant le ribotype hypervirulent (027). Ces caractéristiques uniques ont justifié le séquençage complet de son génome. Par contre, aucun facteur de virulence évident n'a été identifié. À l'opposé, une analyse bio-informatique a permis l'identification d'une région spécifique comportant plusieurs gènes de conversion lysogénique potentiels. L'impact de ces gènes sur la virulence bactérienne reste à être déterminé.

De plus, lorsqu'on introduit le phage ϕ CD38-2 dans la souche sensible CD274, on observe une accumulation plus rapide et plus grande des toxines après 48h dans le surnageant de la culture. Ce phénomène a été confirmé avec des tests ELISA sur des répliques biologiques indépendants ainsi que par un immunodosage avec anticorps spécifiques aux deux toxines. Par ailleurs, une étude transcriptionnelle par PCR en temps réel a permis de constater que le phage ϕ CD38-2 influence également l'expression des gènes *tcdA* et *tcdB* dans le temps. Par contre, l'effet du phage ϕ CD38-2 est variable lorsqu'on l'introduit dans d'autres souches de *C. difficile*.

Donc, les résultats de nos travaux indiquent que certains phages auraient un impact sur la virulence de *C. difficile* en altérant la production et la transcription des gènes de toxines. Nos données laissent toutefois sous-entendre que cet effet peut varier selon les souches de *C. difficile*.

MOTS CLÉS : *Clostridium difficile*, NAP1/027, bactériophage, *Siphoviridae*, toxine

1. INTRODUCTION

1.1 *Clostridium difficile*

De nos jours, les maladies infectieuses représentent une des premières causes de mortalité dans le monde (EVANS et LERBERGHE, 2008). Elles sont causées par des microorganismes parasitaires, c'est-à-dire nuisibles pour la santé de l'homme. L'avènement de la médecine moderne ainsi que la découverte des antibiotiques au début du XXe siècle ont donné l'espoir à une éradication complète des maladies infectieuses causées par les bactéries. Malheureusement, les bactéries, tout comme n'importe quel organisme vivant, sont capables d'évoluer de manière à contourner les contraintes imposées et ainsi permettre la survie et la propagation de l'espèce. Les premiers succès des traitements aux antibiotiques ont incité un usage massif de ces derniers qui, à leur tour, ont provoqué l'apparition des bactéries résistantes (CHEN *et al.*, 2009; HUANG *et al.*, 2009; LISTER *et al.*, 2009; DRAWZ et BONOMO, 2010). Le phénomène de la résistance est très inquiétant, mais il n'est pas le seul problème relié à l'usage massif des antibiotiques. De par leur manque de spécificité et d'un usage mal adapté, il arrive qu'un traitement fasse plus de mal que de bien. Par exemple, un traitement avec des antibiotiques à large spectre peut cibler les microorganismes responsables d'une infection, mais aussi détruire en même temps les bactéries bénéfiques pour l'hôte. Dans cette éventualité, les bactéries favorables peuvent être remplacées par les pathogènes opportunistes qui vont initier une infection et causer des dommages à leur tour.

Clostridium difficile est un bon exemple de pathogène opportuniste. Cette bactérie est un pathogène entérique, anaérobie strict et capable de former des spores. L'infection par cette bactérie se fait généralement lorsque la flore intestinale bactérienne normale est altérée par la prise d'antibiotiques dans le contexte de traitement d'une infection quelconque. Cette altération de la flore intestinale permet à *C. difficile* d'établir une infection qui implique une colonisation massive de l'épithélium entérique par le pathogène. Les manifestations cliniques d'une telle infection se présentent sous forme de douleurs abdominales et diarrhées jusqu'aux symptômes plus sévères tels que les pseudomembranes et le mégacôlon toxique. Le traitement actuel se base principalement sur l'utilisation de deux antibiotiques, soit la métronidazole et la vancomycine (ASLAM *et al.*, 2005; GERDING *et al.*, 2008). Cependant, selon la sévérité de l'infection, des traitements supplémentaires tels que la normalisation d'électrolytes, remplacement des fluides, injection intraveineuse d'immunoglobulines ou encore injection intra colique de vancomycine peuvent s'avérer nécessaires (HALSEY, 2008).

Malgré son caractère anaérobique, *C. difficile* est très persistant dans l'environnement grâce à sa capacité à former des spores. En étant très résistantes aux conditions hostiles, il n'est pas surprenant de trouver des spores de *C. difficile* dans les milieux naturels (sol, rivières), domestiques (salles de bain, surfaces) ainsi que certains produits de consommation (viande, légumes) (BRAZIER, 1998; RODRIGUEZ-PALACIOS *et al.*, 2007). Les hôpitaux et les établissements de soins de longue durée ne font pas exception, et c'est ainsi qu'on les considère comme des réservoirs d'infections (JOHNSON et GERDING, 1998).

La virulence de *C. difficile* va varier en fonction des souches c'est-à-dire de la présence ou absence de facteurs de virulence encodés par la bactérie. Ainsi, on retrouve des souches virulentes qui seront très problématiques et d'autres qui ne seront pas nuisibles. Aujourd'hui, d'une manière générale, *C. difficile* est perçu comme la principale cause des diarrhées associées aux antibiotiques chez les patients hospitalisés.

1.2 Facteurs de virulence de *C. difficile*

Le pouvoir pathogène de *C. difficile* provient essentiellement de deux facteurs de virulence principaux : toxines A (TcdA) et B (TcdB) (GENTH *et al.*, 2008). Ces deux toxines sont produites par tous les isolats virulents, mais certaines souches peuvent produire l'une ou l'autre des deux toxines (VOTH et BALLARD, 2005). Récemment, il a été démontré que la toxine B joue un rôle essentiel dans la virulence de *C. difficile*, du moins dans un modèle murin (LYRAS *et al.*, 2009). Les gènes des deux toxines (*tcdA* et *tcdB*) se trouvent à une place bien précise du chromosome bactérien appelé locus de pathogénicité (PaLoc). Sur ce dernier, on retrouve également un gène codant pour un régulateur positif (TcdR) et négatif (TcdC) qui contrôlent l'expression des toxines dans le temps. Finalement, un gène codant pour une protéine (TcdE) est situé entre *tcdA* et *tcdB* (Fig.1). Les toxines de *C. difficile* ne possédant pas de peptide-signal permettant leur sécrétion, il est à supposer

que TcdE y joue un rôle au niveau de la perméabilité membranaire, ou du moins facilite la sécrétion (TAN *et al.*, 2001).

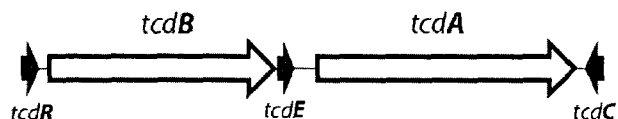


Figure 1 Représentation schématique du locus de pathogénicité de *C. difficile* (~19 kb)

Le régulateur positif TcdR est un facteur σ alternatif qui régule positivement l'expression des gènes de toxines ainsi que l'expression de son propre gène (DUPUY *et al.*, 2008). L'expression des toxines TcdA et TcdB ainsi que le régulateur positif TcdR se fait lorsque la bactérie atteint la phase stationnaire. À l'opposé, le régulateur négatif TcdC est exprimé durant la phase exponentielle. Le produit de ce gène interfère avec le complexe TcdR-ARN polymérase afin d'empêcher la reconnaissance des promoteurs associés aux gènes de toxines (DUPUY *et al.*, 2008). Le gène *tcdE* encode une protéine qui présente une forte homologie avec les holines, protéines de phages impliquées dans la lyse bactérienne suite à un cycle lytique. Les analyses bio-informatiques ont montré que la protéine TcdE partage une similarité en séquence d'acides aminés avec plusieurs holines de phages provenant de différentes espèces bactériennes (*Bacillus subtilis*, *Streptococcus thermophilus*, *S. pneumoniae*, etc.). Également, une structure similaire est observée lorsqu'on compare une holine avec la protéine TcdE en fonction des régions hydrophiles et hydrophobes. Finalement, la surproduction de TcdE à partir d'un vecteur d'expression chez

Escherichia coli, a comme conséquence de diminuer la turbidité de la culture, confirmant ainsi son rôle dans la lyse bactérienne (TAN *et al.*, 2001).

1.2.1 Mode d'action des toxines A et B

De par leur activité spécifique, les toxines A et B ont été classées dans la famille des *Clostridial Glucosylating Toxins* (AKTORIES et JUST, 2005). La synthèse des toxines est dépendante de la phase de croissance bactérienne ainsi que divers signaux externes, incluant la température, la présence de glucose ou certains acides aminés, l'acide butyrique et certains antibiotiques (VOTH et BALLARD, 2005; EMERSON *et al.*, 2008; GERBER *et al.*, 2008).

Tel que discuté plus haut, les toxines sont probablement sécrétées avec l'aide de la protéine TcdE, encodée sur le PaLoc bactérien. La liaison des toxines à un récepteur cellulaire spécifique sur les cellules épithéliales intestinales de l'hôte (différent pour chaque toxine) provoque l'endocytose de ces dernières. Dans la cellule, l'endosome est acidifié ce qui provoque un changement de conformation de la toxine menant à l'autoclivage et au relâchement du domaine actif. Ce dernier, libéré dans le cytoplasme, va effectuer une monoglucosylation des GTPases de la famille des protéines Rho, qui sont impliquées dans la régulation de la synthèse du cytosquelette d'actine. Le résultat, qu'on appelle effet cytopathique, se présente sous forme de la perte d'intégrité et l'arrondissement cellulaire causé par l'apoptose des cellules (GENTH *et al.*, 2008). Récemment, il a été démontré que la toxine A induit la

mort cellulaire par la voie intrinsèque (voie mitochondriale d'apoptose), qui est indépendante des récepteurs de mort cellulaire (*death receptor-independent*) et de la cascade des caspases (MATTE *et al.*, 2009).

Également, l'implication des toxines A et B dans l'induction de la réponse inflammatoire a été étudiée récemment. Ainsi, une stimulation TcdA-TcdB - dépendante des mastocytes a été observée suivi d'un relâchement des cytokines pro-inflammatoires, notamment l'interleukine-8 (IL 8) (WERSHIL *et al.*, 1998; MEYER *et al.*, 2007). D'autres médiateurs d'inflammation, induits par la présence de TcdB, ont été décrits, dont TNF- α et γ -interféron. La contribution de ces derniers dans l'apparition des symptômes résultants d'une infection à *C. difficile* a été mise en évidence (ROCHA *et al.*, 1997; ISHIDA *et al.*, 2004).

1.3 Facteurs de virulence secondaires chez *C. difficile*

Mis à part les toxines A et B qui constituent les principaux facteurs de virulence, on dénombre plusieurs autres gènes qui semblent impliqués à diverses étapes de la pathogenèse.

L'étape préliminaire d'une infection implique une adhésion efficace des bactéries aux cellules épithéliales. Chez *C. difficile*, cette étape se fait via les protéines de surface cellulaire (SLPs, *surface layer proteins*) (CALABI *et al.*, 2001; KARJALAINEN *et al.*, 2001). Un précurseur commun est exprimé par le gène *slpA*

avant de subir des modifications post-traductionnelles et ainsi donner naissance à deux SPLs matures, une de masse moléculaire élevée et l'autre de faible masse (CALABI *et al.*, 2001). D'autres études ont mis en évidence le rôle spécifique des SPLs de masse moléculaire élevée dans l'adhésion bactérienne aux cellules épithéliales humaines (CALABI *et al.*, 2002). Également, il a été démontré que les flagelles de *C. difficile*, en plus de jouer un rôle dans la motilité, sont impliqués dans l'adhésion (TASTEYRE *et al.*, 2001). De plus, l'implication de certaines protéines de surface (e.g. Cwp66) dans l'adhésion a été mise en évidence (WALIGORA *et al.*, 2001).

Ainsi, les SPLs, les flagelles et certaines protéines de surface cellulaires constituent des facteurs importants dans la virulence de *C. difficile*.

Un autre déterminant probablement important dans la pathogenèse bactérienne est la résistance à la phagocytose. Cette résistance provient généralement de la présence d'une capsule de polysaccharides qui entoure les bactéries. La capsule ne joue pas un rôle vital pour la bactérie et en l'absence de celle-ci, la division bactérienne n'est pas affectée. Par contre, la capsule constitue un facteur important dans le pouvoir pathogène, car elle protège la bactérie du système immunitaire de l'hôte et empêche la phagocytose ainsi que l'activation de la voie alterne du système du complément. *C. difficile* possède une capsule qui peut varier en épaisseur, mais qui ne semble pas influencer la virulence *in vitro* (BORRIELLO *et al.*, 1990; DAVIES et BORRIELLO, 1990). L'implication biologique de la présence d'une capsule autour de *C. difficile* n'est pas encore connue, mais on peut supposer qu'elle va jouer un rôle important dans la virulence comme c'est le cas chez d'autres bactéries pathogènes

telles que *S. pneumoniae*, *Neisseria meningitidis* et *Pseudomonas aeruginosa* (AHMED *et al.*, 1993; HAMMERSCHMIDT *et al.*, 1996; VARVIO *et al.*, 2009).

Plusieurs enzymes hydrolytiques et protéolytiques sont exprimés à divers stades de la pathogenèse et sont soupçonnés de contribuer à la virulence de *Clostridium difficile*. De telle manière, des enzymes ayant une activité de sulphatase, hyaluronidase, collagénase, héparinase et autres ont été identifiés chez *C. difficile*. D'une part, ces différentes enzymes peuvent être impliquées dans la destruction du tissu de l'hôte lors de l'invasion bactérienne, ce qui peut faciliter la colonisation et l'établissement d'une infection. D'autre part, en dégradant certaines structures de l'épithélium de l'hôte, ces enzymes hydrolytiques peuvent relâcher des sucres et autres éléments nutritifs que *C. difficile* est capable d'assimiler (SEDDON *et al.*, 1990). Plus spécifiquement, une protéine de surface, la protéase Cwp84, a été caractérisée dernièrement chez *C. difficile* (JANOIR *et al.*, 2007). Il s'agit d'une protéine associée à la surface cellulaire, impliquée dans la dégradation des principaux constituants de la matrice extracellulaire (fibronectine, vitronectine et laminine) chez l'hôte. Cette activité de dégradation peut provoquer des lésions sévères au niveau de l'épithélium entérique et entraîner des inflammations et autres complications, tels que les pseudomembranes.

Le dernier déterminant impliqué dans la pathogenèse de *C. difficile* et pour lequel le rôle n'a pas été étudié de manière exhaustive, est la toxine binaire CDT. À l'opposé des toxines A et B qui sont impliquées dans la glucosylation des GTPases de la famille Rho, la toxine binaire exhibe une activité d'actine-ADP-ribosyltransférase (GONCALVES *et al.*, 2004). Cette activité enzymatique a comme

conséquence d'inhiber la polymérisation des monomères d'actine. La prévalence de cette toxine binaire est estimée à environ 6% parmi les souches de *C. difficile* (GONCALVES *et al.*, 2004; GERIC *et al.*, 2006). Plusieurs autres souches pathogènes possèdent également une actine-ADP-ribosyltransférase, comme c'est le cas chez *C. botulinum* (toxine C2), *C. perfringens* (toxine iota), *C. spiroforme* (toxine CST) et *B. cereus* (*vegetative insecticidal proteins*, VIP) (BARTH, 2004). Une étude récente a démontré que la toxine binaire CDT de *Clostridium difficile* induit la formation de projections membranaires sur les cellules épithéliales grâce à une activité de dépolymérisation de microfilaments cellulaires. Ceci a comme conséquence la formation d'un treillis dense à la surface cellulaire qui piège les bactéries d'une manière non spécifique et augmente l'adhérence bactérienne à l'épithélium entérique (SCHWAN *et al.*, 2009).

Malgré tous les facteurs de virulence mentionnés plus haut, *C. difficile* n'était pas un pathogène particulièrement problématique avant 2001, l'année où on a assisté à l'émergence d'une souche hypervirulente.

1.4 Souche hypervirulente BI/NAP1/027

Un changement épidémiologique inattendu a été observé concernant l'incidence des infections à *C. difficile* vers 2001, en Amérique du Nord. Plusieurs études ont rapporté des statistiques alarmantes quant à l'augmentation brusque du nombre de cas de *C. difficile* de même qu'une sévérité accrue des infections et une mortalité plus élevée parmi les patients atteints d'une infection à *C. difficile* (PEPIN *et al.*, 2004; KUIJPER *et al.*, 2006; KUIJPER *et al.*, 2007; O'CONNOR *et al.*, 2009).

Différentes techniques de typage ont permis de constater que c'est l'émergence d'un clone unique qui est associée aux récentes épidémies (LOO *et al.*, 2005; MCDONALD *et al.*, 2005). Ce clone hypervirulent porte la signature BI/NAP1/027 qui réfère au profil de restriction d'enzyme (REA) BI, le profil en champ pulsé (PFGE) de l'Amérique du Nord NAP1 et le ribotype PCR 027. Certains hôpitaux de la région de Montréal (Québec) font mention d'une prévalence de plus de 80% d'isolats hypervirulents durant la période épidémique (Figure 2), et cette situation était très similaire dans d'autres hôpitaux en Amérique du Nord (KUIJPER *et al.*, 2006; EGGERTSON, 2007; WEISS *et al.*, 2007).

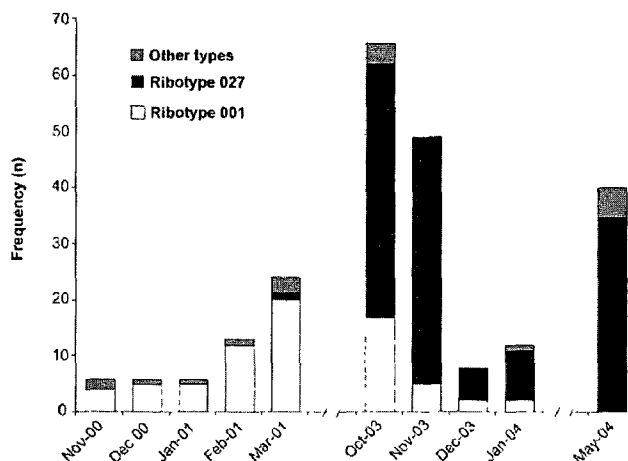


Figure 2 Émergence de la souche hypervirulente. Prévalence des ribotypes PCR parmi les isolats cliniques à l’hôpital Maisonneuve-Rosemont de Montréal pendant la période épidémique (LABBE *et al.*, 2008).

1.5 Caractéristiques de la souche hypervirulente BI/NAP1/027

Les raisons précises de l’émergence de cette souche hypervirulente ne sont pas très bien connues et font encore l’objet de nombreuses études en cours (SPIGAGLIA *et al.*, 2001; LEMEE *et al.*, 2004; STABLER *et al.*, 2006; DAWSON *et al.*, 2009; STABLER *et al.*, 2009; HE *et al.*, 2010). Tel que suggéré dans plusieurs publications, l’utilisation excessive des fluoroquinolones a certainement eu un impact sur la dissémination et l’émergence du clone hypervirulent qui possède une résistance accrue à cet antibiotique. La résistance contre les fluoroquinolones, antibiotiques utilisés largement durant la période pré-épidémique, est généralement acquise par des

mutations dans les gènes bactériens *gyrA* et *gyrB*. Ces derniers encodent les deux sous-unités de l'ADN gyrase, l'enzyme responsable du maintien du taux de superenroulement de l'ADN. Les souches isolées avant les épidémies ne présentent pas de résistance aux fluoroquinolones, ce qui laisse à supposer une acquisition de résistance récente (RAZAVI *et al.*, 2007).

Des tolérances accrues à d'autres antibiotiques ont été rapportées chez *C. difficile* notamment au métronidazole (PELAEZ *et al.*, 2008), érythromycine et clindamycine (SCHMIDT *et al.*, 2007), rifampine (CURRY *et al.*, 2009) ainsi que la tétracycline et chloramphénicol (SPIGAGLIA et MASTRANTONIO, 2004). L'acquisition des résistances aux antibiotiques peut avoir comme effet l'augmentation de la propagation de *C. difficile* en milieu hospitalier (SAMORE *et al.*, 2006).

La virulence apparemment élevée de la souche hypervirulente pourrait s'expliquer, en partie, par une surproduction des toxines A et B. Il a été montré que ces souches produisent de 16 à 23 fois plus de toxines A et B respectivement, *in vitro* (WARNY *et al.*, 2005). Une délétion d'une paire de base en position 117 dans le gène répresseur *tcdC* est soupçonnée d'être à l'origine de cette hyperproduction. La délétion en question provoque un changement du cadre de lecture dans le gène *tcdC*, ce qui a pour effet d'introduire un codon d'arrêt prématuré menant à la production d'une protéine sévèrement tronquée qui serait moins apte à réprimer l'expression des gènes de toxines, d'où la surproduction (DUPUY *et al.*, 2006; CURRY *et al.*, 2007; MATAMOUROS *et al.*, 2007; DUPUY *et al.*, 2008; MURRAY *et al.*, 2009).

Également, les souches épidémiques comme la NAP1/027 semblent posséder une plus grande capacité de sporulation (WILCOX et FAWLEY, 2000; AKERLUND *et al.*, 2008; SAXTON *et al.*, 2009) ce qui faciliterait la dissémination et la survie de ces dernières dans l'environnement hospitalier. À l'opposé, tous les isolats hypervirulents possèdent la toxine binaire CDT, qui semble impliquée dans l'adhésion des bactéries aux cellules épithéliales. Par contre, les implications biologiques de la présence de cette toxine binaire ne sont pas claires et demandent davantage d'études.

À la lumière de ces aspects qui caractérisent la souche hypervirulente, il est difficile de se prononcer sur les raisons précises du changement épidémiologique observé. Certes, dans son ensemble, l'évolution bactérienne est soumise à différents facteurs qui forcent un microorganisme à s'adapter aux conditions imposées. Il peut s'agir d'une adaptation aux changements d'environnement bactérien, d'une niche écologique particulière ou encore d'une adaptation aux facteurs externes tels que la pression de sélection des antibiotiques. Dans le cas de *C. difficile*, l'évolution s'est fait brusquement avec l'émergence d'un clone hypervirulent qui a pu prendre le dessus sur la majorité des souches qui étaient communes à cette époque. Une plus grande production de toxines par certains isolats *in vitro* ne peut pas expliquer entièrement les changements épidémiologiques observés. La régulation de la production des toxines est un processus fort complexe qui ne repose pas exclusivement sur les effets des régulateurs TcdC et TcdR (MURRAY *et al.*, 2009; VERDOORN *et al.*, 2010). Nous sommes donc en mesure de soupçonner l'existence d'autres facteurs qui ont pu influencer l'évolution rapide de *C. difficile*.

1.6 Les bactériophages

Le séquençage du premier génome de la souche CD630 de *C. difficile* (SEBAIHIA *et al.*, 2006) a apporté une multitude de données génétiques qui ont servi de base pour les études moléculaires sur cet organisme. De plus, l'analyse du génome de *C. difficile* a mis en évidence un contenu élevé en éléments génétiques mobiles, qui représentent environ 11% de la totalité du génome (SEBAIHIA *et al.*, 2006). Parmi ces éléments, on dénombre plusieurs transposons conjugatifs, transposons mobilisables, IStons ainsi que deux prophages inductibles. La présence de prophages dans le génome de *C. difficile* n'est pas surprenante puisque des prophages ont été identifiés dans le génome de la majorité des bactéries, incluant de nombreux pathogènes d'importance. Avec l'arrivée des techniques de séquençage à haut débit (de type 454), le nombre de génomes bactériens séquencés a augmenté rapidement. L'analyse de ces séquences a permis de constater que la distribution des prophages au sein des espèces bactériennes est beaucoup plus grande que ce qui était soupçonné (OHNISHI *et al.*, 2001; BANKS *et al.*, 2002; CANCHAYA *et al.*, 2003; CASJENS, 2003).

Les phages, parasites bactériens obligatoires, sont ubiquitaires et représentent l'entité biologique la plus répandue dans la biosphère (SKURNIK et STRAUCH, 2006; HERMOSO *et al.*, 2007). Généralement, on différencie deux classes majeures de phages, une distinction basée sur leur mode de répllication. De telle manière, on retrouve les phages strictement lytiques, qui vont invariablement détruire leur hôte bactérien suite à l'infection. À l'opposé, les phages dits tempérés

peuvent suivre la voie lytique, mais aussi adopter le mode de vie lysogénique. Ce dernier leur permet, suite à l'infection, d'intégrer leur génome dans le chromosome bactérien ou de demeurer latent sous forme épisomale dans le cytoplasme, épargnant ainsi l'hôte et leur permettant de se répliquer en même temps que la bactérie se divise (Figure 3).

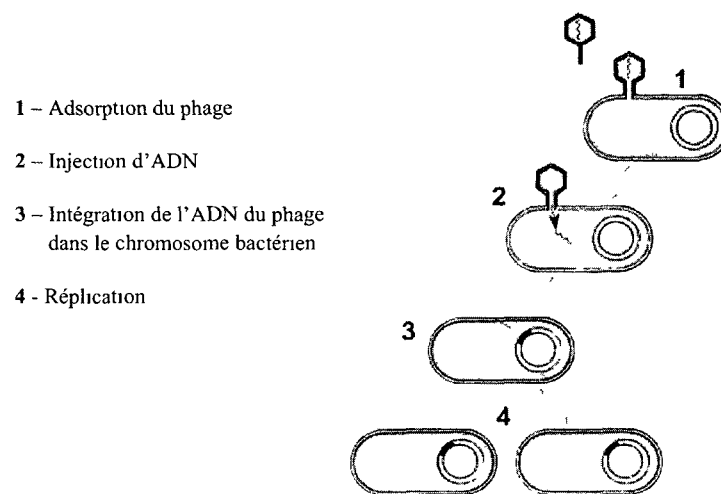


Figure 3 Cycle lysogénique. Représentation schématique de la lysogénéisation d'une souche bactérienne par un phage tempéré.

Certaines souches d'*E. coli* possèdent jusqu'à 18 prophages dans leur génome, ce qui peut représenter jusqu'à 16% d'ADN génomique total (HAYASHI *et al.*, 2001; OHNISHI *et al.*, 2001; ASADULGHANI *et al.*, 2009). Une situation

similaire est observée chez *Streptococcus pyogenes* où on dénombre de 4 à 6 prophages selon les souches ce qui représente jusqu'à 12% d'ADN génomique total (Figure 4) (FERRETTI *et al.*, 2004; AZIZ *et al.*, 2005).

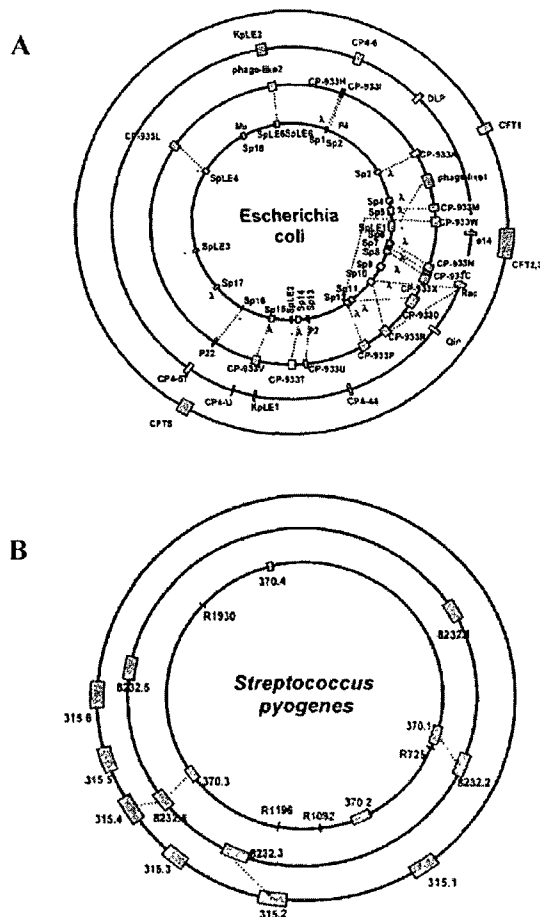


Figure 4 La polylysogénie parmi les espèces bactériennes. Représentation schématique des génomes séquencés alignés (cercles concentriques) d'*E. coli* en A) et de *S. pyogenes* (sérotypes M1, M18 et M3) en B). Les prophages identifiés sur chaque génome sont représentés par des boîtes grises. (CANCHAYA *et al.*, 2003)

Une telle quantité d'ADN étranger ne peut pas passer sans impact sur le génome bactérien. Dans la littérature, on retrouve de nombreux exemples où les phages ont influencé la biologie, l'évolution et la virulence des espèces bactériennes, pathogènes ou non (BRUSSOW *et al.*, 2004). La revue de certains de ces exemples est présentée plus loin dans ce travail. Il va sans dire que ce sont surtout (mais non exclusivement) les phages tempérés qui sont à l'origine de cette influence due à leur capacité d'intégrer les génomes de leurs bactéries hôtes.

1.7 Impact des bactériophages sur la biologie bactérienne

Un élément majeur dans l'émergence de nouveaux isolats pathogéniques et l'évolution bactérienne est l'acquisition d'ADN étranger par un transfert horizontal (CANCHAYA *et al.*, 2003). Très souvent, le matériel génétique ainsi acquis va fournir un avantage sélectif à la bactérie réceptrice ce qui peut avoir des conséquences phénotypiques importantes. De nombreux éléments génétiques mobiles (plasmides, transposons, IStrons, îlots de pathogénicité) ont été associés à l'acquisition de gènes de virulence via un transfert horizontal, mais les bactériophages demeurent le groupe ayant le plus d'impact (BOYD et BRUSSOW, 2002; DAVIS *et al.*, 2002). Par ailleurs, dans la communauté scientifique, les bactériophages sont considérés comme des véritables moteurs évolutifs à court terme (CASJENS, 2003; CANCHAYA *et al.*, 2004).

Les moyens par lesquels les bactériophages peuvent influencer la biologie bactérienne sont nombreux. Il peut s'agir de gènes de virulences encodés directement par les phages et exprimés en mode lysogénique, c.-à-d. lorsque le phage est latent et intégré dans le chromosome bactérien. L'expression des gènes de toxines peut être coréglé avec le processus d'induction de prophages, comme c'est le cas de toxines Shiga chez *E. coli*. D'autres moyens sont plus subtils, mais peuvent toutefois conduire à d'importants changements phénotypiques chez la bactérie hôte. Dans les prochaines sections, une vue d'ensemble sur les différents moyens qu'ont les phages d'influencer la biologie bactérienne sera exposée.

1.7.1 Facteurs de virulence encodés par les bactériophages

Les bactériophages peuvent transporter des gènes codant pour des facteurs de virulence directs et ainsi, suite à la lysogénisation, influencer la virulence de leur bactérie hôte. Le processus par lequel du matériel génétique est acquis par l'intermédiaire des bactériophages s'appelle la transduction. Lorsque ce matériel génétique acquis par transduction est en mesure de modifier le phénotype bactérien, on parle de conversion lysogénique. Ce phénomène sera décrits dans les sections suivantes à travers plusieurs exemples concrets touchant différentes espèces bactériennes.

Un des premiers exemples de la conversion lysogénique a été démontré avec *Corynebacterium diphtheriae*, bactérie Gram +, adaptée à la pathogénèse chez l'homme. Le facteur de virulence principal est la *diphtheria toxin* (DT), capable de détruire les parenchymes du cœur, foie et autres organes vitaux. Cette toxine est encodée par une série de bactériophages tempérés infectant cette bactérie. La première évidence de la conversion d'une souche non-toxinogène en une souche toxinogène, suite à l'infection par le phage β encodant la toxine, a été rapportée au début des années 1950 (FREEMAN, 1951; GROMAN, 1956).

C. diphtheriae est loin d'être le seul exemple où un bactériophage est responsable du caractère virulent d'une bactérie. Un cas fascinant est celui de *Vibrio cholerae*. Ce pathogène entérique est l'agent responsable du Cholera, maladie épidémique contagieuse caractérisée par des symptômes de diarrhées abondantes qui entraînent une déshydratation sévère, insuffisance rénale et éventuellement la mort. Le facteur de virulence principal est la *Cholera toxin* (CT), encodée par le phage filamenteux CTX Φ (WALDOR et MEKALANOS, 1996). Un autre facteur de virulence crucial chez *V. cholerae* est le facteur de colonisation TCP (*toxin coregulated pilus*). Ce dernier, présent sur un locus de pathogénicité (VPI), encode les gènes nécessaires à la formation d'un pilus impliqué dans l'étape initiale de la colonisation de l'intestin (MERRELL *et al.*, 2002). Sur le même locus, on retrouve d'autres gènes impliqués dans l'adhésion ainsi que plusieurs régulateurs transcriptionnels. En 1999, l'évidence sur l'origine phagique du locus VPI a été apportée (KARAOLIS *et al.*, 1999), le renommant ainsi en phage filamenteux VPI Φ . Donc, chez *V. cholerae*, le gène crucial impliqué dans l'adhésion des bactéries aux

entérocytes (TCP) et la toxine responsable des symptômes sévères de la maladie (CT) sont tous les deux encodées par des phages tempérés. L'acquisition de ces deux phages par une souche bénigne de *V. cholerae* va résulter en une conversion lysogénique et l'évolution vers une souche pathogène (BRUSSOW *et al.*, 2004).

Un autre exemple bien connu est celui des toxines Shiga chez *Escherichia coli* O157:H7. Ce sérotype particulier se différencie des autres souches bénignes d'*E. coli* retrouvées dans la flore commensale par son caractère pathogène. *E. coli* O157:H7 est un pathogène intestinal qui produit plusieurs facteurs de virulence dont les principaux sont les toxines Shiga Stx1 et Stx2, toxines de type AB, capables d'inhiber la synthèse protéique. Leur mode d'action se résume au passage de la barrière intestinale et son relâchement dans les vaisseaux sanguins suivi des dommages aux cellules vasculaires du colon, des reins et du système nerveux. La manifestation clinique est représentée principalement par des diarrhées hémorragiques et autres complications plus sévères. Le pouvoir pathogène de ce sérotype provient de l'acquisition relativement récente de prophages encodant deux toxines *stx1* et *stx2*. Les toxines sont maturées dans le périplasme et relâchées via la lyse bactérienne. La régulation de l'expression des toxines peut être assez complexe et impliquer des facteurs bactériens ou phagiques. Il est intéressant de constater qu'une régulation fer-dépendante semble exister chez certains phages encodant la toxine *stx1*. Brièvement, en cas de carence en fer, une certaine proportion des lysogènes va produire la toxine *stx1* qui aura comme conséquence de provoquer l'hémorragie intestinale et ainsi permettre une acquisition du fer via les globules rouges. Le fer rétablit la croissance bactérienne et par conséquent inhibe la production ultérieure de la toxine *stx1*

(CALDERWOOD et MEKALANOS, 1987). De plus, le transfert des phages encodant les toxines Shiga d'une souche de *E. coli* vers une autre a été démontré *in vivo* (ACHESON *et al.*, 1998). Le peroxyde d'hydrogène produit par les neutrophiles semble être le facteur responsable de cette induction des prophages *in vivo* (WAGNER *et al.*, 2001). Également, l'induction des prophages chez *E. coli* a été observée suite à un traitement avec certains antibiotiques (MATSUSHIRO *et al.*, 1999; GAMAGE *et al.*, 2003; HEROLD *et al.*, 2005). Ces observations sont d'une importance considérable au niveau clinique et désormais le pouvoir des bactériophages à encoder ou transmettre des facteurs de virulence n'est plus mis en doute.

Plusieurs autres espèces bactériennes, mis à part celles mentionnées plus haut, tirent leur pouvoir pathogène des toxines ou autres facteurs de virulences cruciaux encodés par les phages. Parfois, il s'agit de toxines comme c'est le cas chez *P. aeruginosa* (cytotoxine *ctx* produite par le phage Φ CTX) (NAKAYAMA *et al.*, 1999), *S. pyogenes* (toxines de type A et C encodées par les phages T12 et CS112) (WEEKS et FERRETTI, 1984; GOSHORN et SCHLIEVERT, 1989) ou encore *Staphylococcus aureus* (entérotoxines A et P encodées par les phages ϕ 13 et ϕ N315 respectivement) (COLEMAN *et al.*, 1989). Il est intéressant de noter que la toxine la plus puissante connue à ce jour, la neurotoxine C1 chez *C. botulinum*, est également encodée par un prophage (BARKSDALE et ARDEN, 1974).

Parfois, les phages vont véhiculer des facteurs de virulence autres que les toxines, ce qui va prodiguer un avantage sélectif aux bactéries lysogènes. Il peut s'agir de protéines capables d'altérer l'antigénicité tels que les gènes *rfb* et *gtr*, encodés par

les phages ϵ^{34} et P22 respectivement, chez *Salmonella enterica* (WRIGHT, 1971; VANDER BYL et KROPINSKI, 2000). Chez la même espèce, certaines enzymes impliquées dans la pathogenèse sont également d'origine phagique, e.g. superoxyde dismutase (*sodC-III*) et neuraminidase (*nanH*) encodées par le phage Fels-1 (FIGUEROA-BOSSI et BOSSI, 1999). Deux autres enzymes, staphylokinase (*sak*) et phospholipase (*sla*) sont encodées par les phages $\phi 13$ et 315.4 chez *S. aureus* et *S. pyogenes* respectivement (BERES *et al.*, 2002).

Cependant, les phages possèdent bien d'autres moyens pour influencer le mode de vie bactérien, mis à part le fait d'encoder des facteurs de virulence directs.

1.7.2 Inactivation des gènes bactériens suite à l'intégration des prophages

Les bactériophages peuvent utiliser des sites d'intégrations fort différents allant de séquences codantes aux espaces intergéniques. Cependant, nombreux sont les prophages, autant chez les bactéries Gram+ que Gram-, qui s'intègrent dans les gènes codant pour des ARN de transfert (ARNt) (CAMPBELL, 1992; CAMPBELL *et al.*, 2002). Une perte de gènes codant pour les ARNt peut être néfaste pour la bactérie. Cependant, il n'est pas rare de constater que le site d'attachement de ces phages (*attP*) va reconstituer la séquence codante du gène en question suite à l'intégration (VENTURA *et al.*, 2003). Par contre, l'inactivation des gènes bactériens par l'intégration des phages a également été observée. Un exemple bien connu est celui d'une glycérol-ester-hydrolase (gène *geh*) chez *S. aureus* inactivé suite à l'intégration

d'un phage (LEE et IANDOLO, 1985; LEE et IANDOLO, 1986). Également chez *S. aureus*, l'insertion du bactériophage phi13 dans le chromosome bactérien résulte en l'inactivation du gène *hly* et la perte de l'expression de la toxine-béta menant à un changement phénotypique majeur (COLEMAN *et al.*, 1991). Le gène *hlyA* chez *S. pyogenes* représente un site d'intégration pour le phage SF370.3, par contre il n'y a pas d'inactivation du gène suite à l'intégration puisque le prophage reconstitue la séquence interrompue (CANCHAYA *et al.*, 2002). Chez *C. difficile*, le myophage tempéré ϕ C2 s'intègre dans un régulateur transcriptionnel *gntR*, menant à son inactivation (GOH *et al.*, 2007). L'impact phénotypique de cette intégration n'est pas connu précisément.

Donc, l'intégration d'un phage dans un gène bactérien peut mener à l'inactivation de ce dernier et apporter un changement phénotypique observable. Cependant, cet impact reste relativement limité puisqu'un bon nombre de phages s'intègre dans les régions intergéniques ou encore il y a complémentation de la séquence interrompue par le phage.

1.7.3 Immunité de l'hôte face aux phages du même groupe

Un avantage, certes discutable, que les bactéries vont tirer de la présence d'un phage dans leur génome est l'immunité que le prophage va procurer à sa bactérie hôte face à l'infection lytique par les phages du même groupe. Afin de maintenir l'état lysogénique, un phage tempéré intégré dans un chromosome bactérien exprime un

répresseur transcriptionnel que l'on nomme habituellement CI en référence à l'homologue retrouvé dans le phage Lambda de *E. coli* et qui est responsable de la répression des gènes impliqués dans le cycle lytique. De telle manière, le phage reste latent dans le génome bactérien jusqu'à l'activation du système SOS bactérien *recA*-dépendant qui est à l'origine de la dégradation du répresseur CI et de l'induction du cycle lytique (Figure 5). Cependant, dans le cas où, durant la lysogénie, un phage apparenté à celui intégré dans le chromosome infecte la cellule, son cycle lytique sera bloqué par le répresseur CI exprimé par le phage intégré. Ceci a pour effet de prévenir la lyse des cellules suite à la surinfection par les phages du même groupe. Dans le cas où le nouveau phage qui infecte une cellule lysogène est suffisamment différent du phage déjà présent, le répresseur CI ne sera pas en mesure de reconnaître les régions promotrices du nouveau phage et donc, le cycle lytique sera possible.

En somme, les bactéries lysogènes ont un certain avantage à conserver un prophage dans leur génome, puisque ça leur procure une protection contre l'infection subséquente par les phages apparentés. Par contre, le danger d'une induction du phage intégré sous l'effet des signaux externes est toujours présent.

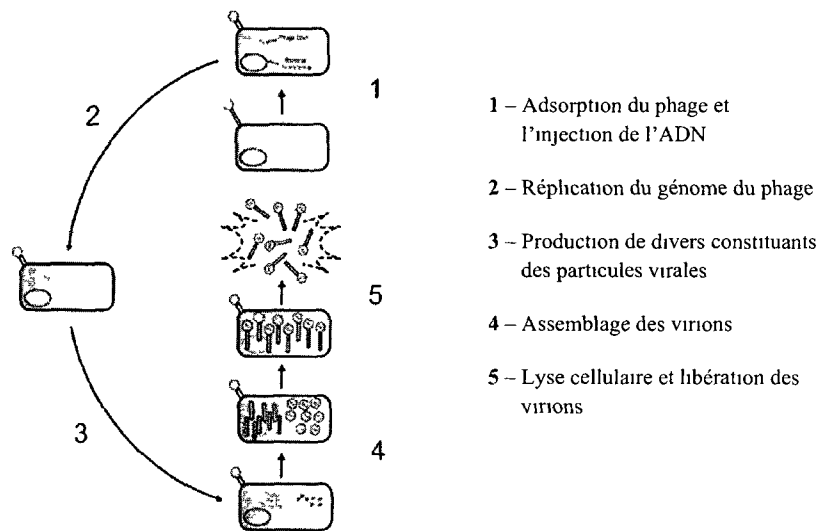


Figure 5 Cycle lytique. Représentation schématique de l'infection d'une bactérie par un phage strictement lytique.

1.7.4 Réarrangements chromosomiques via la recombinaison homologue

La recombinaison homologue intra chromosomique a été observée fréquemment dans les génomes bactériens. Ce phénomène survient lorsque deux régions (éloignées ou non) dans le chromosome bactérien partagent une certaine similarité de séquence; ceci leur permet de recombiner et donne lieu à des réarrangements chromosomiques tels des inversions et des délétions (Figure 6). La duplication de certains gènes peut créer un point chaud de recombinaison, de même que les gènes qui se trouvent naturellement en multiples copies dans le chromosome bactérien (p. ex. ARNr). Il est possible de visualiser le phénomène en alignant deux génomes bactériens (Figure 6). Lorsque les deux génomes partagent une forte

similarité de séquence, on observe une ligne droite suite à l'alignement (Figure 6, A). Par contre, lorsqu'une recombinaison homologue est responsable d'une inversion chromosomique, on observe une forte similarité (ligne droite), mais qui n'est plus colinéaire (Figure 6, B).

Dans l'exemple présenté à la Figure 6 B, l'alignement bio-informatique de deux génomes de *S. pyogenes* a relevé deux événements de recombinaison. La première inversion (en rouge) est due aux gènes bactériens *com*, la deuxième (en bleu) est due à deux prophages similaires, présents dans le chromosome bactérien. Également, comme c'est le cas chez *Xyllela fastidiosa*, la duplication de phages peut provoquer la formation des points de recombinaison (CANCHAYA *et al.*, 2004). Ce type de phénomène peut affecter la biologie bactérienne, sans toutefois avoir des impacts phénotypiques majeurs.

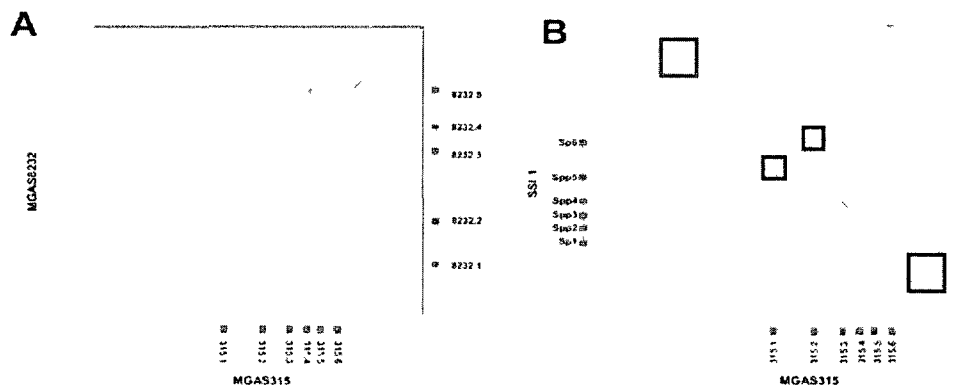


Figure 6 Représentation schématique d'une réorganisation chromosomique. A) alignement de deux souches de *S. pyogenes*, MGAS315 et MGAS8232 partageant une forte similarité. Les différences d'alignement sont dues au contenu en prophage. B) alignement de deux souches de *S. pyogenes*, MGAS315 et SSI-1 présentant deux événements de recombinaison (BRUSSOW *et al.*, 2004)

La revue de ces nombreux exemples nous montre que l'impact des phages sur la biologie bactérienne est un phénomène omniprésent. Cet impact peut se présenter sous forme de facteurs de virulence directs que les phages tempérés vont apporter à la bactérie ou encore de moyens plus subtils tels que les réarrangements au sein des chromosomes bactériens. Également, il semble évident que le transfert horizontal du matériel génétique compte énormément sur la balance de l'évolution bactérienne. Donc, il est d'à-propos de considérer les phages en tant que moteurs évolutifs à court terme chez *C. difficile*.

1.8 Phages de *C. difficile*

Les exemples de bactéries bénignes qui ont pu évoluer vers des pathogènes puissants dus à l'acquisition de phages tempérés dans le temps ne manquent pas. Par contre, chez *C. difficile*, très peu de données sont disponibles quant à l'impact des prophages sur cette espèce.

Les premiers rapports faisant mention de phages chez l'espèce *C. difficile* remontent au milieu de 1980 (SELL *et al.*, 1983; HAWKINS *et al.*, 1984; MAHONY *et al.*, 1985; DEI, 1989; NAGY et FOLDES, 1991). Une tentative de conversion toxigène avec deux des phages identifiés n'a donné aucun résultat positif (MAHONY *et al.*, 1985). Récemment, plusieurs phages de la famille des *Myoviridae*

et *Siphoviridae* ont également été décrits chez *C. difficile* (FORTIER et MOINEAU, 2007).

À ce jour, seulement six génomes de phages de *C. difficile* ont été séquencés et publiés (GOVIND *et al.*, 2006; SEBAIHIA *et al.*, 2006; GOH *et al.*, 2007; MAYER *et al.*, 2008; HORGAN *et al.*, 2010), et jusqu'à très récemment, une seule souche de *C. difficile* avait été séquencée et annotée (SEBAIHIA *et al.*, 2006). Toutefois, plusieurs génomes d'isolats de *C. difficile* ont été séquencés et rendus disponibles sur les bases de données publiques (notamment NCBI, <http://www.ncbi.nlm.nih.gov/>) au cours de la dernière année et l'étude du contenu en prophage de ces génomes sera intéressante.

L'analyse des génomes de phages séquencés a permis de constater une absence, à priori, de facteurs de virulence connus, comme c'était le cas chez plusieurs phages d'autres espèces bactériennes. Donc, jusqu'à présent, les phages de *C. difficile* ne semblent pas encoder de toxines connues, incluant les toxines A et B. Fait à noter, jusqu'à tout récemment, tous les phages de *C. difficile* dont le génome a été séquencé (ϕ C2, ϕ CD27, ϕ CD119 et les deux prophages de la souche CD630, ϕ CD630-1 et 2) font partie d'une même famille de l'ordre des *Caudovirales*, soit celle des *Myoviridae* (ACKERMANN, 2007). Cette famille est caractérisée par des phages ayant une morphologie similaire comprenant principalement une capsidie isométrique et surtout, une queue longue et contractile. Par ailleurs, les phages de type *Myoviridae* chez *C. difficile* partagent une similarité de séquence importante (GOVIND *et al.*, 2006; FORTIER et MOINEAU, 2007; GOH *et al.*, 2007). Cela a pour conséquence de limiter la diversité des séquences génomiques disponibles, et donc l'information utile

qu'on peut en tirer. Un certain nombre de phages faisant partie de la famille des *Siphoviridae*, une autre famille de l'ordre des *Caudovirales*, ont été décrits dans la littérature et ceux-ci sont génétiquement distincts des phages *Myoviridae* et partagent également peu de similarité de séquences entre eux (MAHONY *et al.*, 1985; NAGY et FOLDES, 1991; GOH *et al.*, 2005; FORTIER et MOINEAU, 2007; HORGAN *et al.*, 2010).

Cependant, malgré le fait qu'aucun facteur de virulence n'a été détecté dans les génomes de phages de *C. difficile* séquencés, certains auteurs font mention d'une altération de la production de toxines chez certaines souches lysogènes de *C. difficile*.

La première étude, datant de 2005, rapporte un impact intéressant de trois phages, ϕ C2, ϕ C6 et ϕ C8 sur la production et transcription des gènes de toxines (GOH *et al.*, 2005). La lysogénisation de souches de *C. difficile* par ces trois phages a résulté en une augmentation de la toxine B dans le milieu extracellulaire. Par contre, les analyses de PCR en temps réel n'ont pu démontrer une altération dans la transcription du gène *tcdB*. À l'opposé, le phage ϕ C6 s'est montré capable de diminuer la transcription du gène de la toxine A (*tcdA*), tandis que le phage ϕ C8 a eu comme effet d'augmenter la transcription du même gène. Cette apparente incohérence entre la transcription des gènes de toxines et la quantité de celles-ci dans les surnageants de culture n'a pas pu être expliquée, et le séquençage du génome du phage ϕ C2 en 2007 n'a pas apporté plus d'informations (GOH *et al.*, 2007). De plus, cette étude suggérait que la combinaison phage - hôte avait une influence sur le phénotype observé et donc que l'impact des phages pouvait être hôte - dépendant.

La deuxième étude, plus récente, a apporté un aperçu moléculaire plus précis quant à l'impact des phages sur la régulation de la transcription des gènes du locus de pathogénicité (GOVIND *et al.*, 2009). Le séquençage complet du génome du phage ϕ CD119 n'avait révélé aucun facteur de virulence identifiable (GOVIND *et al.*, 2006). Cependant, la lysogénisation de souches de *C. difficile* par ce phage a résulté en une diminution de la production des toxines A et B. La transcription des gènes du locus de pathogénicité (*tcdA*, *tcdB*, *tcdE*, *tcdR* et *tcdC*) a également été négativement affectée. Une analyse plus poussée a démontré qu'un répresseur du phage, encodé par le gène *repR*, était en mesure de lier la région promotrice en amont du gène *tcdR*, régulateur transcriptionnel positif des gènes du locus de pathogénicité. Cette interaction avait pour conséquence de réprimer la transcription des gènes du PaLoc, affectant directement la production des toxines A et B. Cette étude a démontré que les phages, même en n'encodant aucun facteur de virulence évident, sont en mesure d'influencer des phénotypes bactériens majeurs, tels que la production de toxines.

La ligne directrice du projet présenté dans ce manuscrit est en parallèle avec les deux études mentionnées ci-dessus. Le potentiel des phages d'avoir un impact sur la biologie de *C. difficile* étant bel et bien démontré, il s'agissait d'approfondir les connaissances sur ce groupe de phages en particulier. De plus, le peu de données génomiques disponibles rendait d'autant plus nécessaire l'acquisition de nouvelles séquences génomiques de phages de *C. difficile*, en particulier de phages faisant partie de la famille des *Siphoviridae*.

1.9 Les objectifs de la présente étude

Tel que mentionné précédemment, les phages de *C. difficile* dont les séquences génomiques sont disponibles présentement, font tous partie, sauf une exception, d'une même famille très apparentée sur le plan génétique (*Myoviridae*). Au moment d'écrire ce mémoire (avril 2010), la première séquence d'un phage de *C. difficile* de la famille des *Siphoviridae* a été publiée (HORGAN, 2010). Nous étions sur le point, avec le dépôt de ce mémoire, de soumettre pour publication la séquence d'un phage de cette famille, ce qui aurait pu être le premier exemple chez *C. difficile*. Néanmoins, le séquençage du génome d'un autre phage de cette famille apportera une diversité génétique nécessaire à une meilleure connaissance des phages infectant cette espèce bactérienne.

L'objectif premier de ce projet consistait à isoler des phages tempérés de *C. difficile* à partir d'isolats cliniques, afin d'effectuer une caractérisation détaillée sur un phage de la famille des *Siphoviridae*. Précédemment, plusieurs de phages de *C. difficile* ont été isolés et partiellement caractérisés, dont le phage ϕ CD38-2. Une analyse par microscopie électronique à transmission a montré qu'il s'agissait d'un siphophage. De plus, les analyses de similarité de séquences ont permis de constater que ce phage était génétiquement très distant des autres phages de *C. difficile* (FORTIER et MOINEAU, 2007). Donc, logiquement, le phage ϕ CD38-2 a été choisi pour une caractérisation plus poussée.

D'une manière générale, la caractérisation d'un phage englobe habituellement un aspect microbiologique comportant, entre autre, la détermination de la morphologie des virions par microscopie électronique à transmission. Certains paramètres de la dynamique d'infection (temps de latence, taille de la progéniture) sont des informations utiles lors de la détermination de l'efficacité d'infection d'un phage particulier. Également, la détermination de son spectre d'hôte, i.e. le nombre de souches différentes qu'un phage est capable d'infecter, est une donnée importante dans la mesure où cela peut donner un aperçu de l'impact global que peut avoir ce phage sur la population bactérienne en général. L'aspect moléculaire, quant à lui, comporte le séquençage du génome entier du phage en question incluant l'analyse bioinformatique des séquences obtenues, ainsi qu'une analyse des protéines structurales composant les virions.

Suite à cette première caractérisation globale, le deuxième objectif de ma maîtrise consistait à évaluer l'impact des phages sur la biologie de *C. difficile*. Les phénotypes majeurs et facilement détectables ont été étudiés. Il s'agissait dans un premier temps de vérifier si le phage avait un impact sur la croissance bactérienne *in vitro*, et ensuite d'évaluer son impact sur la production de toxines TcdA et TcdB, incluant l'analyse transcriptionnelle des gènes correspondants. Pour des questions de temps, d'autres phénotypes comme la sporulation et la capacité d'adhésion n'ont pas pu être étudiés, mais cela fera partie des perspectives de ce projet.

Les travaux décrits dans ce mémoire ont été soumis pour publication et permettent d'élargir nos connaissances sur les phages de *C. difficile* qui demeurent, jusqu'à présent, très peu étudiés.

2. ARTICLE

Sekulovic O., Meessen-Pinard M., Fortier L-C., Prophage-stimulated toxin production in *Clostridium difficile* NAP1/027 lysogens, *Journal of Bacteriology*, 2010, soumis

2.1 Avant propos

Ce manuscrit représente l'aboutissement de plus de deux ans de travail dans le cadre de mes études de 2^e cycle, débutées en septembre 2007 et terminées en mai 2010. L'essentiel des expériences effectuées dans cette période fait l'objet du présent manuscrit. La rédaction de l'article est effectuée dans la langue anglaise puisque un journal anglophone (*Journal of Bacteriology*) a été ciblé pour la soumission de la publication.

Ma contribution englobe toutes les expériences effectuées et présentées dans cet article sauf ce qui à trait au ribotypage des souches bactériennes ainsi que les expériences relatives aux inductions spontanées. L'ensemble de ces expériences a été réalisé par Mathieu Meessen-Pinard. Ma contribution dans l'écriture du manuscrit est également majeure, sous la supervision directe de Louis-Charles Fortier.

Auteurs de l'article : Ognjen Sekulovic, Mathieu M. Pinard, Louis-Charles Fortier

**AUTORISATION D'INTÉGRATION
D'UN ARTICLE ÉCRIT EN COLLABORATION
À UN MÉMOIRE OU UNE THÈSE**

Je (ou nous), soussignée(s), soussigné(s), co-auteure(s), co-auteur(s) de l'article intitulé : _____
Prophage-stimulated toxin production in *Clostridium difficile* NAP1/027 lysogens

reconnais (reconnaissons) que ledit article sera inclus comme partie constituante du mémoire de la thèse

de l'étudiant(e) (nom) : **Ognjen Sekulovic**

inscrit(e) au programme de **Maîtrise en Microbiologie**

de la Faculté de **Médecine et des sciences de la santé** de l'Université de Sherbrooke

En foi de quoi, j'ai (nous avons) signé cet engagement en un nombre suffisant d'exemplaires*

Signature _____ Date : **5 mai 2010**

Nom **Ognjen Sekulovic**

Coordonnées **Faculté de médecine et des sciences de la santé
Département de Microbiologie et Infectiologie
3001, 12^e Avenue Nord, J1H 5N4, Sherbrooke (Québec)**

Signature _____ Date : **5 mai 2010**

Nom **Mathieu M. Pinard**

Coordonnées **Faculté de médecine et des sciences de la santé
Département de Microbiologie et Infectiologie
3001, 12^e Avenue Nord, J1H 5N4, Sherbrooke (Québec)**

Signature _____ Date : **5 mai 2010**

Nom **Louis-Charles Fortier**

Coordonnées **Faculté de médecine et des sciences de la santé
Département de Microbiologie et Infectiologie
3001, 12^e Avenue Nord, J1H 5N4, Sherbrooke (Québec)**

Signature _____ Date : _____

Nom _____

Coordonnées _____

* Un exemplaire pour l'étudiante, l'étudiant, un exemplaire pour chaque personne signataire et un exemplaire pour le Service des bibliothèques (à remettre avec le mémoire ou la thèse au moment du dépôt final).

2.2 Résumé de l'article

L'augmentation constante des infections à *Clostridium difficile* dans le monde continue à susciter un intérêt particulier envers ce pathogène entérique. Dans ce sens, de nombreuses études ont été menées dans le but de comprendre la biologie, l'évolution et la pathogénèse de cette espèce bactérienne. Malgré tout, plusieurs aspects touchant la virulence et l'émergence des souches hypervirulentes restent à être clarifiés. Par analogie à d'autres pathogènes, on est en mesure de se questionner à propos de l'impact des phages tempérés sur l'ensemble de la biologie de *C. difficile*. Dans ce manuscrit, nous rapportons la caractérisation microbiologique et moléculaire complète du phage tempéré ϕ CD38-2 (*Siphoviridae*) capable d'infecter et lyser la bactérie *C. difficile*, notamment la grande majorité des souches de type NAP1/027. La lysogénéisation d'une souche hypervirulente avec ce phage a résulté en une stimulation de la transcription des gènes *tcdA* et *tcdB*, ce qui a conduit à une plus grande accumulation des toxines A et B dans le milieu extracellulaire. Par contre, cet effet semblait être hôte-dépendant.

1

2 **Prophage-stimulated toxin production in *Clostridium difficile* NAP1/027 lysogens**

3

4

5 Ognjen Sekulovic, Mathieu Meessen-Pinard and Louis-Charles Fortier*

6

7

8 Département de microbiologie et d'infectiologie,

9 Faculté de médecine et des sciences de la santé,

10 Université de Sherbrooke, Québec, Canada, J1H 5N4,

11

17

18 Running title: Prophage-stimulated toxin production in *C. difficile*

19

20 Keywords: *Clostridium difficile*, NAP1/027, toxins, temperate bacteriophage,

21

Siphoviridae

ABSTRACT

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21

TcdA and TcdB exotoxins are the main virulence factors of *Clostridium difficile*, one of the most deadly nosocomial pathogens. Recent data suggest that prophages can influence the regulation of toxin expression. Here we present the characterization of ϕ CD38-2, a *pac*-type temperate *Siphoviridae* phage that stimulates toxin expression when introduced into the chromosome of *C. difficile*. Host range analysis showed that ϕ CD38-2 was able to infect 99/207 isolates of *C. difficile* representing 11 different PCR ribotypes. Of 89 isolates corresponding to the NAP1/027 hypervirulent strain that recently caused several outbreaks in North America and Europe, 79 (89%) were sensitive to ϕ CD38-2. The complete dsDNA genome was determined and is 41,090 bp long and the overall gene organization is similar to other phages infecting low G+C bacteria. A putative function could be assigned to only 24 of the 55 open reading frames and there was no direct evidence of the presence of genes coding for toxins or other virulence factors. Upon lysogenization with ϕ CD38-2 of CD274, a NAP1/027 representative isolate, 1.6 and 2.1-fold more TcdA and TcdB were detected by ELISA assays and immunodot blots in culture supernatants of the lysogen. In addition, real-time qRT-PCR analyses showed that the levels of *tcdA* and *tcdB* mRNAs were up to 5-fold higher in the CD274 lysogen. Our study provides the first genomic sequence of a new *pac*-type *Siphoviridae* phage family infecting *C. difficile* and brings further evidence supporting the role of prophages in the virulence and toxin production of this important nosocomial pathogen.

INTRODUCTION

1
2 *C. difficile* is a Gram-positive, strictly anaerobic, spore-forming bacillus causing
3 infections with various symptoms ranging from asymptomatic carriage to fulminant
4 diarrhea and pseudomembranous colitis (51). *C. difficile* infection (CDI) is the most
5 frequent cause of antibiotic-associated nosocomial diarrhea in industrialized countries
6 (21). This opportunistic pathogen caused severe outbreaks in North America and Europe
7 over the last 8 years (26, 28, 44). TcdA and TcdB exotoxins are the main virulence
8 factors of *C. difficile* and are encoded on a 19.6 kb chromosome region called the
9 pathogenicity locus (PaLoc) which is found in all toxigenic isolates (51). A hypervirulent
10 epidemic strain, called BI/NAP1/027, was shown to produce 16 times more TcdA and 23
11 times more TcdB, *in vitro* (52). The increased toxin production is thought to be
12 responsible for the greater disease severity and higher mortality rates reported for patients
13 infected with this particular strain (26, 28, 44).

14
15 The expression of *C. difficile* toxins is growth phase-dependent. This regulation is
16 achieved through the expression of TcdR, an alternative sigma factor that acts as a
17 positive regulator of toxin expression and TcdC, an early expressed anti-sigma factor that
18 prevents the TcdR-containing RNA polymerase from binding to toxin promoters (11, 34,
19 35, 37). A number of deletions were reported in the *tcdC* gene from various clinical
20 isolates (8). A particular 1-bp deletion causing a -1 frameshift mutation and the
21 expression of a truncated protein could possibly explain the increased toxin production
22 observed *in vitro* for the NAP1/027 epidemic strain (8, 30, 37, 52).

23

1 Temperate bacteriophages (or simply phages) have played a determinant role in the
2 virulence and evolution of major bacterial pathogens (5, 7). Temperate phages can lead to
3 lysogeny, which occurs when the phage integrates into the bacterial chromosome and
4 remains as a “*latent*” prophage. During this lysogenic cycle, prophages sometimes
5 modify the phenotype of their host, for example by expressing highly potent toxins, like
6 the *Shiga* toxins (Stx) in *Escherichia coli*, the cholera toxin (CT) in *Vibrio cholerae*, or
7 the botulinum neurotoxins (BoNTs) in *C. botulinum* (5).

8

9 A number of phages infecting *C. difficile* have been isolated and partially characterized
10 so far (9, 13, 18, 19, 24, 31, 39, 42, 46), but all of them are temperate. Besides the two
11 prophages that were identified in the genome of *C. difficile* strain 630 (45), only four
12 phages have been characterized at the molecular level, including complete genome
13 sequencing, namely ϕ CD119 (19), ϕ C2 (17), ϕ CD27 (39) and ϕ CD6356 (24). All these
14 phages are members of the *Myoviridae* family except ϕ CD6356, which is the first and
15 only *Siphoviridae* member for which a complete genome sequence is currently available
16 (24). Hence, there is a clear lack of genomic data for this group of phages, especially
17 those of the *Siphoviridae* family. So far, *C. difficile* phages have never been found to
18 encode proven virulence factors or to convert non-toxigenic *C. difficile* isolates into
19 toxin-producing lysogens (17, 19, 24, 39). Nevertheless, two recent studies suggest that
20 phages may contribute somehow to the regulation of toxin production in *C. difficile* (16,
21 20), but the clear lack of data regarding phages of *C. difficile* makes it difficult to
22 appreciate the real impact of prophages on *C. difficile* life style and virulence.

23

1 In a previous study, we identified ϕ CD38-2, a temperate *Siphoviridae* phage induced
2 from a *C. difficile* clinical isolate (13). Here, we provide the full characterization of this
3 phage including whole genome sequencing and phenotypic characterization of lysogens.
4 We also provide additional evidences supporting that prophages contribute to the
5 virulence of this important nosocomial pathogen.

6

7

MATERIALS AND METHODS

8 **Bacteria and growth conditions**

9 All strains used in this study were isolated from human fecal samples kindly provided by
10 Dr. Louis Valiquette of the Université de Sherbrooke. Strain CD274, the host strain for
11 ϕ CD38-2, has all the characteristics of the B1/NAP1/027 hypervirulent strain (binary
12 toxin-positive, PCR ribotype 027, *tcdC* deletion at position 117). Bacteria were routinely
13 grown inside a ThermoForma model 1025 anaerobic chamber (Fisher Scientific) under
14 anaerobic atmosphere (10% H₂, 5% CO₂ and 85% N₂) at 37°C in pre-reduced brain heart
15 infusion (BHI) broth (Oxoid) or in TY broth (3% tryptose, 2% yeast extract, pH 7.4).

16 **Bacterial DNA extraction and PCR ribotyping**

17 Three milliliters of an overnight *C. difficile* culture were centrifuged and total genomic
18 DNA was extracted using the Illustra Bacterial Genomic DNA extraction kit following
19 the manufacturer's recommendations (GE Healthcare). PCR ribotyping was performed on
20 an Eppendorf Mastercycler with 20 ng purified DNA and primers published by Bidet
21 *et al.* (3), with modifications described previously (13). Band patterns were analyzed and
22 compared using GelComparII (Applied Maths).

1 **Prophage induction and phage propagation**

2 Phage ϕ CD38-2 was isolated from a mitomycin C induction lysate (13). Three rounds of
3 purification from single plaques were performed using the double agar overlay method
4 (15) and 0.5 mL of a log-phase culture (optical density at 600nm, $OD_{600nm} = 0.4$) of
5 *C. difficile* strain CD274 as the sensitive host. The addition of 10 mM $CaCl_2$ and 0.4 M
6 $MgCl_2$ into the soft agar was required to obtain plaques. For routine prophage induction,
7 10 μ L of serial ten-fold dilutions of an overnight culture of *C. difficile* were spotted onto
8 a BHI agar plate and incubated for 4 hours at 37°C under anaerobic atmosphere to allow
9 cells to reach the log-phase. Plates were then irradiated under UV light (302nm) for 10
10 seconds on a standard UV Transilluminator (GE Healthcare). A soft agar overlay was
11 then poured on top of the plates as above. Clear zones in the bacterial lawn after an
12 overnight incubation at 37°C under anaerobic atmosphere were indicative of a successful
13 prophage induction. Phages were then purified from an agar plug as described above.
14 Standard procedures were used for amplification in BHI broth. $CaCl_2$ and $MgCl_2$ were
15 added to a final concentration of 10 mM each and phage titers were determined by the
16 soft agar overlay method described above. Titters $\geq 10^9$ plaque-forming units per mL
17 (PFU/mL) were easily obtained with this method.

18 **Transmission electron microscopy (TEM)**

19 Phage particles were washed with 0.1 M ammonium acetate pH 7.5, deposited onto 400-
20 mesh Formvar/carbon-coated copper grids (Cedarlane Laboratories) and negatively
21 stained with 2% uranyl acetate (UA) as described before (13). The grids were observed at
22 60 kV with a Hitachi H-7500 transmission electron microscope equipped with a 1K x 1K

1 Hammamatsu 1 megapixel digital camera controlled with the AMT software (Advanced
2 Microscopy Techniques).

3 **Host range determination and one step growth curve assays**

4 A spot test on soft agar overlays prepared as describe above was used to determine the
5 host range of ϕ CD38-2 using 10 μ L of a 10-fold diluted phage lysate and a collection of
6 207 clinical isolates representing 41 different PCR ribotypes. For one step growth curve
7 assays, cells were grown in pre-reduced BHI until the OD_{600nm} reached 0.8. Then, a
8 phage aliquot was added to 2 mL of bacterial culture to obtain an MOI= 0.05. $CaCl_2$ and
9 $MgCl_2$ were added to a final concentration of 10 mM each followed by 5 min incubation
10 at 37°C to allow adsorption. To remove non-adsorbed phages, 1 mL of the cell
11 suspension was washed three times with pre-reduced BHI and then serial ten-fold
12 dilutions were made in 10 mL BHI containing 10 mM $CaCl_2$ and $MgCl_2$ and incubated at
13 37°C under anaerobic atmosphere. Aliquots were taken at fixed intervals during 180 min
14 and phage titers were determined as described above. The burst size was calculated as
15 follows: (final phage titer – initial phage titer) / (initial phage titer).

16 **Analysis of structural proteins by SDS-PAGE and mass spectrometry**

17 Phage particles from a 1 L cleared lysate ($\sim 10^9$ PFU/mL) were purified by two successive
18 rounds of discontinuous cesium chloride gradient, as described previously (14). Twenty
19 microliters of purified phage particles (5×10^{11} PFU/mL) were analyzed on a 12%
20 denaturing SDS-polyacrylamide gel as before (14). After Coomassie blue staining,
21 protein bands were cut out of the gel, digested with trypsin and analyzed by liquid
22 chromatography-tandem mass spectrometry (LC-MS/MS) at the Proteomics platform of
23 G enome Qu ebec Innovative center at McGill University (Montr eal, Canada).

1 **Phage DNA purification and restriction analysis**

2 Small-scale preparations of whole phage DNA were obtained from cleared lysates using a
3 rapid mini-prep protocol described elsewhere (40). For larger preparations, the Maxi
4 Lambda DNA purification kit was used following the manufacturer's recommendations
5 (QIAGEN). Phage DNA was digested with restriction enzymes (Roche) and the digested
6 products were heated at 80°C for 10 min and immediately run through a 0.8% agarose
7 gel. Gels were stained with ethidium bromide, exposed to UV, and photographed using
8 the ImageQuant 300 gel documentation system (GE Healthcare).

9 **Phage genome sequencing and bioinformatics analysis**

10 Whole phage genome sequencing and assembly was performed on a Roche 454 GS-FLX
11 platform using the Titanium chemistry at the Génome Québec Innovation Center of
12 McGill University (Montréal, Canada). Additional sequencing reactions were done
13 directly on purified phage DNA with specific primers on an Applied Biosystems ABI
14 3730xl sequencer at the genomic platform of the CHUL research center (Québec,
15 Canada). Additional sequence assembly was done using the Gap v4.10 application of the
16 Staden package v1.6.0. Some editing was also done using BioEdit v7.0.5.3 and Artemis
17 11.22. Putative open reading frames (ORFs) were predicted using GeneMark.hmm for
18 Prokaryotes v2.4 and Glimmer v3.02. The predicted proteins were compared with the
19 BLASTp tools of the NCBI (2) and ACLAME (27) databases. Structural features and
20 domains in predicted proteins were identified using InterProScan.

21 **Isolation of lysogens**

22 Lysogens were created using a modified soft agar overlay method. Briefly, 0.1 mL of a
23 ϕ CD38-2 lysate ($\sim 10^8$ PFU/mL) was incorporated into a BHI soft agar containing CaCl_2

1 and MgCl₂ that was then poured over BHI agar plates. Serial ten-fold dilutions of a log-
2 phase (OD_{600nm} = 0.4) sensitive host were spread over this phage lawn and incubated
3 overnight at 37°C under anaerobic atmosphere. Five phage-resistant colonies were picked
4 and re-streaked 3 times onto BHI plates without phages to purify the lysogens. The
5 presence of the prophage into each lysogen was confirmed by southern hybridization and
6 its functionality was assessed by UV induction followed by phage isolation, DNA
7 extraction and Hind III restriction profiling as described above.

8 **Detection of toxins A and B in culture supernatants**

9 An overnight pre-culture of *C. difficile* in TY broth was used to inoculate a fresh tube of
10 the same broth (3% inoculum). Cells were grown as before and the OD_{600nm} was
11 monitored over a 24 h period. Aliquots, each representing a maximum of 2.5% of the
12 total culture volume, were taken at 2, 8, 12, 18, 24 and 48 h post-inoculation. Cells were
13 removed by centrifugation and cleared supernatants were stored at -20°C until analysis.
14 Detection of the toxins was done using an ELISA assay (Premier™ Toxins A and B kit,
15 Meridian Biosciences), as recommended by the supplier. An immunodot blot method was
16 used to selectively detect TcdA and TcdB. For this, culture supernatants were serially
17 diluted in TY broth and directly spotted (0.1 mL) onto nitrocellulose membranes using a
18 96-well dotblotter apparatus. All wells were washed twice with phosphate-buffer (PBS),
19 after which the membrane was allowed to air-dry for 30 min. Toxins were detected with
20 monoclonal anti-TcdA or anti-TcdB mouse antibodies (Meridian Life Science) at a
21 1:3000 or 1:1000 dilution, respectively. A secondary anti-mouse IgG HRP-linked
22 antibody (Cell Signaling) was used at a 1:3000 dilution and the membranes were revealed
23 with the ECL Plus Western blotting detection system (GE Healthcare) as recommended

1 by the manufacturer, followed by exposition to Hyperfilm ECL autoradiography films
2 (GE Healthcare) (38). Spot intensities were compared using the ImageJ 1.42q software
3 (<http://rsbweb.nih.gov/ij/>).

4 **RNA extraction and gene expression analysis**

5 Total RNA was extracted from 10 mL culture samples after two successive treatments
6 with TRIzol (Invitrogen). Cells were broken during the first treatment by adding glass
7 beads ($\leq 106 \mu\text{m}$, Sigma) and using a FastPrep apparatus (MP Biomedicals) with the
8 following parameters: 2 cycles of 45 seconds at a speed of 4m/second. Total RNA was
9 dissolved in RNase-free water and treated with RNase-free DNase I (Invitrogen) for 1 h
10 at room temperature. A final extraction with TRIzol was performed to obtain pure RNA
11 preparations in RNase-free water containing RNase-OUT (Invitrogen). The absence of
12 contaminating genomic DNA was verified by performing a 40-cycle PCR reaction with
13 primers targeting the 16S rDNA gene in the presence of 800 ng total RNA. First-strand
14 cDNA synthesis was performed on 3 μg total RNA using SuperScriptII RT (Invitrogen)
15 with random primers (Promega) according to the manufacturer's specifications. Real-
16 time quantitative RT-PCR reactions were performed on a Mastercycler ep realplex
17 (Eppendorf) in a total volume of 10 μL with the following components: 1X iTaq SYBR
18 Green Supermix (Bio-Rad Laboratories), 150 ng of template cDNA and one of the primer
19 sets specific for *tcdA* (0.25 μM), *tcdB* (0.5 μM) or 16S rRNA (0.25 μM) (Table S1). The
20 cycling conditions were the following: 95°C for 2 min followed by 40 cycles of 95°C for
21 15 s and 60°C for 1 min. The $\Delta\Delta\text{Ct}$ comparative method was used to calculate the relative
22 gene expression of *tcdA* and *tcdB* using the 16S rRNA gene as the reference.

23

1 **Table S1. Primers used for real-time qRT-PCR assays**

Target gene	Strand	Primer	Sequence (5'- 3')	Product size (bp)	Reference
<i>tcdA</i>	Fwd	LCF 398	AGCTTTCGCTTTAGGCAGTGTATAAATCAAGC	139	This study
	Rev	LCF 399	TATCAGACTCTATGGCTGGGTTAAGGTG		
<i>tcdB</i>	Fwd	LCF 400	AGCATATGCGGCAGCTTATCAAGA	158	This study
	Rev	LCF 401	GCATCGTCAAATGACCATAAGCTAGCCATT		
16S <i>rRNA</i>	Fwd	LCF 408	GGGAGACTTGAGTGCAGGAG	120	Denève et al., 2009
	Rev	LCF 409	GTGCCTCAGCGTCAGTTACA		

2

3 **Nucleotide sequence accession number**

4 The complete genome sequence has been submitted to the GenBank database under the
5 accession number HM568888.

6

7

RESULTS8 **Phage isolation**

9 We have previously reported the induction of two temperate phages, ϕ CD38-1 and
10 ϕ CD38-2, after mitomycin C treatment of CD38, a clinical isolate of *C. difficile* (13). In
11 our screening to find sensitive bacterial hosts, we found that 0.4 M MgCl₂ was required
12 in addition to 10 mM CaCl₂ to obtain phage plaques of ϕ CD38-2 on bacterial lawns.
13 Using this high salt concentration in soft agar overlay assays, several *C. difficile* isolates
14 were found to be sensitive to ϕ CD38-2, including CD274, which has all the common
15 characteristics of the hypervirulent NAP1/027 strain that caused several outbreaks in
16 North America and Europe. Phage ϕ CD38-2 was purified from single plaques using
17 CD274 as the host and further propagated in BHI broth to $\geq 10^9$ PFU/mL. TEM analysis
18 (not shown) confirmed that the isolated phage was a member of the *Siphoviridae* family

1 of the order *Caudovirales* (1), and corresponded to the ϕ CD38-2 phage that we described
 2 before (13). Also, previous DNA-DNA hybridizations suggested that ϕ CD38-2 was
 3 genetically distinct from other *C. difficile* *Myoviridae* phages characterized so far (13) so
 4 we decided to fully characterize this phage, including complete genome sequencing.

5 **Host range and lytic growth cycle**

6 The host range of ϕ CD38-2 was determined using a collection of *C. difficile* clinical
 7 isolates and spot tests on soft agar overlays. ϕ CD38-2 infected 99 of the 207 isolates
 8 tested (48%), among which 79 isolates (80%) corresponded to the NAP1/027 epidemic
 9 strain (PCR ribotype 027). The other non-NAP1/027 sensitive isolates represented 10
 10 different PCR ribotypes (Table 1). The lytic growth cycle of ϕ CD38-2 was determined on
 11 strain CD274. The latent period was 95 min and the burst size was 35 ± 11 PFU per
 12 infected cell, which is in the range reported for other *C. difficile* phages (18, 31).

13

14

Table 1: Host range analysis of ϕ CD38-2.

PCR ribotype	Number of isolates	Number of sensitive isolates
001	2	1
006	3	2
007	2	2
014	22	6
016	3	2
019	1	1
020	3	1
023	5	1
024	13	2
027*	89	79
035	2	2

15

16

* Corresponds to the hypervirulent strain BI/NAP1/027

1 **Genome sequence**

2 Full genome sequencing was performed on a 454 GS-FLX platform. A total of 17,659
3 sequence reads were obtained and assembled in two large contigs totalizing 40,468 bp,
4 with an average coverage of ~90-fold. The two contigs were joined and gaps were filled
5 after performing sequencing reactions directly on the phage DNA. The ϕ CD38-2 genome
6 is composed of a dsDNA molecule of 41,090 bp with a G+C content of 30.83%, which is
7 a little above, but in the range reported for other *C. difficile* phages (28.7-29.4%) (17, 19,
8 39) and of *C. difficile* strain 630 (29.06%) (45). Digestion of the purified phage DNA
9 with several restriction enzymes gave profiles perfectly corresponding to a circular
10 genomic map, except for a faint submolar fragment of ~1 kb that was observed with
11 EcoRV. Heating the digestions at 80°C for 10 min prior to loading on the agarose gel
12 didn't reveal cohesive termini. Moreover, southern hybridization of the restricted DNA
13 with a probe covering nucleotides 39,450 to 40,663, which we suspected to be the region
14 containing the *pac*-site, revealed the expected submolar fragments in all digestions (data
15 not shown). Thus, our data indicate that ϕ CD38-2 is a *pac*-type phage that packages its
16 DNA using a headful mechanism and that the *pac*-site is located somewhere between
17 *orf53* and *orf54*. To our knowledge, this represents the first example of a *pac*-type
18 *Siphoviridae* phage infecting *C. difficile*.

19 **DNA homology and other similar prophages**

20 BLASTn analyses against the non-redundant nucleotide database at NCBI revealed the
21 existence of bacterial genomic fragments nearly identical to ϕ CD38-2 in two *C. difficile*
22 isolates currently being sequenced at McGill University: strain QCD-37x79 (contig
23 NZ_ABHG02000044) and QCD-63q42 (contigs NZ_ABHD02000046,

1 NZ_ABHD02000048, NZ_ABHD02000049 and NZ_ABHD02000057). The presence of
2 two prophages highly similar to ϕ CD38-2 in these two *C. difficile* isolates thus makes no
3 doubt. Intriguingly, the QCD-37x79 contig was shorter than ϕ CD38-2 by only 558 bp
4 and the missing part covered a region within *orf20* that contains 2.5 copies of a perfect
5 tandem repeat of 228 bp. The presence of these repeats has been confirmed by direct
6 sequencing on the ϕ CD38-2 phage DNA. Notably, only phage-related DNA
7 corresponding to ϕ CD38-2 (less 558 bp within *orf20*) was present on the QCD-37x79
8 and QCD-63q42 contigs and bacterial DNA could not be detected.

9 **Gene products and annotation**

10 Fifty-five (55) putative open reading frames (*orfs*) encoding proteins \geq 30 amino acids
11 were identified by automated GeneMark.hmm and Glimmer analyses using standard
12 (ATG) and alternative start codons (GTG, TTG, CTG). Manual validation of each *orf*
13 was then performed and the most probable start codon was selected based on the presence
14 of a suitable ribosome-binding site complementary to the 3' end of the 16S rDNA gene of
15 *C. difficile* 630 (45). Most *orfs* are encoded on the forward strand, except *orfs* 24, 28, 29,
16 30-33 and 39 that are in the reverse orientation, and of which 7 are located next to the
17 lysis cassette, between *orfs* 23 and 34. Overall, the total coding percentage was 89.4%
18 and the intergenic DNA covered 4,391 bp, with the largest intergenic region covering 450
19 bp between *orfs* 32 and 33. A genomic map of ϕ CD38-2 is presented in Fig. 1.

20

21 All predicted *orfs* were translated into proteins and compared against non-redundant
22 protein sequences from GenBank and ACLAME databases using BLASTp. Not
23 surprisingly, most ORFS had identical or nearly identical protein homologs in *C. difficile*

1 strains QCD-37x79 and QCD-63q42, as mentioned above. Table S2 shows the best
2 BLAST hits corresponding to gene products identified in strain QCD-37x79, along with a
3 second relevant hit from another source, when available (excluding hits from strain QCD-
4 63q42). Putative functions were attributed to each ORF based on BLAST results, by
5 comparison with homologous proteins found in the ACLAME database, and on the
6 presence of conserved domains found through searches in the conserved domain database
7 (CDD) at NCBI and by InterProScan analyses. Overall, a putative function was attributed
8 to only 23 of the 55 ORFS (42%) (Table S2). Moreover, only a few hits corresponded to
9 gene products from other phages, including *C. difficile* phages ϕ CD27 and ϕ CD119, thus
10 confirming that ϕ CD38-2 is completely different from the temperate *Myoviridae* phages
11 ϕ C2, ϕ CD119, ϕ CD27, ϕ CD630-1 and ϕ CD630-2 described previously (17, 19, 39, 45).
12 The complete genome sequence of ϕ CD6356, the first *cos*-type temperate *Siphoviridae*
13 phage from *C. difficile*, has been published recently (24). Phage ϕ CD6356 showed strong
14 protein similarity to the *C. difficile* strain QCD-63q42 over the lysis, lysogeny control
15 and DNA replication, recombination and modification modules (24). Thus, ϕ CD6356 and
16 ϕ CD38-2 are related over these modules, but unrelated over their structural genes. The
17 gene organization and modular structure of ϕ CD38-2 genome is similar to other
18 *C. difficile* phages and other temperate phages infecting low G+C bacteria (29). The only
19 difference in ϕ CD38-2 is the apparent lack of a clearly defined lysogeny control module
20 between the lysis and DNA replication modules (see below).

21

1

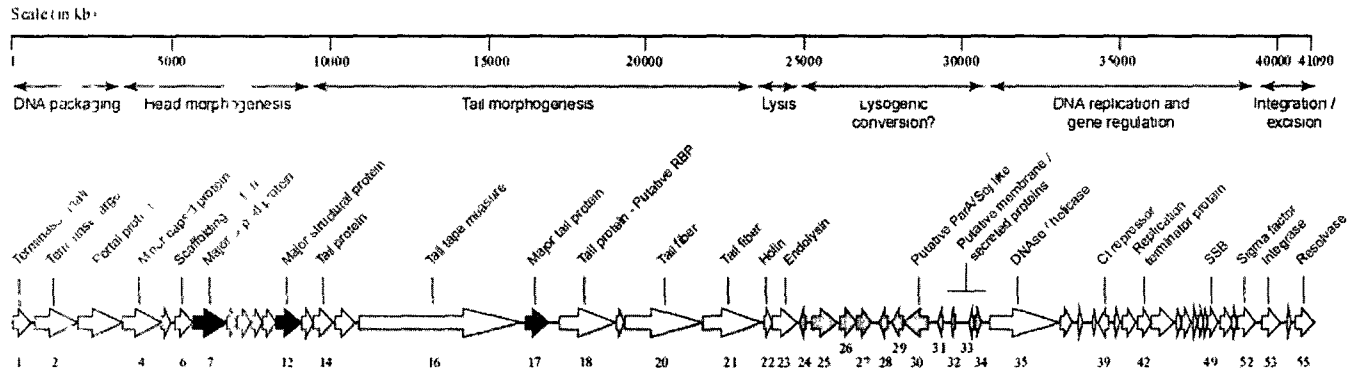


FIG. 1. Genetic organization of the complete ϕ CD38-2 genome (41,090 bp). Predicted ORFs with their respective orientation are represented by arrows. Functional assignments are indicated above each ORF, along with functional modules that were inferred based on gene annotation and whole genomic organization. Black arrows correspond to proteins identified by LC-MS/MS analysis while gray arrows refer to a possible horizontally acquired genomic region. RBP: Receptor-Binding Protein, SSB: Single-Strand Binding protein.

TABLE S2. Putative coding sequences from ϕ CD38-2 and bioinformatics analyses of the deduced proteins

ORF	Start	End	Size (bp)	Size (aa)	RBS	Predicted fonction ^a	BLAST hit accession number [source organism] ^b	Alignment (%ID)	Size (aa) ^c	E-Value
1	29	736	708	235	gaaaggagTTG	Terminase (small)	ZP_05399284 [C. difficile QCD-37x79] YP_001308765.1 [C. beijerinckii NCIMB 8052]	235/235 (100%) 98/236 (41%)	235 260	6E-133 6E-35
2	723	2045	1320	440	agaagaaatagaggcagaaatatATG	Terminase (large)	ZP_05399283 [C. difficile QCD-37x79] ZP_02640253 [C. perfringens CPE str. F4969]	440/440 (100%) 287/440 (65%)	440 443	0E+00 3E-166
3	2046	3458	1413	470	aagaggtgactaaATG	Minor capsid protein	ZP_05399282 [C. difficile QCD-37x79] NP_695169 [Lactobacillus phage phig1e]	470/470 (100%) 111/434 (25%)	470 508	0E+00 3E-35
4	3459	4691	1233	410	gacaaggggtgatgtagATG	Minor capsid protein	ZP_05399281 [C. difficile QCD-37x79] YP_001974361 [Streptococcus phage PH15]	410/410 (100%) 64/274 (23%)	410 381	0E+00 1E-05
5	4684	4968	285	94	aaaggagagaaaaATG	Phage-related protein	ZP_05399280 [C. difficile QCD-37x79] ZP_01821446 [Streptococcus pneumoniae SP6-BS73]	94/94 (100%) 37/88 (42%)	94 87	6E-48 1E-10
6	5098	5670	573	190	aggaggttttATG	Scaffolding protein	ZP_05399279 [C. difficile QCD-37x79] NP_469454 [Listeria innocua Clip11262]	190/190 (100%) 61/196 (31%)	190 189	6E-101 9E-09
7	5681	6754	1074	357	aggaggtagaaaATG	Phage coat protein	ZP_05399278 [C. difficile QCD-37x79] YP_001391656 [C. botulinum F str. Langeland]	357/357 (100%) 95/345 (27%)	357 321	0E+00 3E-23
8	6758	7090	333	110	gagagggtaggctATG		ZP_05399277 [C. difficile QCD-37x79] ZP_02640259 [C. perfringens CPE str. F4969]	110/110 (100%) 47/114 (41%)	110 114	5E-57 1E-12
9	7092	7526	435	144	gaggtgttagtGTG		ZP_05399276 [C. difficile QCD-37x79] ZP_02616048 [C. botulinum Bf]	144/144 (100%) 58/143 (40%)	144 136	2E-76 2E-21
10	7533	7880	348	115	ggaggcagaataatgggATG		ZP_05399275 [C. difficile QCD-37x79] ZP_02633339 [C. perfringens E str. JGS1987]	115/115 (100%) 59/115 (51%)	115 117	2E-61 1E-22
11	7877	8308	432	143	ggaggtagaagggaATG		ZP_05399274 [C. difficile QCD-37x79] ZP_02616050 [C. botulinum Bf]	143/143 (100%) 70/141 (49%)	143 143	2E-78 5E-37
12	8311	9111	800	266	aaaggagaaatgtaaaATG		ZP_05399273 [C. difficile QCD-37x79] ZP_02640263 [C. perfringens CPE str. F4969]	266/266 (100%) 73/156 (46%)	266 168	2E-152 3E-29
13	9127	9489	363	120	aagaggggctaaggaaATG		ZP_05399272 [C. difficile QCD-37x79] ZP_02616053 [C. botulinum Bf]	120/120 (100%) 49/123 (39%)	120 114	3E-63 5E-14
14	9489	10094	606	201	aacgtagagctaaaaataATG		ZP_05399271 [C. difficile QCD-37x79] ZP_02616054 [C. botulinum Bf]	201/201 (100%) 119/202 (58%)	201 197	3E-113 3E-63
15	10168	10809	642	213	gggggggattgttATG		ZP_05332086 [C. difficile QCD-37x79] ZP_02949175 [C. butyricum 5521]	213/213 (100%) 31/72 (43%)	213 362	8E-119 1E-06
16	10920	16070	5151	1716	aaaaaggggtgtagATG	Tail tape measure	ZP_05399269 [C. difficile QCD-37x79] ZP_02640311 [C. perfringens CPE str. F4969]	1716/1716 (100%) 178/702 (25%)	1716 933	0E+00 4E-62
17	16191	16925	735	244	gagaggaggcataaagctATG	Tail component protein	ZP_05399268 [C. difficile QCD-37x79] YP_001787259 [C. botulinum A3 str. Loch Maree]	244/244 (100%) 87/241 (36%)	244 225	3E-137 4E-29
18	17285	19054	1770	589	agaaagtttgaaTTG	Tail protein /	ZP_05399267 [C. difficile QCD-37x79] ZP_02621516 [C. botulinum C str. Eklund]	589/589 (100%) 81/245 (33%)	589 568	0E+00 1E-24
19	19081	19323	243	80	aaaaaggagaagaagtATG		ZP_05399266 [C. difficile QCD-37x79] YP_699982 [C. phage phiSM101]	80/80 (100%) 27/75 (36%)	80 82	3E-38 0.022

ORF	Start	End	Size (bp)	Size (aa)	RBS	Predicted fonction ^a	BLAST hit accession number [source organism] ^b	Alignment (%ID)	Size (aa) ^c	E-Value
20	19338	21800	2463	820	aaaaggagtgataatTTG	Tail fiber protein	ZP_05399319 [C difficile QCD-37x79] ZP_782154 [C tetani E88]	476/478 (99%) 45/149 (30%)	478 660	0E+00 3E-05
21	21800	23650	1851	616	agaagaggtgagtttaATG	Tail fiber protein	ZP_05399318 [C difficile QCD-37x79] YP_719584 [Haemophilus somnus 129PT]	616/616 (100%) 67/238 (28%)	616 762	0E+00 4E-13
22	23736	23999	264	87	aggaggaaagttATG	Holin	ZP_05399317 [C difficile QCD-37x79] YP_001087874 [C difficile 630]	87/87 (100%) 75/83 (90%)	87 86	3E-41 7E-35
23	24000	24806	807	268	agaagaaggcttttaATG	Endolysine	ZP_05399316 [C difficile QCD-37x79] YP_002290910 [C difficile phage phiCD27]	268/268 (100%) 148/277 (53%)	268 270	1E-154 2E-67
24	25067	24855	213	70	aagagaggtgtgaaaaATG		ZP_05399315 [C difficile QCD-37x79] ZP_02086761 [C bolteae ATCC BAA 613]	70/70 (100%) 21/68 (30%)	70 70	5E-29 4E-06
25	25246	26076	831	276	aagggggttaagcaagaATG		ZP_05399314 [C difficile QCD-37x79] YP_699976 [Clostridium phage phiSM101]	276/276 (100%) 50/127 (39%)	276 137	1E-152 2E-13
26	26155	26661	507	168	ggaggattgaagtATG		ZP_05399313 [C difficile QCD-37x79] ZP_05399427 [C difficile QCD-37x79]	168/168 (100%) 154/168 (91%)	168 168	7E-90 2E-81
27	26681	27151	471	156	ggaggataaaaaATG		ZP_05331942 [C difficile QCD-63q42]	156/156 (100%)	156	9E-84
28	27665	27384	282	93	aaaaaggggttgaaagATG		ZP_05399311 [C difficile QCD-37x79] YP_001088968 [C difficile 630]	92/93 (98%) 31/85 (36%)	93 491	2E-46 2E-03
29	28141	27809	333	110	agaaggaggtgttagATG		ZP_05399310 [C difficile QCD-37x79]	106/110 (96%)	110	3E-51
30	28951	28142	807	269	gaggtggagaaaATG	ParA/SOJ-like protein	ZP_05331939 [C difficile QCD-37x79] YP_313580 [Spiroplasma citri]	265/269 (98%) 76/268 (28%)	269 256	1E-151 2E-19
31	29420	29238	183	60	aaaaaggagctaataATG		ZP_05399308 [C difficile QCD-37x79] ZP_05350752 [C difficile ATCC 43255]	60/60 (100%) 28/57 (49%)	60 61	3E-27 3E-08
32	29788	29660	129	42	agaaggaggaacgggaATG	Membrane protein ^d	ZP_05399307 [C difficile QCD-37x79]	42/42 (100%)	42	8E-13
33	30391	30239	153	50	aaaggaggtgatgaagtATG	Membrane protein ^d	ZP_05399306 [C difficile QCD-37x79] YP_001087437 [C difficile 630]	50/50 (100%) 36/53 (67%)	50 53	5E-20 4E-11
34	30432	30617	183	61	gggagataatcttTTG	Secreted protein ^d	ZP_05399305 [C difficile QCD-37x79]	61/61 (100%)	61	5E-24
35	30911	33076	2163	721	ggagggataaagATG	DNase	ZP_05399304 [C difficile QCD-37x79] YP_659583 [C difficile 630]	721/721 (100%) 375/703 (53%)	721 869	0E+00 0E+00
36	33123	33470	348	115	aaggagataaaaaggATG		ZP_05399303 [C difficile QCD-37x79]	115/115 (100%)	115	4E-57
37	33683	33793	111	36	agaaaaacagaggttaTTG		ZP_05331922 [C difficile QCD-63q42]	36/36 (100%)	36	2E-10
38	34150	34245	93	31	aggggatgattggTTG		ZP_05399302 [C difficile QCD-37x79]	31/31 (100%)	31	3E-07
39	34643	34290	354	117	gaaaggagtttaaaatg	CI phage repressor	ZP_05399301 [C difficile QCD-37x79] ZP_02633304 [C perfringens E str JGS1987]	117/117 (100%) 35/99 (35%)	117 129	1E-60 2E-09
40	34816	34998	183	60	aggaggtgtttctATG		ZP_05399300 [C difficile QCD-37x79]	60/60 (100%)	60	3E-24
41	35065	35490	426	141	aaggaggtggaagATG		ZP_05399299 [C difficile QCD-37x79]	141/141 (100%)	141	5E-74
42	35565	35969	405	134	agaagggggttttaaaagATG	Replication terminator	ZP_05399298 [C difficile QCD-37x79] YP_535158 [Lactobacillus phage Sal2]	134/134 (100%) 42/120 (35%)	134 138	1E-70 5E-10
43	35992	36705	714	237	aaaaaggagttttaatATG		ZP_05399297 [C difficile QCD-37x79] NP_782698 [C tetani E88]	237/237 (100%) 120/228 (52%)	237 238	1E-133 1E-55
44	36761	36925	165	54	aaaaaggaggtttattTTG		ZP_05399296 [C difficile QCD-37x79]	54/54 (100%)	54	5E-22

ORF	Start	End	Size (bp)	Size (aa)	RBS	Predicted fonction ^a	BLAST hit accession number [source organism] ^b	Alignment (%ID)	Size (aa) ^c	E-Value
45	36922	37233	312	103	aggaggctgggcaATG		ZP_05399295 [C. difficile QCD-37x79]	103/103 (100%)	103	2E-50
46	37279	37428	150	49	aggaggctgggcaATG		ZP_05399294 [C. difficile QCD-37x79]	49/49 (100%)	49	1E-19
47	37459	37632	174	57	gaatggagggtataaaATG		ZP_05399293 [C. difficile QCD-37x79]	57/57 (100%)	57	2E-25
48	37632	37748	117	38	gaaaggaatgacgaaatataATG		YP_529609 [C. difficile phage phi CD119]	31/57 (54%)	57	2E-13
							ZP_05399292 [C. difficile QCD-37x79]	38/38 (100%)	38	5E-13
49	37745	38071	327	108	aagggggagggttgaaATG	SSB	ZP_05401366 [C. difficile QCD-23m63]	22/38 (57%)	40	6E-05
							ZP_05399291 [C. difficile QCD-37x79]	108/108 (100%)	108	7E-56
50	38163	38528	363	121	aagtaggtgattaaATG		YP_001090187 [C. difficile 630]	73/108 (67%)	144	4E-37
							ZP_05399290 [C. difficile QCD-37x79]	121/121 (100%)	121	6E-65
51	38512	38685	174	57	aagtaggtgattaaATG		YP_002290935 [C. difficile phage phiCD27]	49/122 (40%)	121	2E-10
							ZP_05399289 [C. difficile QCD-37x79]	57/57 (100%)	57	3E-23
52	38697	39269	573	190	aggaagtgaatttATG	Sigma70/SimgaF-like	ZP_05399288 [C. difficile QCD-37x79]	190/190 (100%)	190	4E-103
53	39447	40055	609	202	gagaaggtagaataaGTG	Integrase	ZP_05399287 [C. difficile QCD-37x79]	202/202 (100%)	202	3E-113
							YP_001885139 [C. botulinum B str. Eklund 17B]	122/198 (61%)	212	6E-63
54	40265	40372	108	35	ggtgaagcttaaggacATG		ZP_05332102 [C. difficile QCD-63q42]	43/43 (100%)	43	4E-16
							ZP_05399286 [C. difficile QCD-37x79]	34/35 (97%)	43	2E-11
55	40521	41090	567	189	agtaaggggtgttttttATG	Resolvase	ZP_05399285 [C. difficile QCD-37x79]	189/189 (100%)	189	2E-107
							YP_001090102 [C. difficile 630]	114/185 (61%)	189	1E-58

^a Inferred from BLASTp analyses on protein sequences in GenBank and ACLAME databases, and by the presence of conserved domains found through searches in the conserved domains database (CDD) at NCBI and by InterProScan analysis

^b The best hit is shown along with another relevant hit when available (excluding hits from the QCD-63q42 strain)

^c Size in amino acids (aa) of the BLAST hit (which may be used to calculate the overall %ID)

1 **Structural proteins**

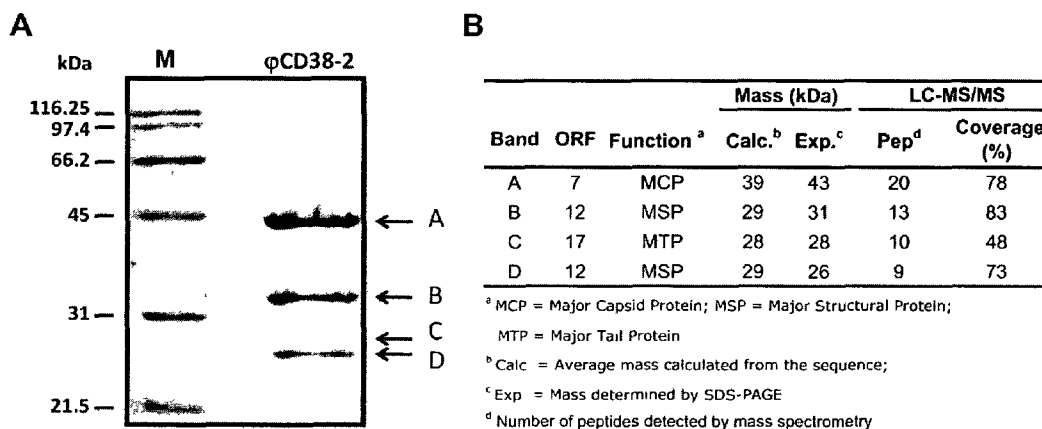
2 The *orfs* encoding structural proteins were clearly identified and were bordered on the
3 left side by the packaging module comprising ORF1, ORF2 and ORF3 that code for the
4 small and large terminase, and the portal protein, respectively, and on the right side by
5 the lysis module comprising ORF22 and ORF23 that code for a putative holin and an
6 endolysin, respectively (described below). Excluding the similar prophages found in the
7 genomes of *C. difficile* strains QCD-63q42 and QCD-37x79, protein BLAST analyses
8 showed that most structural proteins were related to other *Clostridium* genomic
9 sequences, including *C. perfringens*, *C. botulinum* and *C. tetani* (Table S2). ORF18
10 corresponds to a tail protein with a putative endopeptidase activity and conserved
11 domains (TIGR01665, pfam06605, COG4926) were found suggesting that *orf18*
12 probably encodes the phage receptor-binding protein (RBP) responsible for host
13 specificity.

14

15 Structural proteins from CsCl-purified ϕ CD38-2 particles were separated by denaturing
16 SDS-PAGE, followed by LC-MS/MS analysis of the trypsin-digested proteins. Seven
17 bands were cut out of the gel and analyzed by mass spectrometry (Fig. 2). Three major
18 bands were observed on SDS-PAGE (Fig. 2A): band A corresponded to ORF7, with 78%
19 protein coverage, bands B and D corresponded to ORF12, with 83% and 73% coverage,
20 respectively while band C corresponded to ORF17. The experimental and calculated
21 masses were similar for the 4 bands and peptide mapping didn't reveal any evidence of
22 post-translational proteolytic processing (Fig. 2B). Thus, based on local genomic
23 organization and BLAST analyses, ORF7 was annotated as the major capsid protein

1 (MCP) and ORF17 as the major tail protein (MTP). Because the boundary between the
 2 capsid and tail morphogenesis modules could not be clearly defined, ORF12 could be
 3 either a capsid or a tail protein and as a consequence, it was annotated as a major
 4 structural protein (MSP) without reference to any particular virion structure (Fig. 2B).
 5 Three minor bands corresponding to ribosomal proteins were identified, likely the result
 6 of a contamination with bacterial proteins despite two successive rounds of discontinuous
 7 CsCl gradient.

8



9

10 FIG. 2. Analysis of ϕ CD38-2 structural proteins. (A) Coomassie brilliant blue staining
 11 of a 12% SDS-polyacrylamide gel showing ϕ CD38-2 structural proteins along with a
 protein molecular mass marker (lane M). Arrows and letters on the right correspond to
 12 proteins from panel (B) that were identified by LC-MS/MS analysis.

12

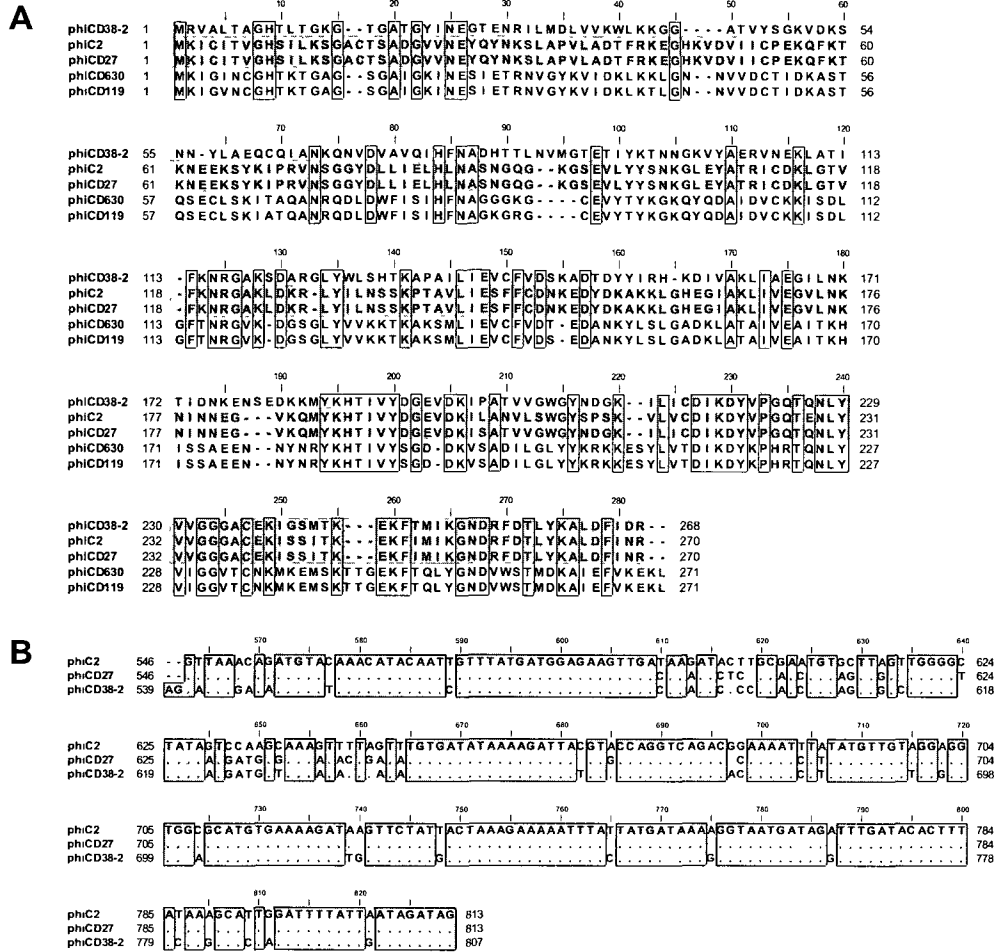
13 Lysis module

14 No putative function could be readily attributed to ORF22 based on BLAST hits alone.

15 However, the size of the protein, the presence of a transmembrane segment and several
 16 charged amino acids in C-terminus, as well as its location right upstream of the

1 endolysin, are all signatures of phage holins. In line with this, conserved domains
2 [MurNAc-LAA (cd02696) and Amidase_3 (pfam01520)] were found in ORF23, which is
3 likely a phage endolysin with N-acetylmuramoyl-L-alanineamidase activity (NAM-
4 amidase). ORF23 was also found to be similar to the endolysin from phage ϕ CD27 (53%
5 identity, 148/277 residues aligned, $2e-67$). A multiple protein alignment with endolysins
6 from *C. difficile* phages for which a sequence was available revealed two distinct
7 homology groups: ϕ CD630-2 and ϕ CD119 endolysins are nearly identical and form one
8 group, while ϕ C2 and ϕ CD27 endolysins are also nearly identical and form another
9 distinct group (Fig. 3A). Interestingly, the N-terminal part of ϕ CD38-2 endolysin
10 diverges from the other 4 endolysins but the last ~60 C-terminal residues are almost
11 100% identical to the C-terminal domain of ϕ C2 and ϕ CD27 endolysins. In line with this,
12 the last 267 nucleotides in the 3' end of the endolysin genes were very well conserved
13 (Fig. 3B). Taken together, these results suggest that a recombination event occurred
14 between different phage endolysins, leading to domain swapping (either the C-terminal
15 CBD or the N-terminal catalytic domain) and supporting the modular structure of phage
16 endolysins (22).

17



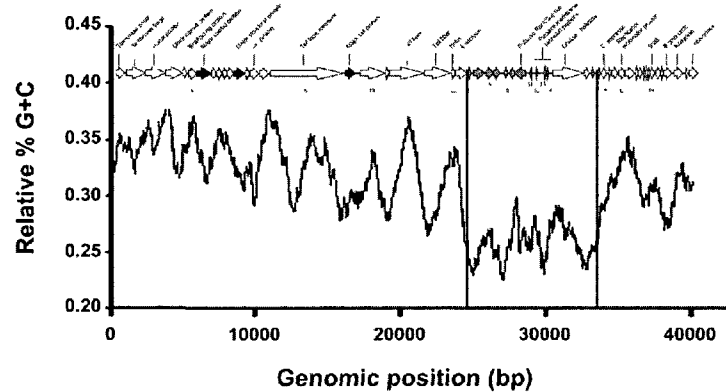
1

2

FIG. 3. (A) Multiple protein alignment of phage endolysins. Boxed residues are identical in all 5 endolysins aligned. Grey shading was used to highlight the similarity between ϕ C2, ϕ CD27 and the C-terminal part of ϕ CD38-2 endolysins. (B) DNA alignment of the last 267 nucleotides of the 3' end of ϕ C2, ϕ CD27 and ϕ CD38-2 endolysins. Dots represent identical nucleotides when compared to the ϕ C2 sequence as the reference. Boxed nucleotides are 100% identical in the 3 genes aligned.

1 DNA replication, transcription and gene regulation

2 As opposed to the structural genes for which a putative function could be assigned to 15
3 of the 23 *orfs* (65%), only a few of the 32 non-structural genes (*orf24* to *orf55*) had
4 homologs in public databases (except the two contigs from McGill described earlier).
5 Hence, a putative function could be assigned to only 8 ORFS (25%), of which 4 could be
6 related to DNA replication, transcription or gene regulation. ORF35 contains several
7 conserved domains (DEAD, pfam00270 3e-6; ResIII, pfam04851, 2e-5; SSL2,
8 COG1061, 6e-8) that are found in DNA helicases, and also in type III restriction
9 enzymes. Several BLASTp hits corresponding to deoxyribonucleases (DNase) were
10 retrieved, including CDP07 (YP_659583, 53% ID, 375/703 residues aligned, E-value =
11 0.0) (Table S2), a putative DNase found on plasmid pCD630 (NC_008226.1) carried by
12 *C. difficile* strain 630 (45). We performed a nucleotide BLAST analysis of this region on
13 the NCBI server and found a region extending from position 31,112 to 32,991 in ϕ CD38-
14 2 sharing 65% identity with a region from plasmid pCD630 extending from nucleotides
15 3,580 to 4,568, which corresponds to $\sim 2/3$ of the gene coding for a DNase. The same
16 analysis was performed with ϕ C2, ϕ CD27 and ϕ CD119 to see if other *C. difficile* phages
17 contained plasmid-related DNA. Only ϕ CD119 shared significant homology (68%
18 nucleotide identity over 794 bp) with plasmid pCKL555A (CP000674.1) from *C. kluyveri*
19 DSM 555. This region from the phage covers nucleotides 46,728 to 47,522 and codes for
20 a methyltransferase (19). In addition, the ϕ CD38-2 genome shows a marked deviation in
21 its G+C content from *orf24* to *orf34*, where the average G+C content is $25.6\% \pm 1.2$
22 while it is $31.6\% \pm 3.1$ in the rest of the genome (Fig. 4). Such deviations are often
23 indicators of recent horizontal DNA acquisition events.



1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

FIG. 4. Relative G+C content of ϕ CD38-2. The average % G+C was calculated on a 10-base window along the whole phage genome and values were plotted against the nucleotide position on the genome. The region with a lower %G+C compared with the rest of the genome is indicated by vertical lines and the corresponding region in the ϕ CD38-2 genome is shown above the graph line.

ORF42 was similar to a phage replication terminator protein, and ORF49 contained conserved domains (SSB_OBF, cd04496 9e-22; SSB, pfam00436, 7e-25) suggesting that it is a single-strand DNA binding protein (SSB). ORF52 contained a conserved domain corresponding to region 3 and 4 of the Sigma70 family of proteins (IPR013324; Sigma70-ECF, TIGR02937, 6e-5). Region 3 interacts with DNA upstream of the -10 element of certain “extended-10” promoters (43), like in Gram-positive promoters where a conserved TG dinucleotide is found upstream of the -10 element (35). Region 4 interacts with the -35 element (43). The role of this putative Sigma factor remains to be determined.

Lysogeny control

A lysogeny module could not be clearly defined in ϕ CD38-2, as opposed to other low G+C temperate phages and *C. difficile* phages ϕ C2, ϕ CD27 and ϕ CD119, where the

1 phage repressors CI, Cro, and other antirepressors, as well as the integrase and resolvase
2 are clustered together between the lysis module and the DNA replication/modification
3 module (17, 19, 29, 39). In ϕ CD38-2, a different organization was observed. For example
4 ORF39, a putative CI repressor based on BLAST results and on the presence of DNA-
5 binding domains (HTH_XRE, cd00093, 4e-06; HTH_3, pfam01381, 7e-6) was found
6 approximately 10 kb downstream of the lysis cassette. ORF53, a phage integrase of the
7 tyrosine recombinase/integrase family (INT_SG3, cd01186, 3e-29; Phage integrase,
8 pfam00589, 5e-11), was found 5 kb downstream of *orf39* near ORF55, a protein of the
9 SR serine recombinase family (SR_ResInv, cd03768, 8e-31; Resolvase, pfam00239, 6e-
10 30) that is likely a resolvase. Finally, we could not find a putative Cro-like repressor
11 despite scrutinizing the genome for related proteins with HTH-containing DNA binding
12 domains.

13 **Putative lysogenic conversion genes**

14 Most of the differences found between ϕ CD38-2 and the QCD-37x79 and QCD-63q42
15 contigs were located in a region of about 1.5 kb extending from nucleotide 27,500 to
16 29,000 of the ϕ CD38-2 genome. This region encompasses a part of *orf28*, *orf29*, *orf30*
17 and the end of *orf31*. BLAST analyses failed to show extensive similarity with proteins
18 from public databases, except for the already mentioned phage sequences from strain
19 QCD-37x79 and QCD-63q42 (Table S2). Interestingly, ORF30 contained several
20 conserved domains (ParA cd02042, 1e-3; Soj COG1192, 1e-18; CbiA pfam01656, 4e-9;
21 SopA PHA02519, 3e-6) found in proteins with ATPase activity generally associated with
22 chromosome partitioning (cell division) as well as plasmid maintenance in bacteria. A
23 *parA* homolog was also reported in the genomes of ϕ C2 and ϕ CD6356 (17, 24) but the

1 exact function of this gene in *C. difficile* phages is unknown. Also, InterProScan analyses
 2 predicted the presence of a signal peptide and/or transmembrane regions within ORF32,
 3 ORF33, and ORF34. This suggests that ORF32 and ORF33 could potentially be targeted
 4 to the membrane whereas ORF34 could be secreted.

5

6 Putative Sigma70-like promoter sequences were found in ϕ CD38-2 upstream of a
 7 number of genes in the region from *orf24* to *orf40*, as well as upstream of the integrase
 8 (ORF53) and resolvase (ORF55) (Table 2).

9

Table 2: Putative sigma70-like promoters identified in the genome of ϕ CD38-2

<i>orf</i> ^a	Coordinates	Strand	Promoter sequence (5'→3') ^b
Gram-positive consensus ^c			(A-rich) TTGACA -- (16-18bp) --- TG N TATAAT
24	25091 - 25119	-	TT TAC TTTTAATTTTTTATG T TAT TAT
25	25182 - 25210	+	gaaaaagt gTTGAC TTAGATATTCATATAGAT TATAAT
32	29845 - 29872	-	aaaaataggagaaaact aTTG TTTTGACGACCCGTC ATGG TATAAT
34	29944 - 29974	+	aaaa TTG TATCTTTTTTAAAAAATG TGA TATAAT
34	30151 - 30180	+	aaag TTG TATCTTTTAAAGAAAATG TGA TATAAT
35	30771 - 30799	+	TTG TTTTCTACTTATAT T TAT TAT TAT
33	30795 - 30823	-	aaaaccaa TTG TGTTTTCTACACTT TGT CT TATAAT
38	34049 - 34077	+	TTG TATATATCGAACAAAAGGG GG TATAAT
39	34667 - 34695	-	TTGAC TTACGGGCAACTTACTG TAA AAT
40	34749 - 34777	+	aaaagtta TTGAC TTGCCCTTATTACGCAC TATAAT
53	39391 - 39418	+	TTG GGCTTTTTATAATGCAAAT TAA TAT AT
55	40439 - 40467	+	TTGAC TACGATTTAATTATG TAG TAA AAT

^a *orf* located downstream the putative promoter.

^b Nucleotides that fit the consensus sequence are boxed.

^c Based on the report by Mani *et al.* (29). A typical TG dinucleotide is often found upstream the -10 box in Gram-positive bacteria, as well as an A-rich region upstream the -35 box.

10

11

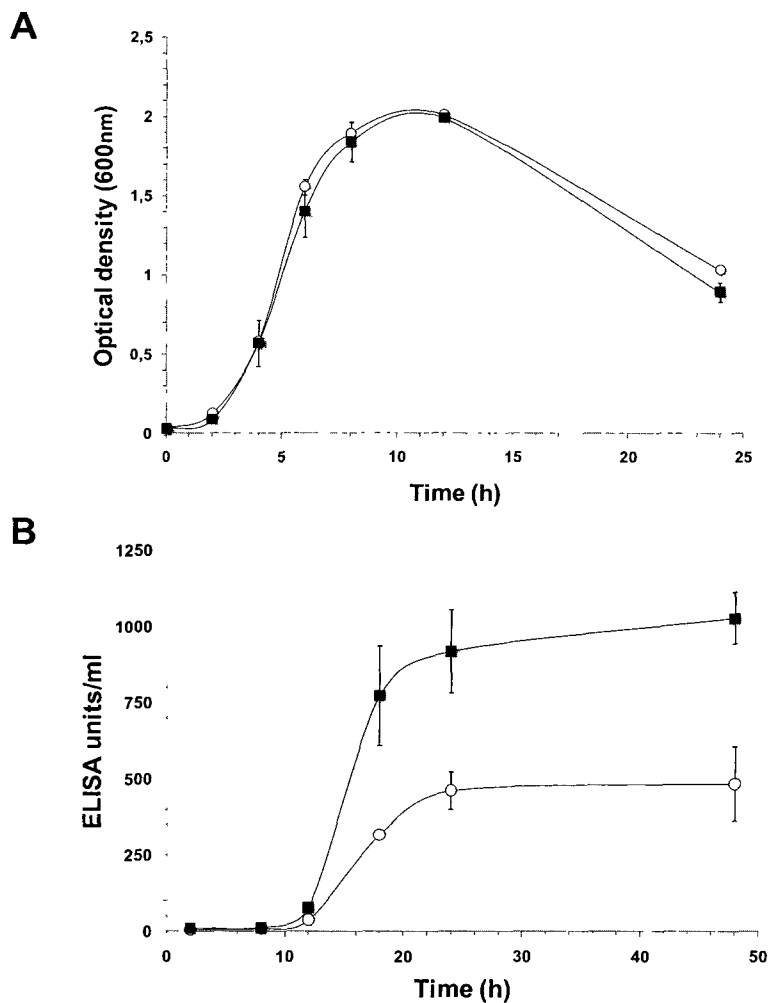
12 The presence of these putative promoters suggests that expression of the downstream
 13 genes could be constitutive during the lysogenic cycle under normal growth. To test this
 14 hypothesis, we extracted total RNA from exponential and stationary phase cells of

1 CD274/φCD38-2 lysogens, and performed RT-PCR on a number of genes downstream of
2 these promoters, including *orfs* 25, 27, 29, 30, 42 and 53 to see if they were expressed.
3 We included *orfs* 7, 16, and 23 as controls since structural and lysis genes should
4 normally not be transcribed when the prophage is latent. All genes turned out to be
5 expressed under normal growth conditions (data not shown), including the structural
6 genes. The most probable explanation for this is that a fraction of the prophages had
7 undertaken the lytic growth cycle. Spontaneous induction is well known to occur with
8 many temperate phages (4, 6, 33). After growing a CD274/φCD38-2 lysogen for 8 h at
9 37°C in BHI broth, up to $\sim 10^5$ PFU/mL were detected in filtered culture supernatants (not
10 shown), confirming that φCD38-2 was able to readily excise and initiate a lytic growth
11 cycle without any particular inducing treatment. Thus, the background spontaneous
12 prophage induction made impossible to confirm whether the identified vegetative
13 promoters were active or not during the lysogenic state. A similar observation was
14 reported for phage LE1 from *Leptospira biflexa* (4).

15 **Prophage-stimulated toxin production in φCD38-2 lysogens**

16 Previous reports have shown that toxin production in *C. difficile* was increased upon
17 lysogenization with phages φC2, φC6 and φC8 (16) and was decreased in the presence of
18 φCD119 (20). In order to test whether φCD38-2 could also influence toxin production in
19 *C. difficile*, we introduced φCD38-2 into the chromosome of its host strain CD274, which
20 is a representative member of the hypervirulent strain BI/NAP1/027. The growth profile
21 in TY broth and total biomass yield after 24 h were not significantly different between
22 the CD274/φCD38-2 lysogen and the wild-type parental strain (Fig. 5A).

23



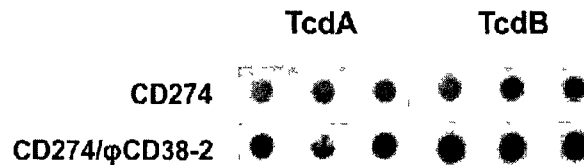
1
2 FIG. 5. Growth and toxin production in the wild-type CD274 (open circles) and a
3 CD274/φCD38-2 lysogen (black squares). (A) Growth in TY broth was monitored by
4 taking the optical density at 600 nm over a 24 h period. (B) The amounts of TcdA and
5 TcdB toxins in cleared culture supernatants were detected by a Toxin A/B ELISA
6 assay. Data represent the mean \pm SEM (Standard Error of the Mean) of three
7 independant biological replicates.

5

6 Aliquots of culture supernatants were collected at different time intervals over a 48 h
7 period and toxins were detected using a commercial ELISA assay. Total toxins
8 accumulated faster and to a higher level in culture supernatants from the lysogen
9 compared to the wild-type strain. After 12 h of growth, supernatants from the

1 CD274/ ϕ CD38-2 lysogen contained 2.1-fold more toxins compared to wild-type CD274
 2 (Fig. 5B). We then used an immunodot blot approach to specifically detect TcdA and
 3 TcdB. The production of both toxins was increased in the lysogen after 24 h of growth,
 4 and TcdB seemed to accumulate to a slightly higher level (Fig. 6). Densitometry analysis
 5 of the dots showed that the amount of TcdA and TcdB was \sim 1.6- and 2.1-fold higher in
 6 the CD274/ ϕ CD38-2 lysogen compared to the wild-type CD274 strain, respectively. This
 7 was consistent with the results obtained with the ELISA assay (Fig. 5B).

8

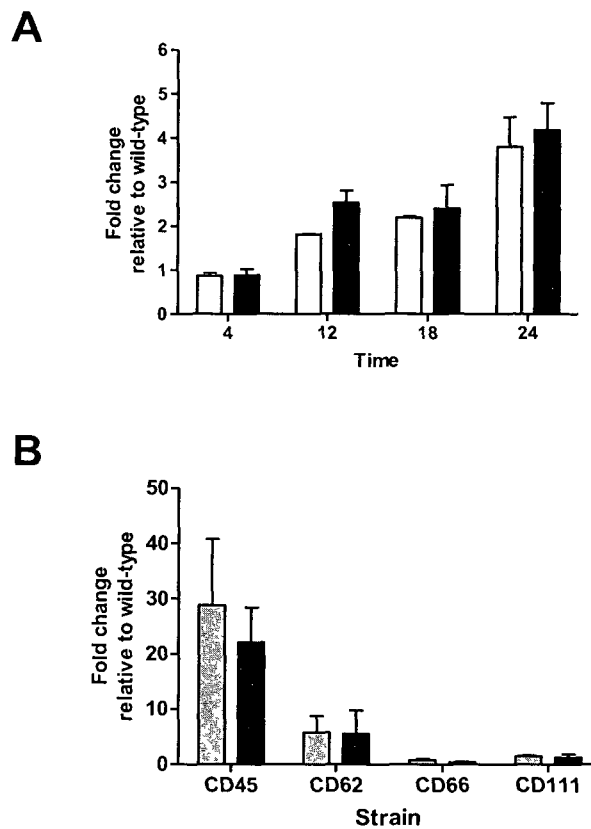


9

10 FIG. 6. Immuno-dot blot detection of toxins in cleared supernatants from 24 h cultures of
 11 wild-type CD274 and a CD274/ ϕ CD38-2 lysogen. TcdA and TcdB toxins were detected
 12 using monoclonal anti-TcdA and anti-TcdB antibodies. Each spot of a series represents
 13 an independent biological experiment.

13

14 We then performed a relative quantitative real-time RT-PCR assay to compare the levels
 15 of *tcdA* and *tcdB* mRNA in both strains. Total RNA was extracted from the
 16 CD274/ ϕ CD38-2 lysogen and the wild-type parental strain CD274 at 4, 12, 18 and 24 h
 17 post-inoculation. The levels of *tcdA* and *tcdB* mRNA were similar in both lysogenic and
 18 wild-type strains after 4 h of growth (ratio of \sim 1) but after 12 h, the levels in the lysogen
 19 were 1.8 and 2.5 higher for *tcdA* and *tcdB*, respectively (Fig. 7A). After 24 h, the lysogen
 20 expressed 3.8 and 4.2-fold more *tcdA* and *tcdB* mRNA, respectively. These data are in
 21 agreement with the toxin levels that were detected by ELISA and immunodot blot.



1

2

3

4

5

6

7

8

9

10

FIG.7. Relative levels of *tcdA* (gray bars) and *tcdB* (black bars) mRNA in ϕ CD38-2 lysogens versus their respective wild-type parental strain, as determined by qRT-PCR analysis. (A) Toxin expression in wild-type CD274 versus a CD274/ ϕ CD38-2 lysogen at different time points. Data are expressed as the fold change in gene expression in the ϕ CD38-2 lysogen relative to the wild-type strain and the data represent the mean \pm SEM of two independent biological replicates. A value of 1 means there is no difference in mRNA levels in the lysogen compared to the wild-type strain. (B) *tcdA* and *tcdB* mRNA levels at 24 h in four different lysogens relative to their wild-type parental strain. Data are presented the same way as in (A).

Four additional ϕ CD38-2 lysogens were created in different genetic backgrounds to verify if the above observations made with the CD274 lysogen could be extended to other *C. difficile* isolates. Phage ϕ CD38-2 was introduced into strains CD45 (PCR ribotype

1 035), CD62 (PCR ribotype 014), as well as CD66 and CD111 (both PCR ribotype 027).
2 Toxins A and B were detected in culture supernatants using the ELISA assay and only
3 the CD45 lysogen showed a 2.5-fold increase in extracellular toxins after 48 h compared
4 to the non-lysogenic parental strain. Toxin levels in the supernatants of the other three
5 lysogens were not significantly different (data not shown). The amounts of *tcdA* and *tcdB*
6 transcripts were determined by real-time qRT-PCR analysis of total RNA extracted from
7 these lysogens and the corresponding parental strain after 24 h of growth in TY broth.
8 The detected mRNA levels were higher in the CD45/ ϕ CD38-2 lysogen with a 22 and 29-
9 fold increase in *tcdA* and *tcdB* mRNA, respectively (Fig. 7B). The CD62, CD66 and
10 CD111 lysogens showed on average a 5, 0.5 and 1.5-fold difference in toxin A/B mRNA
11 levels compared to the parental strain, respectively. Five colonies were picked-up at the
12 end of each culture experiment and the presence of the prophage was confirmed by PCR
13 in all cases, thus excluding prophage loss as a possible explanation for the observed
14 variability in toxin production from one lysogen to another. Taken together, our results
15 demonstrate that ϕ CD38-2 can stimulate toxin production in some lysogens, including
16 the NAP1/027 epidemic strain, by increasing mRNA transcription and/or stability, and
17 this effect seems to be strain dependant.

18

19

20

DISCUSSION

21 The molecular basis for the hypervirulence and hypertoxin-producing phenotype of the
22 NAP1/027 epidemic strain is still unclear and it is reasonable to presume that prophages
23 might be part of the answer. Here we report the microbiological and molecular

1 characterisation of ϕ CD38-2, a temperate phage of the *Siphoviridae* family infecting
2 *Clostridium difficile* and our study provides further evidence that temperate phages play a
3 significant role in the evolution and virulence of this important pathogen.

4
5 So far, only temperate phages have been isolated and characterized after mitomycin C or
6 UV induction of *C. difficile* lysogens and strictly lytic phages have not yet been
7 described. *C. difficile* is found mainly in its sporulated form in the environment and
8 spores are insensitive to phage infection. It is thus very unlikely for a phage to infect
9 *C. difficile* outside a mammalian host, which probably explains why they have adopted a
10 temperate life style for better survival. *C. difficile* phages seem to have a rather narrow
11 host spectrum, and a given phage generally infects only a few different strains (32, 46). In
12 our study, ϕ CD38-2 was able to infect *C. difficile* isolates representing 11 different PCR
13 ribotypes and 48% of a collection of 207 *C. difficile* isolates (Table 1). A more detailed
14 analysis revealed that only a limited number of isolates within each PCR ribotype were
15 sensitive to ϕ CD38-2. What was most interesting however, was that ϕ CD38-2 was able
16 to infect 89% of all NAP1/027 isolates tested in our collection (79/89). The fact that 10%
17 of the NAP1/027 isolates were insensitive to ϕ CD38-2 suggests that significant genetic
18 differences exist between the sensitive and insensitive isolates. It is now well established
19 that a given PCR ribotype, including 027, can be divided into subtypes using highly
20 discriminatory typing methods such as MLVA (12, 48, 49). Thus, it is possible that the
21 isolates that were insensitive to ϕ CD38-2 corresponded to different subtypes. A very
22 likely explanation would be that insensitive isolates were lysogenized by another related
23 prophage, thus making them immune to ϕ CD38-2 infection. Another possible

1 explanation would be that the bacterial receptors for ϕ CD38-2 adsorption were absent in
2 the insensitive isolates. Nevertheless, these results further support the idea that PCR
3 ribotyping, although a very rapid, reliable and cheap method, is insufficiently
4 discriminatory to reveal subtle genetic differences that may significantly alter the host
5 phenotype. Our study also suggests that ϕ CD38-2 represents a very interesting candidate
6 to study phage-host interactions in a clinically relevant *C. difficile* strain.

7

8 **First genome sequence of a *pac*-type *Siphoviridae* phage infecting *C. difficile***

9 The genomes of only a few phages of *C. difficile* have been fully sequenced so far. The
10 diversity of genomic sequences available at the moment is thus very limited and any new
11 phage sequence is highly desirable. In addition, all phage sequences represent members
12 of the *Myoviridae* family that are related genetically (17, 19, 39, 45). The only exception
13 is phage ϕ CD6356, a *cos*-type phage of the *Siphoviridae* family that was recently
14 characterized and sequenced (23). This phage was shown to be genetically unrelated to
15 the previously identified *Myoviridae* phages. In a previous study, we have shown by
16 DNA hybridization experiments that ϕ CD38-2 was genetically distinct from *Myoviridae*
17 phages that seemed to form an homogenous group (13). The complete genomic sequence
18 of ϕ CD38-2 was determined and comparisons using BLAST tools confirmed our
19 previous observation. We also found ϕ CD38-2 in 3 other isolates (out of 30 analyzed),
20 suggesting that the corresponding prophage is relatively frequent in *C. difficile*.
21 Interestingly, an almost identical prophage was also identified in the chromosome of two
22 *C. difficile* strains being sequenced by McGill University (QCD-37x79 and QCD-63q42).
23 Only a few genomic differences were found between these prophages and ϕ CD38-2.

1 Also noteworthy to mention, almost half of the ϕ CD6356 phage (from *orf27* to *orf56*)
2 was also found to be very similar to the prophage in strain QCD-63q42, but with lower
3 identity values compared to ϕ CD38-2 (41-95% amino acid identity) (24). Thus, ϕ CD38-2
4 and ϕ CD6356 are related over their lysis and DNA replication / gene regulation modules,
5 but different in terms of structural genes. The fact that ϕ CD38-2 is a *pac*-type
6 *Siphoviridae* phage whereas ϕ CD6356 is a *cos*-type phage also suggests that they are part
7 of two distinct phage families that package their DNA using two different mechanisms.
8 Interestingly, comparison of the endolysin gene of ϕ CD38-2 with that of other *C. difficile*
9 phages revealed the possibility of a past recombination event within this gene (Fig. 3).
10 The endolysin and integrase genes are generally well conserved and they probably act as
11 recombination hot spots. Thus, *C. difficile* phages could potentially exchange whole
12 genomic fragments corresponding to the lysis, lysogeny and DNA replication / gene
13 regulation modules using endolysin and integrase / recombinase genes as anchor points.
14 Supporting this, the similarity of ϕ CD6356 with the QCD-63q42 starts around the
15 endolysin gene (*orf27*) and stops at a site-specific recombinase (*orf57*).

16 **Putative lysogenic conversion genes**

17 In other *C. difficile* phages and phages infecting low G+C Gram-positive bacteria, the
18 lysogeny module is generally located between the lysis cassette and the DNA replication
19 and regulation module. This region generally encodes Cro and CI repressors,
20 transcriptional regulators and antirepressors as well as the integrase (17, 19, 24, 29, 39).
21 In ϕ CD38-2, lysogeny-related genes were not clustered within a particular region and
22 putative *cI* repressor and integrase genes were found, but far outside this genomic region
23 (Fig. 1). Instead of a lysogeny module, we found a number of *orfs* that could potentially

1 represent lysogenic conversion genes. In addition, we identified a number of putative
2 vegetative promoters that could potentially drive transcription of these genes during
3 lysogeny (Table 2). However, our transcriptional analyses using RT-PCR were
4 unsuccessful due to background levels of spontaneous induction of ϕ CD38-2 that led to
5 the expression of lytic genes. A similar situation was reported for phage LE1 of
6 *Leptospira biflexa* (4).

7

8 A ParA homolog (ORF30) similar to a *Spiroplasma citri* Soj-like protein was found in
9 ϕ CD38-2 (Table S2). ParA/Soj-like proteins, Spo0J/ParB-like proteins and *parS*
10 centromere-like sequences are involved in chromosome segregation and plasmid
11 maintenance (36). A functional *parABS* system was also reported in the genome of the
12 temperate phage LE1 from *Leptospira biflexa*(4). Interestingly, ORF41 and ORF42 of
13 *C. difficile* phage ϕ C2 were found to share some low similarity with ORF1 and ORF2,
14 the ParA and ParB homologs in phage LE1, respectively (17). The genes were located
15 between the lysis cassette and the DNA replication / gene regulation module (17). It was
16 proposed that this gene product could help maintain the ϕ C2 prophage in a
17 pseudolysogenic state. In ϕ CD6356, a ParA homolog was also identified, but no ParB
18 (like in ϕ CD38-2). The authors suggested that this ParA/Soj homolog could participate in
19 sporulation, since Soj acts as a sporulation initiation inhibitor in the absence of a Spo0J
20 (ParB homolog) (24, 25). Additional experiments will be needed to verify if either of
21 these two hypotheses is true. Interestingly, the presence near *parA* of a 1.5 kb fragment
22 sharing 65% identity at the DNA level with the pCD630 plasmid from *C. difficile* strain

1 630 suggests that this possible lysogenic conversion region in ϕ CD38-2 could have been
2 acquired through a past recombination event with a plasmid.

3 **Prophage-stimulated toxin production in NAP/027 lysogens**

4 TcdR and TcdC are positive and negative regulators of toxin production in *C. difficile*,
5 respectively (51). A number of deletions were identified in *tcdC*, and in particular a 1-bp
6 deletion at position 117 that leads to the synthesis of a severely truncated TcdC protein in
7 the NAP1/027 epidemic strain (8, 30). This deletion is thought to be responsible for the
8 increased toxin production reported in this strain (52). However, recent studies suggest
9 that deletions in *tcdC* alone cannot explain hypertoxin production and hypervirulence of
10 NAP1/027 isolates (41, 50). The regulation of toxin production in *C. difficile* thus seems
11 to be complex and other mechanisms are likely involved in this process. A previous
12 report by Goh *et al.* suggested that lysogens carrying the temperate phages ϕ C2, ϕ C6 and
13 ϕ C8 could modify toxin production in *C. difficile*. They noted that toxin mRNA levels
14 and the amount of toxins detected by ELISA in culture supernatants were not always
15 correlating (16). Interestingly, the PaLoc shares some sequence similarity with phage
16 proteins, in particular TcdE, suggesting that it is probably the remains of an ancient
17 prophage. It also suggests that phage regulatory networks could possibly be intertwined
18 with those of the PaLoc (16, 47). Further evidence supporting a possible interconnection
19 between prophages and the PaLoc was recently provided by Govind *et al.* who showed
20 that during lysogeny, the RepR transcriptional regulator encoded by ϕ CD119 was able to
21 bind promoter regions in the PaLoc, causing a downregulation of the expression of all
22 PaLoc genes, including TcdA and TcdB toxins (20).

23

1 Lysogenization of CD274, a *C. difficile* isolate corresponding to the hypervirulent strain
2 NAP1/027, with ϕ CD38-2 led to a 1.6 and 2.5-fold increase in toxin A and B detected in
3 culture supernatants, respectively (Fig. 5 and 6). Furthermore, we demonstrated by a real
4 time qRT-PCR approach that *tcdA* and *tcdB* mRNA levels were higher in lysogens
5 carrying ϕ CD38-2 (Fig. 7). Our results also suggest that depending on the genetic
6 background of the host strain, the impact of ϕ CD38-2 on toxin transcription and
7 production is variable. These observations are in agreement with data reported by Goh *et*
8 *al.* who also demonstrated variable toxin production in different hosts lysogenized with
9 the same phage (16). Our study and those of Govind *et al.* and Goh *et al.* also show that
10 depending on the phage-host system, the impact of different phages on toxin production
11 is variable. Unfortunately, we were unsuccessful in our attempts to identify the ϕ CD38-2
12 integration site in these lysogens, probably because of spontaneous prophage induction
13 yielding phage DNA concatemers that prevent identification of phage-host DNA
14 junctions using inverted PCR and hybridization methods. As proposed by Goh *et al.* (17),
15 we cannot rule out the possibility that ϕ CD38-2 leads to pseudolysogeny, where the
16 prophage would be maintained as a circular DNA molecule or would alternate between
17 an integrated and an episomic form. The presence of a ParA homolog and of plasmid-
18 related sequences in ϕ CD38-2 supports this hypothesis. So it is currently impossible to
19 explain the observed strain-dependant variation in toxin production. We cannot exclude
20 the possibility that ϕ CD38-2 integrated at different sites in these lysogens, thereby
21 affecting the hosts differently. It is also possible that other endogenous prophages
22 (expressing a RepR-like repressor for example) could interfere with toxin gene regulation
23 too, thereby masking the effect of ϕ CD38-2. The exact mechanism by which ϕ CD38-2

1 influences toxin production is currently under investigation. Nevertheless, the present
2 study and that of Govind *et al.* clearly demonstrate that prophages can influence the
3 transcription and expression of virulence genes in *trans*.

4 **Conclusion**

5 In summary, we characterized and sequenced the first genome of a *pac*-type *Siphoviridae*
6 phage infecting *C. difficile*. This phage is able to infect several NAP1/027 isolates
7 corresponding to the hypervirulent strain that recently caused severe outbreaks in North
8 America and Europe. Complete genome sequencing didn't reveal the presence of
9 identifiable virulence factors, but lysogenization of NAP1/027 with ϕ CD38-2 can lead to
10 increased toxin production. ϕ CD38-2 is able to infect several other PCR ribotypes and
11 was found in a number of other *C. difficile* isolates. This phage is capable of spontaneous
12 induction and thus, can disseminate across the *C. difficile* species. Because ϕ CD38-2 has
13 the capacity to increase *C. difficile* virulence through modulation of toxin expression
14 makes of this phage a very interesting model to study phage-host interactions in a
15 clinically relevant bacterial pathogen.

16

17

ACKNOWLEDGMENTS

18 We are grateful to Louis Valiquette for providing clinical isolates of *C. difficile*. We also
19 thank Pier-Luc Dudemaine for helping with the host range analysis. This study was
20 supported by the Centre de recherch  Clinique  tienne-Le Bel, by a discovery grant from
21 the Natural Sciences and Engineering Council of Canada (NSERC) and by a seed grant

1 from the Canadian Institutes of Health Research (CIHR). LCF is a research scholar from
2 the Fonds de la Recherche en Santé du Québec (FRSQ).

3

4

REFERENCES

- 5 1. **Ackermann, H.** 2007. 5500 Phages examined in the electron microscope. *Arch.*
6 *Viro.* **152**:227-243.
- 7
- 8 2. **Altschul, S., T. Madden, A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D.**
9 **Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein
10 database search programs. *Nucleic Acids Res.* **25**:3389-3402.
- 11
- 12 3. **Bidet, P., F. Barbut, V. Lalande, B. Burghoffer, and J. Petit.** 1999.
13 Development of a new PCR-ribotyping method for *Clostridium difficile* based on
14 ribosomal RNA gene sequencing. *FEMS Microbiol. Lett.* **175**:261-266.
- 15
- 16 4. **Bourhy, P., L. Frangeul, E. Couvé, P. Glaser, I. Saint Girons, and M.**
17 **Picardeau.** 2005. Complete nucleotide sequence of the LE1 prophage from the
18 spirochete *Leptospira biflexa* and characterization of its replication and partition
19 functions. *J. Bacteriol.* **187**:3931-3940.
- 20
- 21 5. **Brussow, H., C. Canchaya, and W. Hardt.** 2004. Phages and the evolution of
22 bacterial pathogens: from genomic rearrangements to lysogenic conversion.
23 *Microbiol. Mol. Biol. Rev.* **68**:560-602.
- 24
- 25 6. **Calendar, R.** 2006. *The bacteriophages*, R. Calendar Ed. ed. Oxford University
26 Press, New York.
- 27
- 28 7. **Canchaya, C., G. Fournous, and H. Brussow.** 2004. The impact of prophages
29 on bacterial chromosomes. *Mol. Microbiol.* **53**:9-18.
- 30
- 31 8. **Curry, S., J. Marsh, C. Muto, M. O'Leary, A. Pasculle, and L. Harrison.**
32 2007. *tdcC* genotypes associated with severe TcdC truncation in an epidemic
33 clone and other strains of *Clostridium difficile*. *J. Clin. Microbiol.* **45**:215-221.
- 34
- 35 9. **Dei, R.** 1989. Observations on phage-typing of *Clostridium difficile*: preliminary
36 evaluation of a phage panel. *Eur. J. Epidemiol.* **5**:351-354.
- 37
- 38 10. **Denève, C., S. Bouttier, B. Dupuy, F. Barbut, A. Collignon, and C. Janoir.**
39 2009. Effects of subinhibitory concentrations of antibiotics on colonization factor
40 expression by moxifloxacin-susceptible and moxifloxacin-resistant *Clostridium*
41 *difficile* strains. *Antimicrob. Agents. Chemother.* **53**:5155-5162.

- 1 11. **Dupuy, B., R. Govind, A. Antunes, and S. Matamouros.** 2008. *Clostridium*
2 *difficile* toxin synthesis is negatively regulated by TcdC. *J. Med. Microbiol.*
3 **57**:685-689.
4
- 5 12. **Fawley, W., P. Parnell, P. Verity, J. Freeman, and M. Wilcox.** 2005.
6 Molecular epidemiology of endemic *Clostridium difficile* infection and the
7 significance of subtypes of the United Kingdom epidemic strain (PCR ribotype 1).
8 *J Clin Microbiol* **43**:2685-96.
9
- 10 13. **Fortier, L.-C., and S. Moineau.** 2007. Morphological and genetic diversity of
11 temperate phages in *Clostridium difficile*. *Appl. Environ. Microbiol.* **73**:7358-
12 7366.
13
- 14 14. **Fortier, L., A. Bransi, and S. Moineau.** 2006. Genome sequence and global
15 gene expression of Q54, a new phage species linking the 936 and c2 phage
16 species of *Lactococcus lactis*. *J. Bacteriol.* **188**:6101-6114.
17
- 18 15. **Fortier, L., and S. Moineau.** 2009. Phage production and maintenance of stocks,
19 including expected stock lifetimes. *Methods Mol. Biol.* **501**:203-219.
20
- 21 16. **Goh, S., B. Chang, and T. Riley.** 2005. Effect of phage infection on toxin
22 production by *Clostridium difficile*. *J. Med. Microbiol.* **54**:129-135.
23
- 24 17. **Goh, S., P. Ong, K. Song, T. Riley, and B. Chang.** 2007. The complete genome
25 sequence of *Clostridium difficile* phage phiC2 and comparisons to phiCD119 and
26 inducible prophages of CD630. *Microbiology* **153**:676-685.
27
- 28 18. **Goh, S., T. Riley, and B. Chang.** 2005. Isolation and characterization of
29 temperate bacteriophages of *Clostridium difficile*. *Appl. Environ. Microbiol.*
30 **71**:1079-1083.
31
- 32 19. **Govind, R., J. Fralick, and R. Rolfe.** 2006. Genomic organization and molecular
33 characterization of *Clostridium difficile* bacteriophage phiCD119. *J. Bacteriol.*
34 **188**:2568-2577.
35
- 36 20. **Govind, R., G. Vedyappan, R. Rolfe, B. Dupuy, and J. Fralick.** 2009.
37 Bacteriophage mediated toxin gene regulation in *Clostridium difficile*. *J. Virol.*
38 **83**:12037-12045.
39
- 40 21. **Gravel, D., M. Miller, A. Simor, G. Taylor, M. Gardam, A. McGeer, J.**
41 **Hutchinson, D. Moore, S. Kelly, D. Boyd, M. Mulvey, and C. N. I. S.**
42 **Program.** 2009. Health care-associated *Clostridium difficile* infection in adults
43 admitted to acute care hospitals in Canada: a Canadian Nosocomial Infection
44 Surveillance Program Study. *Clin. Infect. Dis.* **48**:568-576.
45

- 1 22. **Hermoso, J. A., J. L. García, and P. García.** 2007. Taking aim on bacterial
2 pathogens: from phage therapy to enzybiotics. *Curr. Opin. Microbiol.* **10**:461-472.
3
- 4 23. **Horgan, M., O. O'Sullivan, A. Coffey, G. F. Fitzgerald, D. van Sinderen, O.**
5 **McAuliffe, and R. P. Ross.** 2010. Genome analysis of the *Clostridium difficile*
6 phage PhiCD6356, a temperate phage of the Siphoviridae family. *Gene*.
7
- 8 24. **Horgan, M., O. O'Sullivan, A. Coffey, G. F. Fitzgerald, D. van Sinderen, O.**
9 **McAuliffe, and R. P. Ross.** 2010. Genome analysis of the *Clostridium difficile*
10 phage phiCD6356, a temperate phage of the *Siphoviridae* family. *Gene:In press*.
11
- 12 25. **Ireton, K., N. Gunther, and A. Grossman.** 1994. *spo0J* is required for normal
13 chromosome segregation as well as the initiation of sporulation in *Bacillus*
14 *subtilis*. *J. Bacteriol.* **176**:5320-5329.
15
- 16 26. **Kuijper, E., B. Coignard, and P. Tull.** 2006. Emergence of *Clostridium*
17 *difficile*-associated disease in North America and Europe. *Clin. Microbiol. Infect.*
18 **12 Suppl 6**:2-18.
19
- 20 27. **Leplae, R., A. Hebrant, S. J. Wodak, and A. Toussaint.** 2004. ACLAME: a
21 CLAssification of Mobile genetic Elements. *Nucleic Acids Res.* **32**:D45-49.
22
- 23 28. **Loo, V., L. Poirier, M. Miller, M. Oughton, M. Libman, S. Michaud, A.**
24 **Bourgault, T. Nguyen, C. Frenette, M. Kelly, A. Vibien, P. Brassard, S. Fenn,**
25 **K. Dewar, T. Hudson, R. Horn, P. Rene, Y. Monczak, and A. Dascal.** 2005. A
26 predominantly clonal multi-institutional outbreak of *Clostridium difficile*-
27 associated diarrhea with high morbidity and mortality. *N. Engl. J. Med.* **353**:2442-
28 2449.
29
- 30 29. **Lucchini, S., F. Desiere, and H. Brussow.** 1999. Similarly organized lysogeny
31 modules in temperate *Siphoviridae* from low GC content gram-positive bacteria.
32 *Virology* **263**:427-435.
33
- 34 30. **MacCannell, D. R., T. J. Louie, D. B. Gregson, M. Laverdiere, A.-C. Labbe,**
35 **F. Laing, and S. Henwick.** 2006. Molecular analysis of *Clostridium difficile* PCR
36 ribotype 027 isolates from Eastern and Western Canada. *J. Clin. Microbiol.*
37 **44**:2147-2152.
38
- 39 31. **Mahony, D., P. Bell, and K. Easterbrook.** 1985. Two bacteriophages of
40 *Clostridium difficile*. *J. Clin. Microbiol.* **21**:251-254.
41
- 42 32. **Mahony, D., J. Clow, L. Atkinson, N. Vakharia, and W. Schlech.** 1991.
43 Development and application of a multiple typing system for *Clostridium difficile*.
44 *Appl. Environ. Microbiol.* **57**:1873-1879.
45

- 1 33. **Maiques, E., C. Ubeda, S. Campoy, N. Salvador, I. Lasa, R. P. Novick, J.**
2 **Barbé, and J. R. Penadés.** 2006. Beta-lactam antibiotics induce the SOS
3 response and horizontal transfer of virulence factors in *Staphylococcus aureus*. *J.*
4 *Bacteriol.* **188**:2726-2729.
5
- 6 34. **Mani, N., and B. Dupuy.** 2001. Regulation of toxin synthesis in *Clostridium*
7 *difficile* by an alternative RNA polymerase sigma factor. *Proc. Natl. Acad. Sci.*
8 *USA* **98**:5844-5849.
9
- 10 35. **Mani, N., D. Lyras, L. Barroso, P. Howarth, T. Wilkins, J. Rood, A.**
11 **Sonenshein, and B. Dupuy.** 2002. Environmental response and autoregulation of
12 *Clostridium difficile* TxeR, a sigma factor for toxin gene expression. *J. Bacteriol.*
13 **184**:5971-5978.
14
- 15 36. **Marston, A., and J. Errington.** 1999. Dynamic movement of the ParA-like Soj
16 protein of *B. subtilis* and its dual role in nucleoid organization and developmental
17 regulation. *Mol. Cell* **4**:673-682.
18
- 19 37. **Matamouros, S., P. England, and B. Dupuy.** 2007. *Clostridium difficile* toxin
20 expression is inhibited by the novel regulator TcdC. *Mol. Microbiol.* **64**:1274-
21 1288.
22
- 23 38. **Matte, I., D. Lane, E. Côté, A. Asselin, L. Fortier, C. Asselin, and A. Piché.**
24 2009. Anti-apoptotic proteins Bcl-2/Bcl-XL inhibit *Clostridium difficile* toxin A-
25 induced cell death in human epithelial cells. *Infect. Immun.* **77**:5400-5410.
26
- 27 39. **Mayer, M. J., A. Narbad, and M. J. Gasson.** 2008. Molecular characterization
28 of a *Clostridium difficile* bacteriophage and its cloned biologically active
29 endolysin. *J. Bacteriol.* **190**:6734-6740.
30
- 31 40. **Moineau, S., S. Pandian, and T. R. Klaenhammer.** 1994. Evolution of a lytic
32 bacteriophage via DNA acquisition from the *Lactococcus lactis* chromosome.
33 *Appl. Environ. Microbiol.* **60**:1832-1841.
34
- 35 41. **Murray, R., D. Boyd, M. Mulvey, P. Levett, and M. Alfa.** 2009. Truncation in
36 the *tcdC* region of the *Clostridium difficile* PathLoc of clinical isolates does not
37 predict increased biological activity of Toxin B or Toxin A. *BMC Infect. Dis.*
38 **9**:103.
39
- 40 42. **Nagy, E., and J. Foldes.** 1991. Electron microscopic investigation of lysogeny of
41 *Clostridium difficile* strains isolated from antibiotic-associated diarrhea cases and
42 from healthy carriers. *Acta Pathol. Microbiol. Immunol. Scand.* **99**:321-326.
43
- 44 43. **Paget, M. S. B., and J. D. Helmann.** 2003. The sigma70 family of sigma factors.
45 *Genome Biol.* **4**:203.

- 1 44. **Pepin, J., L. Valiquette, and B. Cossette.** 2005. Mortality attributable to
2 nosocomial *Clostridium difficile*-associated disease during an epidemic caused by
3 a hypervirulent strain in Quebec. *CMAJ* **173**:1037-1042.
4
- 5 45. **Sebahia, M., B. Wren, P. Mullany, N. Fairweather, N. Minton, R. Stabler, N.**
6 **Thomson, A. Roberts, A. Cerdeno-Tarraga, H. Wang, M. Holden, A. Wright,**
7 **C. Churcher, M. Quail, S. Baker, N. Bason, K. Brooks, T. Chillingworth, A.**
8 **Cronin, P. Davis, L. Dowd, A. Fraser, T. Feltwell, Z. Hance, S. Holroyd, K.**
9 **Jagels, S. Moule, K. Mungall, C. Price, E. Rabbino-witsch, S. Sharp, M.**
10 **Simmonds, K. Stevens, L. Unwin, S. Whithead, B. Dupuy, G. Dougan, B.**
11 **Barrell, and J. Parkhill.** 2006. The multidrug-resistant human pathogen
12 *Clostridium difficile* has a highly mobile, mosaic genome. *Nat. Genet.* **38**:779-
13 786.
14
- 15 46. **Sell, T., D. Schaberg, and F. Fekety.** 1983. Bacteriophage and bacteriocin
16 typing scheme for *Clostridium difficile*. *J. Clin. Microbiol.* **17**:1148-1152.
17
- 18 47. **Tan, K., B. Wee, and K. Song.** 2001. Evidence for holin function of *tcdE* gene in
19 the pathogenicity of *Clostridium difficile*. *J. Med. Microbiol.* **50**:613-619.
20
- 21 48. **Tanner, H. E., K. J. Hardy, and P. M. Hawkey.** 2010. Co-existence of multiple
22 MLVA sub-types of *Clostridium difficile* PCR-Ribotype 027 strains within faecal
23 specimens. *J. Clin. Microbiol.* *In press*
24
- 25 49. **van den Berg, R., I. Schaap, K. Templeton, C. Klaassen, and E. Kuijper.**
26 2007. Typing and Subtyping of *Clostridium difficile* Isolates by Using Multiple-
27 Locus Variable-Number Tandem-Repeat Analysis. *J Clin Microbiol* **45**:1024-8.
28
- 29 50. **Verdoorn, B., R. Orenstein, J. Rosenblatt, L. Sloan, C. Schleck, W. Harmsen,**
30 **L. Nyre, and R. Patel.** 2009. High prevalence of *tcdC* deletion-carrying
31 *Clostridium difficile* and lack of association with disease severity. *Diagn.*
32 *Microbiol. Infect. Dis.* **66**:24-28.
33
- 34 51. **Voth, D. E., and J. D. Ballard.** 2005. *Clostridium difficile* toxins: mechanism of
35 action and role in disease. *Clin. Microbiol. Rev.* **18**:247-263.
36
- 37 52. **Warny, M., J. Pepin, A. Fang, G. Killgore, A. Thompson, J. Brazier, E. Frost,**
38 **and L. McDonald.** 2005. Toxin production by an emerging strain of *Clostridium*
39 *difficile* associated with outbreaks of severe disease in North America and Europe.
40 *The Lancet* **366**:1079-1084.
41
42
43

3 DISCUSSION

3.1 Rappel des objectifs du projet

La bactérie *Clostridium difficile* est maintenant considérée comme la principale cause de diarrhées associées aux infections bactériennes en milieu hospitalier (BARBUT *et al.*, 1996; JOHNSON et GERDING, 1998). En mars 2003, un changement épidémiologique important quant au taux d'incidence des infections à *C. difficile* (ICD) a été observé (PEPIN *et al.*, 2004). Ce changement dans l'épidémiologie a été relié à l'émergence d'une souche hypervirulente, BI/NAP1/027, associée à un taux de mortalité accru chez les patients hospitalisés ainsi qu'à des symptômes de la maladie plus sévères (PEPIN *et al.*, 2004; LOO *et al.*, 2005). Cette émergence d'un clone hypervirulent a fait l'objet de plusieurs études visant à déterminer les facteurs impliqués dans l'évolution de *C. difficile* (STABLER *et al.*, 2006; STABLER *et al.*, 2009; HE *et al.*, 2010). Le séquençage du génome de la souche de *C. difficile* CD630 a révélé une multitude d'éléments génétiques mobiles, témoignant de l'acquisition fréquente d'ADN étranger par transfert horizontal (SEBAIHIA *et al.*, 2006). Parmi les éléments génétiques mobiles, on dénombre des transposons conjugatifs, transposons mobilisables, IStrons ainsi que deux prophages inducibles. Les bactériophages tempérés sont reconnus comme de véritables moteurs évolutifs à court terme au sein des espèces bactériennes (BRUSSOW et HENDRIX, 2002; CANCHAYA *et al.*, 2003). Leur impact au niveau de la biologie, l'évolution et

la virulence chez de nombreuses espèces, pathogènes ou non, ont été très bien documentées (BOYD et BRUSSOW, 2002; BRUSSOW *et al.*, 2004). Par contre, très peu de données concrètes sont disponibles en ce qui concerne les bactériophages chez *C. difficile* et leur contribution sur la virulence bactérienne est encore mal comprise. Nous avons donc voulu explorer davantage cette voie en tentant de démontrer que les phages de *C. difficile* ont un rôle à jouer dans la biologie bactérienne. Dans cet ordre d'idées, plusieurs prophages ont été induits à partir de nombreux isolats cliniques de *C. difficile*. Un phage en particulier, le siphophage ϕ CD38-2, a été choisi pour une étude microbiologique et moléculaire complète. Son impact sur la croissance et la production de toxines chez *C. difficile* a été évalué *in vitro* et nos résultats appuient l'hypothèse que les phages ont réellement un impact sur la virulence de cet important pathogène.

3.2 Isolement de phages tempérés de *Clostridium difficile*

Des études préliminaires ont montré une forte prévalence de phages tempérés chez *C. difficile*. Par exemple, huit prophages ont été induits et isolés à partir de six isolats cliniques de *C. difficile* suite aux traitements avec la mitomycine C (FORTIER et MOINEAU, 2007). Un phage en particulier, ϕ CD38-2, a retenu notre attention due à quelques caractéristiques intéressantes. Premièrement, il s'agit d'un phage faisant partie de la famille des *Siphoviridae*; les membres de cette famille n'ont été que très peu caractérisés chez *C. difficile* à ce jour. De plus, jusqu'à tout

récemment, il n'y avait pas de séquence génomique disponible pour ce type de phage dans les bases de données publiques. Deuxièmement, les inductions préliminaires aux rayons UV ont révélées la présence de ce phage dans plusieurs isolats cliniques différents, suggérant une forte capacité de dissémination. Finalement, le phage ϕ CD38-2 est capable d'infecter et de lyser la souche hypervirulente BI/NAP1/027, ce qui fait de ce phage un modèle très intéressant pour étudier l'impact des phages chez une souche importante de *C. difficile*.

3.3 Caractérisation microbiologique et moléculaire du phage ϕ CD38-2

Le premier volet de cette étude impliquait la caractérisation de ce phage au niveau microbiologique et moléculaire. Préalablement, il a été établi que le phage ϕ CD38-2 faisait partie de la famille des *Siphoviridae* et que sa séquence en acides nucléiques différait considérablement des autres phages isolés auparavant (FORTIER et MOINEAU, 2007). Suite à ces observations, nous avons isolé et purifié le phage ϕ CD38-2 et une étude du spectre d'hôte a été réalisée à l'aide d'une collection de plus de 200 isolats cliniques de *C. difficile* dans le but de déterminer la proportion des souches sensibles à ce phage.

3.3.1 Spectre d'hôte et dynamique d'infection du phage ϕ CD38-2

Le spectre d'hôte du phage ϕ CD38-2 a été déterminé au sein d'une collection de 207 isolats cliniques de *C. difficile*, repartis en 41 ribotypes différents. Il s'est avéré que près de la moitié des souches testées (99/207, 48%) étaient sensibles au phage ϕ CD38-2 et de ce nombre, la grande majorité (79 isolats, 80%) faisait partie du ribotype 027 correspondant à la souche hypervirulente. Ainsi, le phage ϕ CD38-2 est le premier phage décrit chez *C. difficile* à posséder la caractéristique intéressante de pouvoir infecter un pourcentage très élevé d'isolats de ribotype 027. Le reste des souches sensibles (20 isolats, 20%) étaient repartis dans 10 ribotypes différents. Dans la littérature, il est fréquent de constater que les phages ont un spectre d'hôte plutôt restreint et très souvent (mais pas exclusivement) limité à une seule espèce bactérienne. Ceci est dû à la nécessité des phages à reconnaître un récepteur particulier sur la surface bactérienne. Ainsi, un récepteur commun, reconnaissable par le phage ϕ CD38-2, semble présent sur la majorité des isolats de ribotype 027, permettant une infection généralisée de ce groupe particulier. L'insensibilité au phage ϕ CD38-2 de certains isolats de ribotype 027 pourrait s'expliquer par la présence possible d'un phage apparenté dans leur génome. Par ailleurs, plusieurs phages ont pu être induits à partir de divers isolats de ribotype 027, dont ϕ CD6 (FORTIER et MOINEAU, 2007) ainsi que ϕ CD111 et ϕ CD146 partiellement caractérisés par notre laboratoire (données non publiées). Tel que discuté dans l'introduction, la présence d'un phage apparenté dans un chromosome bactérien permet de bloquer l'infection de la souche lysogène par d'autres phages similaires. D'ailleurs, il semble maintenant

évident que le ribotypage ne donne qu'une vue partielle du contenu génétique d'une souche. En effet, plusieurs sous-types ont été identifiés au sein du ribotype 027 en utilisant des méthodes de typage plus discriminantes comme le MLVA (VAN DEN BERG *et al.*, 2007; FAWLEY *et al.*, 2008; TANNER *et al.*, 2010).

Il est à noter que l'étude du spectre d'hôte ne fournit aucune information quant à la dynamique d'infection du phage ϕ CD38-2. Les paramètres de la dynamique d'infection d'un phage (temps de latence, taille de la progéniture, etc.) servent à déterminer l'efficacité d'infection sur une souche bactérienne en particulier. Les paramètres de la dynamique d'infection dépendent de certains facteurs spécifiques tels que la reconnaissance adéquate du récepteur bactérien par le phage ou encore la présence ou l'absence de divers systèmes de résistance aux infections chez les bactéries (système de restriction/modification, système d'avortement de l'infection) (BOUCHARD *et al.*, 2002; ROBERTS *et al.*, 2007; DOMINGUES *et al.*, 2008; FINERAN *et al.*, 2009). Dans ce sens, les paramètres de la dynamique d'infection peuvent varier considérablement d'une souche à une autre. Dans notre cas, nous avons utilisé un isolat de ribotype 027 (CD274), puisque le phage ϕ CD38-2 est capable d'infecter la grande majorité des souches de ce groupe.

Sur cette souche particulière, il a été montré que le phage ϕ CD38-2 a un temps de latence de 95 minutes. Ce paramètre nous renseigne quant au temps nécessaire au phage pour compléter un cycle, de l'étape d'adsorption jusqu'au début de la lyse bactérienne et de la libération des premiers virions. Cette étape peut varier

considérablement d'un phage à un autre, et selon les conditions physiologiques de l'hôte et les conditions expérimentales (température de croissance, type de milieu de culture etc.). Chez *C. difficile*, des temps de latence variant entre 30 et 120 minutes ont été rapportés pour certains phages (MAHONY *et al.*, 1985; GOH *et al.*, 2005). Dans ce sens, le phage ϕ CD38-2 se situe dans la moyenne par rapport aux autres phages caractérisés.

À partir d'une courbe de croissance du phage représentant la dynamique d'infection, on peut également calculer la taille de la progéniture, c'est-à-dire le nombre de particules virales relâchées par bactérie infectée (ufp/cell), lorsque le phage est en mode lytique. Ce paramètre varie selon différentes conditions, tout comme le temps de latence. Pour le phage ϕ CD38-2, il a été calculé que chaque bactérie infectée relâche environ 50 particules virales à la fin du cycle lytique. D'autres phages tempérés de *C. difficile* semblent avoir des tailles de progéniture se situant de 5 à 122 ufp/cell (MAHONY *et al.*, 1985; GOH *et al.*, 2005). Le coliphage tempéré λ relâche environ 100 nouvelles particules virales suite à l'infection d'une cellule d'*E. coli* (CALENDAR, 2006). En opposé, les phages lytiques d'*E. coli* (série T1 à T7) ont des tailles de progéniture sensiblement plus élevées, se situant entre 120 et 300 ufp/cell, ce qui peut être expliqué par leur caractère strictement lytique et non tempéré (BROCK, 1990).

3.3.2 Séquençage et comparaisons génomiques du phage ϕ CD38-2

Les hybridations de type ADN-ADN ont montré la grande divergence génétique du phage ϕ CD38-2 par rapport aux autres phages tempérés de *C. difficile*. Donc, le séquençage du génome entier du ϕ CD38-2 s'est avéré utile afin d'élargir nos connaissances quant à la diversité génétique de ce groupe particulier de phages.

Le séquençage a permis de définir la taille du génome à 41,09 kb, ce qui est inférieur par rapport aux autres phages *Myoviridae* de *C. difficile* caractérisés jusqu'à présent (50,9kb à 56,5kb) (GOH *et al.*, 2005; GOVIND *et al.*, 2006; MAYER *et al.*, 2008) mais légèrement supérieur par rapport au siphophage ϕ CD6356 récemment caractérisé (37,7kb) (HORGAN *et al.*, 2010). L'écart entre les *Siphoviridae* et les *Myoviridae* est en grande partie dû aux différences dans le nombre de gènes composant les modules structuraux. Les *Myoviridae* ont généralement plus de gènes structuraux, entre autre à cause de la queue contractile qui est plus complexe.

Une comparaison de la séquence nucléotidique contre la base de données GenBank a révélé la présence de deux prophages quasi identiques (quelques nucléotides polymorphiques seulement (SNPs)) dans deux souches de *C. difficile* en cours de séquençage (QCD-37x79 and QCD-63q42). Cette observation suggère que le phage ϕ CD38-2 est relativement mobile et donc capable de passer d'une souche sensible de *C. difficile* à une autre. Sa distribution précise au sein d'une collection de souches de *C. difficile* reste à être déterminée. Des expériences préliminaires ont

permis de détecter un phage probablement identique à ϕ CD38-2 dans au moins trois autres isolats sur 30 analysés.

Au moment d'écrire ce mémoire, un phage (ϕ CD6356) de la famille *Siphoviridae* a été caractérisé chez *C. difficile* (HORGAN et al., 2010). Puisque sa séquence nucléotidique n'est pas disponible dans les bases de données à ce jour, la comparaison génomique précise avec le phage ϕ CD38-2 est impossible à effectuer. Par contre, la séquence nucléotidique du ϕ CD6356 est probablement très différente de celle du phage ϕ CD38-2 puisque l'annotation génomique semble différer considérablement. De plus, la taille du génome du ϕ CD6356 (37 664 pb) ne correspond pas avec celle du ϕ CD38-2 (41090 pb). Également, d'autres différences au niveau de la structure du génome sont observables entre les deux phages. Le phage ϕ CD6356 possède des extrémités cohésives (phage de type *cos*) qui servent de point de reconnaissance lors de l'encapsidation du génome. Le phage ϕ CD38-2 quant à lui utilise un moyen de type « *head-full packaging* » ce qui le place dans la catégorie des phages de type *pac*.

D'autre part, la faible similarité de séquence observée par rapport aux autres phages de *C. difficile* dont la séquence est disponible confirme les résultats d'hybridations obtenus précédemment (FORTIER et MOINEAU, 2007). Également, ces résultats suggèrent que le phage ϕ CD38-2 fait partie d'un sous-groupe de siphophages qui dérivent d'un ancêtre différent de celui des myophages caractérisés chez *C. difficile*. Ce siphophage est également distinct des siphophages ϕ CD24-1, ϕ CD8-1 et ϕ CD8-2 identifiés récemment (FORTIER et MOINEAU, 2007). La faible similarité de séquence avec les myophages a dû restreindre la recombinaison

homologue entre ces deux familles. Le séquençage d'autres phages de la famille des *Siphoviridae*, tel que ϕ CD24-1, ϕ CD8-1 et ϕ CD8-2 serait intéressant et permettrait de mieux évaluer la diversité et l'évolution des phages chez *C. difficile*. Toutefois, il faudra trouver des souches hôtes capables de propager ces phages, souches qui sont actuellement non disponibles.

3.3.3 Analyse globale du génome par bioinformatique

L'identification de 55 cadres de lectures ouverts (*orfs*) a été possible suite aux analyses bioinformatiques effectuées. Une grande proportion du génome (89,4%) est constituée de régions codantes, ce qui est en accord avec les observations rapportées préalablement, puisque les phages tendent à maximiser l'utilité de leurs génomes.

La traduction des *orfs* en séquences protéiques correspondantes a permis une comparaison basée sur la similarité de séquence en acides aminés. La recherche de domaines protéiques conservés avec les bases de données CCD (NCBI) et InterProScan a permis de mieux cerner la fonction de certaines protéines. Suite à l'ensemble de ces analyses, une fonction putative a pu être attribuée à une minorité de cadres de lectures (23/55), dont la grande partie implique des protéines structurales. Les analyses bioinformatiques n'ont pas pu prédire de fonction précise pour la grande majorité des *orfs* impliqués dans la réplication de l'ADN, la régulation de l'expression

des gènes ainsi que le contrôle de la lysogénie. Il est intéressant de noter que la grande majorité des *orfs* pour lesquels une fonction a pu être attribuée partagent une certaine similarité avec des gènes présents chez d'autres espèces de *Clostridium*. Dans cette optique, on pourrait penser que le phage ϕ CD38-2 a, par le passé, traversé la barrière des espèces et initié des échanges génétiques menant à la création de phages chimères dont on détecte des reliquats dans différentes espèces bactériennes de nos jours. Il serait aussi plausible que le phage ϕ CD38-2 descende d'un ancêtre commun qui était présent dans une espèce de *Clostridium* ancestrale qui a par la suite dérivé en des espèces différentes au cours de l'évolution, dont l'espèce *C. difficile* (FORD et al., 1998; HENDRIX et al., 1999; HENDRIX et al., 2003). Encore une fois, le séquençage d'autres siphophages chez *C. difficile* et éventuellement d'autres espèces de *Clostridium* et leur comparaison génomique servira à raffiner le lien existant entre les phages des différentes espèces de *Clostridium*.

3.3.4 Analyse détaillée des différents modules

En dépit de la faible similarité avec les séquences présentes dans les bases de données, la structure du génome du ϕ CD38-2 est relativement similaire aux autres phages caractérisés chez *C. difficile* ainsi que des phages tempérés infectant des bactéries avec un contenu en G+C faible (LUCCHINI *et al.*, 1999; CANCHAYA *et al.*, 2003). Une organisation modulaire permet aux phages de mieux contrôler

l'expression de certains gènes dans le temps ainsi que de synchroniser la production de protéines/enzymes spécifiques par rapport au stade du cycle lytique.

La particularité du ϕ CD38-2 par rapport aux myophages de *C. difficile* caractérisés précédemment, réside dans l'organisation du module du contrôle de la lysogénie. Ce module particulier n'est pas clairement défini dans le cas du ϕ CD38-2, et ressemble plutôt à celui du coliphage λ , également un *Siphoviridae* (LUCCHINI *et al.*, 1999). De plus, le siphophage de *C. difficile* ϕ CD6356 partage la même particularité (HORGAN *et al.*, 2010). Cependant, à la base de seulement deux séquences de type *Siphoviridae* chez *C. difficile*, il n'est pas possible de tirer une conclusion adéquate quant aux différences d'organisation modulaire des phages tempérés.

Le module comportant les gènes structuraux (formation de la queue et capsid) suit la même organisation que celle observée chez les myophages décrits chez *C. difficile*. Par contre, leurs homologues de séquence se retrouvent plutôt chez d'autres espèces de *Clostridium* incluant *C. perfringens*, *C. botulinum*, *C. tetani*, *C. beijerinckii* et *C. butyricum* (Table 1). Tel que discuté précédemment, il est possible que les phages infectant ces souches ainsi que le ϕ CD38-2 dérivent d'un ancêtre commun, ce qui expliquerait une certaine conservation des séquences protéiques. Le fait que ce soit les protéines de structure qui aient changé le moins au fil du temps n'est pas surprenant en soit. Logiquement, les protéines de structure, de par leur fonction spécifique, ont très peu de pression d'adaptation lorsqu'un phage change d'hôte. À l'inverse, les protéines impliquées dans les événements tels que la réplication et la transcription doivent être adaptées à l'hôte puisque ces processus font

intervenir le détournement de la machinerie cellulaire qui peut varier d'un hôte bactérien à un autre. Il en va de même pour le module de lysogénie où l'équilibre dans l'expression et l'activité de certains régulateurs est directement relié aux voies de régulation de la bactérie hôte (ex : répresseur CI versus réponse SOS) (QUINONES *et al.*, 2005; WALDOR et FRIEDMAN, 2005; ROKNEY *et al.*, 2008).

De plus, on constate que l'aspect morphologique est, somme toute, très bien conservé parmi les phages des différentes espèces. À titre d'exemple, 96% des phages caractérisés (sur plus de 5500) font partie de l'ordre *Caudovirales* qui comprend seulement trois familles (*Myoviridae*, *Siphoviridae* et *Podoviridae*) dont la caractéristique principale repose sur la présence d'une queue (ACKERMANN, 2007). Donc, malgré une grande diversité de séquence, la très grande majorité des phages tendent à garder une structure commune. Dans le même ordre d'idée, il n'est pas rare de trouver des phages infectant différentes espèces bactériennes, et qui partagent une forte similarité de séquence au niveau des protéines de structure, mais peu ou pas de similarité dans les autres modules (DUPUIS et MOINEAU, 2010).

L'analyse des séquences composant le module de lyse a mis en évidence la possibilité d'un événement de recombinaison survenu dans cette région. Cet événement aurait eu comme conséquence la permutation de régions codantes, c.-à-d. l'échange d'une partie du module de lyse menant à la création d'une endolysine chimère. Ce phénomène a été mis en évidence par l'alignement des résidus d'acides aminés de plusieurs endolysines encodées par les phages de *C. difficile*. Deux endolysines, provenant des phages ϕ C2 et ϕ CD27, partagent une très forte similarité avec la partie C-terminale de l'endolysine du phage ϕ CD38-2. Cependant, la partie N-

terminale de l'endolysine du ϕ CD38-2 ne partage pas de similarité significative avec les autres endolysines. Généralement, la partie C-terminale encode le domaine de liaison tandis que la partie N-terminale porte le domaine catalytique. Les endolysines, impliquées à la fin du cycle lytique, sont des protéines dont les gènes sont relativement conservés parmi les différents phages, et peuvent parfois servir de points de recombinaison lors d'une coinfection ou entre deux prophages intégrés. Un tel événement de recombinaison semble s'être produit avec le phage ϕ CD38-2, ce qui a mené à l'échange de domaines et création de protéines hybrides.

Tel que discuté précédemment, la plus grande divergence au niveau de la séquence se retrouve dans le module encodant les gènes impliqués dans la régulation, répllication et transcription. Il est intéressant de noter qu'une région d'environ 1900 pb située dans cette région possède une forte homologie de séquence avec le plasmide de la souche CD630 de *C. difficile*, couvrant en partie une région encodant une DNase. De plus, le contenu en G+C dans cette région particulière est plus faible que dans le reste du génome, suggérant que le phage ϕ CD38-2 a potentiellement recombiné avec un plasmide de *C. difficile* menant à l'acquisition du matériel génétique. L'analyse du contenu en G+C est souvent l'indicatif d'une acquisition horizontale du contenu génétique (HACKER *et al.*, 1997; FERRETTI *et al.*, 2004; MANN et CHEN, 2010). De telle manière, une chute ou une augmentation brusque du pourcentage de G+C dans un génome peut indiquer le point de recombinaison menant à l'acquisition de l'ADN étranger. Les trois plasmides caractérisés chez *C. difficile*, pCD630, pCD6 et pCDB11 ont tous un contenu en G+C relativement bas (entre 24-28%), celui du phage ϕ CD38-2 étant légèrement plus élevé (31,8%). De plus, un autre gène retrouvé

généralement sur les plasmides (*parA*) est également encodé sur cette région particulière. ParA est impliqué en combinaison avec ParB et une séquence nucléotidique spécifique (*parS*) dans le processus de la partition des plasmides (BOURHY *et al.*, 2005). Ces résultats viennent davantage appuyer l'hypothèse d'une recombinaison possible du phage ϕ CD38-2 avec un plasmide. Dans ce cas, les gènes encodés dans cette région particulière pourraient être considérés comme les gènes potentiels de conversion lysogénique. D'ailleurs, les analyses bioinformatiques ont montré que certains gènes de cette région possèdent des motifs particuliers, tels que des peptides-sigaux et/ou des régions transmembranaires. Il est à supposer que les produits de ces gènes pourraient être des protéines membranaires ou encore des peptides sécrétés. Il n'est pas exclu que ces protéines pourraient participer à l'adhésion bactérienne aux cellules épithéliales ou agir comme superantigènes (BANKS *et al.*, 2002; LLEWELYN et COHEN, 2002), mais il est plus vraisemblable qu'il s'agisse de protéines impliqués dans l'immunité des lysogènes. Dans certains cas, les phages, tel que le coliphage T5, peuvent exprimer des protéines qui ont la fonction de bloquer leur propre récepteur membranaire, empêchant ainsi la surinfection par les phages reconnaissant le même récepteur (PEDRUZZI *et al.*, 1998). Dans d'autres cas, les phages peuvent encoder des systèmes d'exclusion de la surinfection qui consistent à empêcher l'injection de l'ADN du phage dans la bactérie (KLIEM et DREISEIKELMANN, 1989; LU et HENNING, 1994; MOAK et MOLINEUX, 2000). Généralement, ces systèmes sont composés de protéines localisées au niveau de la membrane. Ainsi, dans le cas du phage ϕ CD38-2, les gènes

qui possèdent des peptides-sigaux et/ou des régions transmembranaires pourraient constituer de tels systèmes d'immunité.

Plusieurs promoteurs canoniques potentiels ont pu être identifiés en amont de ces gènes, suggérant une expression constitutive, par contre l'étude d'expression de ces gènes n'a pas pu être effectuée dû à l'induction spontanée du phage ϕ CD38-2 et à une certaine proportion des bactéries qui contiennent des phages en mode lytique et donc qui expriment des gènes qui ne sont normalement pas exprimés chez le prophage latent. L'impact de ces gènes sur la biologie bactérienne reste donc à être déterminé.

3.4 Caractérisation de l'impact du phage ϕ CD38-2 sur la biologie bactérienne

La lysogénéisation de plusieurs souches de *C. difficile* avec le phage ϕ CD38-2 a été effectuée afin de caractériser l'impact potentiel de ce phage sur la biologie bactérienne. Dans un premier temps, le phage ϕ CD38-2 a été introduit dans la souche CD274 (NAP1/027), et son impact sur quelques phénotypes différents a été déterminé. Ensuite, afin de déterminer s'il s'agissait d'un impact général ou plutôt spécifique à la souche utilisée, le phage ϕ CD38-2 a également été introduit dans quatre autres souches et les lysogènes ont été testés pour les mêmes phénotypes.

3.4.1 Impact du phage sur la croissance bactérienne

Une caractéristique susceptible d'être modifiée lors de la lysogénisation d'une souche est celle de la dynamique de croissance de la souche bactérienne.

Il s'est avéré que le phage ϕ CD38-2, suite à la lysogénisation, ne semble avoir aucun impact au niveau de la croissance bactérienne, du moins dans les milieux et conditions testés. D'une part, ceci suggère que le phage ne s'est pas intégré dans un gène qui est vital pour la bactérie, sans quoi la croissance en aurait été grandement affectée. D'autre part, le phage ne semble rien apporter à la bactérie qui pourrait stimuler sa croissance *in vitro*, dans les milieux testés.

3.4.2 Impact du phage sur le locus de pathogénicité

Le phage ϕ CD38-2 n'ayant pas d'effets au niveau de la croissance bactérienne, nous avons procédé à l'étude de son impact sur le locus de pathogénicité, en particulier la transcription des gènes *tcdA* et *tcdB* et la production de toxines A et B.

Les essais ELISA ont montré qu'environ deux fois plus de toxines étaient détectées dans le surnageant de la souche lysogène (CD274/ ϕ CD38-2) versus la souche sauvage (CD274) après 12h de croissance, suggérant une stimulation de la

production de toxines par le phage. L'immunodétection (dot blot) des toxines A et B à l'aide d'anticorps monoclonaux spécifiques a confirmé ces résultats en montrant une légère prédominance de la toxine B dans les surnageants de culture. Par ailleurs, la transcription des gènes de toxines dans la souche CD274 semble également stimulée par la présence du phage ϕ CD38-2. Cet effet n'était pas généralisé, dans le sens où une variation considérable a été observée lorsque le phage ϕ CD38-2 était introduit dans d'autres souches de *C. difficile*.

Le mécanisme exact de ce phénomène n'étant pas connu, seulement quelques hypothèses ont pu être avancées. Tout d'abord, il est possible qu'un régulateur du phage ϕ CD38-2 soit en mesure de reconnaître la région promotrice des gènes de toxines et ainsi stimuler leur transcription. Un effet similaire a été observé avec les lysogènes du phage ϕ CD119 (GOVIND *et al.*, 2009). En accord avec cette hypothèse, certains gènes du phage ϕ CD38-2 ont été identifiés en tant que régulateurs transcriptionnels potentiels. Également, il est possible que le produit d'un gène du phage soit capable d'interagir avec le répresseur TcdC et ainsi diminuer son efficacité. À l'inverse, une interaction avec le régulateur positif TcdR menant à une stimulation de la reconnaissance des promoteurs des gènes de toxines est également envisageable. Dans un cas comme dans l'autre, ceci aurait pour effet une transcription accrue des gènes *tcdA* et *tcdB* ce qui se traduirait par une augmentation de la production des toxines. La nature de ce gène hypothétique ne peut pas être déterminée avec précision puisque plus de la moitié des gènes du phage ϕ CD38-2 n'ont pas d'homologue dans les bases de données. Dans la même mesure, le phage pourrait affecter d'une quelconque façon la stabilité des ARNm des toxines, ce qui aurait pour effet

d'augmenter la quantité de transcrits détectée. Dans le même ordre d'idées, le site d'intégration du phage dans le chromosome bactérien peut expliquer le phénotype observé. Dans le cas où l'intégration du phage se fait à proximité du locus de pathogénicité ou encore dans la région promotrice des gènes *tcdA* ou *tcdB*, il est possible d'envisager une altération de la transcription/production des toxines. Le même phénomène sera observé si l'intégration du phage survient dans le répresseur *tcdC* ou un autre gène impliqué dans la régulation de l'expression / production des toxines. Par ailleurs, l'intégration des phages dans des gènes régulateurs a été observée auparavant (GOH *et al.*, 2007). Plusieurs tentatives d'identification du site d'intégration du phage ϕ CD38-2 dans diverses souches bactériennes ont été effectuées mais sans résultat concluant. Certaines de nos analyses suggèrent que le phage ϕ CD38-2 pourrait ne pas s'intégrer dans le chromosome bactérien suite à la lysogénisation, mais qu'il pourrait possiblement rester sous forme épisomale dans le cytoplasme de la bactérie.

L'ensemble de ces observations confirme que l'influence du ϕ CD38-2 sur la production des toxines est dépendante du contenu génétique bactérien. Un phénomène similaire a été observé auparavant avec d'autres phages de *C. difficile*, où l'influence des phages sur la transcription/production de toxines était hôte-dépendante (GOH *et al.*, 2005). Le ribotypage par PCR ne semble pas assez discriminant quant à l'effet du phage sur une souche bactérienne donnée puisque trois souches de ribotype 027 (CD66, CD111 et CD274) affichent des phénotypes différents lorsque lysogénisées avec le phage ϕ CD38-2. Il est possible que ces souches possèdent d'autres prophages dans leur génome et qu'une interaction puisse exister entre les

phages d'origine et le ϕ CD38-2 de manière à amplifier ou diminuer un effet. D'ailleurs, nous avons confirmé que les souches CD274 et CD111 possèdent deux prophages distincts et inductibles dans leur génome.

À la lumière des différentes observations discutées dans ce rapport, il semble évident que les phages de *C. difficile* jouent un rôle dans la biologie bactérienne en altérant certains phénotypes importants de la bactérie. Nous avons caractérisé au niveau microbiologique et moléculaire le deuxième phage tempéré de la famille des *Siphoviridae* chez *C. difficile*. Sa séquence génomique partage très peu de similarité avec les phages déjà caractérisés chez cette bactérie. Cependant, la structure de son génome est conforme à celle retrouvée chez la majorité des phages infectant des bactéries ayant un faible pourcentage G+C, mis à part le module de la lysogénie qui semble différer dans l'organisation de certains gènes. La capacité de ce phage à initier un cycle lytique spontanément lui procure l'avantage de dissémination lors d'une coinfection. Par ailleurs, ce phénomène est plutôt généralisé chez les phages tempérés. De telle manière, l'induction spontanée a été observée chez les lysogènes de *Salmonella enterica* (FIGUEROA-BOSSI et BOSSI, 1999; GARCIA-RUSSELL *et al.*, 2009), *Lactococcus lactis* (LUNDE *et al.*, 2003; LUNDE *et al.*, 2005) et *E. coli* (LIVNY et FRIEDMAN, 2004). Par la suite, l'impact du ϕ CD38-2 sur la biologie bactérienne a été évalué *in vitro*. Tandis qu'il ne semblait avoir aucun impact sur la dynamique de la croissance bactérienne (malgré l'induction spontanée), il s'est avéré capable d'augmenter la production de toxines par une stimulation de la transcription/stabilisation des gènes encodant les toxines.

L'ensemble de ces résultats ont soulevé plusieurs hypothèses et perspectives. Dans un premier temps, il sera intéressant d'évaluer les implications biologiques de la lysogénie du phage ϕ CD38-2 dans un modèle *in vivo*. Il est possible que l'effet du phage sur la production de toxines observé *in vitro* soit différente lorsque évalué *in vivo*. Également, d'un point de vue moléculaire, il sera intéressant de déterminer le mécanisme exact par lequel le phage ϕ CD38-2 arrive à stimuler la transcription/production des toxines. Dans ce sens, la détermination de son site d'intégration dans le chromosome bactérien est primordiale. Dans la même ligne d'idées, il sera intéressant d'évaluer la capacité du phage ϕ CD38-2 à influencer d'autres phénotypes relié à la pathogenèse bactérienne tels que l'adhésion, sporulation, mobilité, acquisition du fer etc. La présence de gènes codant pour des protéines ayant des passages transmembranaires et/ou peptides signaux peut suggérer une implication dans l'adhésion aux cellules épithéliales. Également, il est possible que la présence du gène *parA* dans le génome du ϕ CD38-2 soit reliée à l'inhibition de l'initiation de la sporulation, tel qu'observé auparavant chez *Bacillus subtilis* (IRETON *et al.*, 1994; QUISEL et GROSSMAN, 2000; AUTRET *et al.*, 2001; MCLEOD et SPIEGELMAN, 2005). Deux autres phages de *C. difficile* semblent disposer du gène *parA* sans toutefois encoder sa contrepartie *parB* (GOH *et al.*, 2007; HORGAN *et al.*, 2010), donc il sera d'autant plus intéressant d'établir sa fonction précise au sein de ces génomes. Finalement, une étude sur la distribution du phage ϕ CD38-2 dans une large collection d'isolats de *C. difficile* sera intéressante puisque ce phage semble avoir un impact au niveau de la transcription/production des toxines.

Une large distribution du phage ϕ CD38-2 pourra suggérer qu'il procure un avantage compétitif par rapport aux souches n'ayant pas intégré le phage.

4 REMERCIEMENTS

Tout d'abord, je tiens à remercier Dr. Louis-Charles Fortier de m'avoir donné l'opportunité de poursuivre des études de 2^e cycle dans son laboratoire et de m'avoir guidé tout au long de la réalisation de ce projet. Un grand merci à Dr. Raymund Wellinger et Dr. Vincent Burrus pour avoir accepté de faire partie du comité d'évaluation.

Je tiens également à remercier mes collègues de travail, étudiants et stagiaires, avec un accent particulier à Mathieu Pinard pour toute l'aide technique et surtout pour les bonnes discussions, encouragements et support moral. Word up! Un merci très spécial à Stéphanie Mauler sans qui les longues années de maîtrise n'auraient pas été aussi agréables. À Kind et OB, you are family...

Je remercie toutes les personnes au département de Microbiologie qui m'ont aidé et/ou conseillé dans les différents aspects de mes études avec une mention spéciale à Nancy Laterreur, Simon Lévesque, David Faucher, Jean-François Noël, Laetitia Michelle, Carole Picard et Mylène Ledoux.

Finalement, je remercie ma famille pour leur support indéfectible. Family over everything. One love.

5 BIBLIOGRAPHIE

Acheson, D. W., J. Reidl, X. Zhang, G. T. Keusch, J. J. Mekalanos and M. K. Waldor (1998). "In vivo transduction with shiga toxin 1-encoding phage." Infect Immun **66**(9): 4496-4498.

Ackermann, H. W. (2007). "5500 Phages examined in the electron microscope." Arch Virol **152**(2): 227-243.

Ahmed, K., T. C. Dai, A. Ichinose, H. Masaki, T. Nagatake and K. Matsumoto (1993). "Neutrophil response to *Pseudomonas aeruginosa* in respiratory infection." Microbiol Immunol **37**(7): 523-529.

Akerlund, T., I. Persson, M. Unemo, T. Noren, B. Svenungsson, M. Wullt and L. G. Burman (2008). "Increased sporulation rate of epidemic *Clostridium difficile* Type 027/NAP1." J Clin Microbiol **46**(4): 1530-1533.

Aktorics, K. and I. Just (2005). "Clostridial Rho-inhibiting protein toxins." Curr Top Microbiol Immunol **291**: 113-145.

Asadulghani, M., Y. Ogura, T. Ooka, T. Itoh, A. Sawaguchi, A. Iguchi, K. Nakayama and T. Hayashi (2009). "The defective prophage pool of *Escherichia coli* O157: prophage-prophage interactions potentiate horizontal transfer of virulence determinants." PLoS Pathog **5**(5): e1000408.

Aslam, S., R. J. Hamill and D. M. Musher (2005). "Treatment of *Clostridium difficile*-associated disease: old therapies and new strategies." Lancet Infect Dis **5**(9): 549-557.

Autret, S., R. Nair and J. Errington (2001). "Genetic analysis of the chromosome segregation protein Spo0J of *Bacillus subtilis*: evidence for separate domains involved in DNA binding and interactions with Soj protein." Mol Microbiol **41**(3): 743-755.

Aziz, R. K., R. A. Edwards, W. W. Taylor, D. E. Low, A. McGeer and M. Kotb (2005). "Mosaic prophages with horizontally acquired genes account for the emergence and diversification of the globally disseminated MIT1 clone of *Streptococcus pyogenes*." J Bacteriol **187**(10): 3311-3318.

Banks, D. J., S. B. Beres and J. M. Musser (2002). "The fundamental contribution of phages to GAS evolution, genome diversification and strain emergence." Trends Microbiol **10**(11): 515-521.

Barbut, F., G. Corthier, Y. Charpak, M. Cerf, H. Monteil, T. Fosse, A. Trevoux, B. De Barbeyrac, Y. Boussougant, S. Tigaud, F. Tytgat, A. Sedallian, S. Duborgel, A. Collignon, M. E. Le Guern, P. Bernasconi and J. C. Petit (1996). "Prevalence and pathogenicity of *Clostridium difficile* in hospitalized patients. A French multicenter study." Arch Intern Med **156**(13): 1449-1454.

Barksdale, L. and S. B. Arden (1974). "Persisting bacteriophage infections, lysogeny, and phage conversions." Annu Rev Microbiol **28**(0): 265-299.

Barth, H. (2004). "Uptake of binary actin ADP-ribosylating toxins." Rev Physiol Biochem Pharmacol **152**: 165-182.

Beres, S. B., G. L. Sylva, K. D. Barbian, B. Lei, J. S. Hoff, N. D. Mammarella, M. Y. Liu, J. C. Smoot, S. F. Porcella, L. D. Parkins, D. S. Campbell, T. M. Smith, J. K.

McCormick, D. Y. Leung, P. M. Schlievert and J. M. Musser (2002). "Genome sequence of a serotype M3 strain of group A *Streptococcus*: phage-encoded toxins, the high-virulence phenotype, and clone emergence." Proc Natl Acad Sci U S A **99**(15): 10078-10083.

Borriello, S. P., H. A. Davies, S. Kamiya, P. J. Reed and S. Seddon (1990). "Virulence factors of *Clostridium difficile*." Rev Infect Dis **12 Suppl 2**: S185-191.

Bouchard, J. D., E. Dion, F. Bissonnette and S. Moineau (2002). "Characterization of the two-component abortive phage infection mechanism AbiT from *Lactococcus lactis*." J Bacteriol **184**(22): 6325-6332.

Bourhy, P., L. Frangeul, E. Couve, P. Glaser, I. Saint Girons and M. Picardeau (2005). "Complete nucleotide sequence of the LE1 prophage from the spirochete *Leptospira biflexa* and characterization of its replication and partition functions." J Bacteriol **187**(12): 3931-3940.

Boyd, E. F. and H. Brussow (2002). "Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved." Trends Microbiol **10**(11): 521-529.

Brazier, J. S. (1998). "The epidemiology and typing of *Clostridium difficile*." J Antimicrob Chemother **41 Suppl C**: 47-57.

Brock, T. D. (1990). The emergence of bacterial genetics. New York, Cold Spring Harbor Laboratory Press.

Brussow, H., C. Canchaya and W. D. Hardt (2004). "Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion." Microbiol Mol Biol Rev **68**(3): 560-602.

Brussow, H. and R. W. Hendrix (2002). "Phage genomics: small is beautiful." Cell **108**(1): 13-16.

Calabi, E., F. Calabi, A. D. Phillips and N. F. Fairweather (2002). "Binding of *Clostridium difficile* surface layer proteins to gastrointestinal tissues." Infect Immun **70**(10): 5770-5778.

Calabi, E., S. Ward, B. Wren, T. Paxton, M. Panico, H. Morris, A. Dell, G. Dougan and N. Fairweather (2001). "Molecular characterization of the surface layer proteins from *Clostridium difficile*." Mol Microbiol **40**(5): 1187-1199.

Calderwood, S. B. and J. J. Mekalanos (1987). "Iron regulation of Shiga-like toxin expression in *Escherichia coli* is mediated by the *fur* locus." J Bacteriol **169**(10): 4759-4764.

Calendar, R. (2006). The bacteriophages. New York, Oxford university press Inc.

Campbell, A., A. del-Campillo-Campbell and M. L. Ginsberg (2002). "Specificity in DNA recognition by phage integrases." Gene **300**(1-2): 13-18.

Campbell, A. M. (1992). "Chromosomal insertion sites for phages and plasmids." J Bacteriol **174**(23): 7495-7499.

Canchaya, C., G. Fournous and H. Brussow (2004). "The impact of prophages on bacterial chromosomes." Mol Microbiol **53**(1): 9-18.

Canchaya, C., G. Fournous, S. Chibani-Chennoufi, M. L. Dillmann and H. Brussow (2003). "Phage as agents of lateral gene transfer." Curr Opin Microbiol **6**(4): 417-424.

Canchaya, C., C. Proux, G. Fournous, A. Bruttin and H. Brussow (2003). "Prophage genomics." Microbiol Mol Biol Rev **67**(2): 238-276, table of contents.

Casjens, S. (2003). "Prophages and bacterial genomics: what have we learned so far?" Mol Microbiol **49**(2): 277-300.

Chen, L. F., T. Chopra and K. S. Kaye (2009). "Pathogens resistant to antibacterial agents." Infect Dis Clin North Am **23**(4): 817-845, vii.

Coleman, D. C., D. J. Sullivan, R. J. Russell, J. P. Arbuthnott, B. F. Carey and H. M. Pomeroy (1989). "*Staphylococcus aureus* bacteriophages mediating the simultaneous lysogenic conversion of beta-lysin, staphylokinase and enterotoxin A: molecular mechanism of triple conversion." J Gen Microbiol **135**(6): 1679-1697.

Curry, S. R., J. W. Marsh, C. A. Muto, M. M. O'Leary, A. W. Pasculle and L. H. Harrison (2007). "tcdC genotypes associated with severe TcdC truncation in an epidemic clone and other strains of *Clostridium difficile*." J Clin Microbiol **45**(1): 215-221.

Curry, S. R., J. W. Marsh, K. A. Shutt, C. A. Muto, M. M. O'Leary, M. I. Saul, A. W. Pasculle and L. H. Harrison (2009). "High frequency of rifampin resistance identified in an epidemic *Clostridium difficile* clone from a large teaching hospital." Clin Infect Dis **48**(4): 425-429.

Davies, H. A. and S. P. Borriello (1990). "Detection of capsule in strains of *Clostridium difficile* of varying virulence and toxigenicity." Microb Pathog **9**(2): 141-146.

Davis, B. M., H. H. Kimsey, A. V. Kane and M. K. Waldor (2002). "A satellite phage-encoded antirepressor induces repressor aggregation and cholera toxin gene transfer." EMBO J **21**(16): 4240-4249.

Dawson, L. F., E. Valiente and B. W. Wren (2009). "*Clostridium difficile*--a continually evolving and problematic pathogen." Infect Genet Evol **9**(6): 1410-1417.

Dei, R. (1989). "Observations on phage-typing of *Clostridium difficile*: preliminary evaluation of a phage panel." Eur J Epidemiol **5**(3): 351-354.

Domingues, S., S. McGovern, D. Plochocka, M. A. Santos, S. D. Ehrlich, P. Polard and M. C. Chopin (2008). "The lactococcal abortive infection protein AbiP is membrane-anchored and binds nucleic acids." Virology **373**(1): 14-24.

Drawz, S. M. and R. A. Bonomo (2010). "Three decades of beta-lactamase inhibitors." Clin Microbiol Rev **23**(1): 160-201.

Dupuis, M. E. and S. Moineau (2010). "Genome organization and characterization of the virulent lactococcal phage 1358 and its similarities to *Listeria* phages." Appl Environ Microbiol **76**(5): 1623-1632.

Dupuy, B., R. Govind, A. Antunes and S. Matamouros (2008). "*Clostridium difficile* toxin synthesis is negatively regulated by TcdC." J Med Microbiol **57**(Pt 6): 685-689.

Dupuy, B., S. Raffestin, S. Matamouros, N. Mani, M. R. Popoff and A. L. Sonenshein (2006). "Regulation of toxin and bacteriocin gene expression in *Clostridium* by interchangeable RNA polymerase sigma factors." Mol Microbiol **60**(4): 1044-1057.

Eggertson, L. (2007). "*C. difficile* outbreaks in Gatineau, Sault Ste. Marie." CMAJ **177**(4): 334-335.

Emerson, J. E., R. A. Stabler, B. W. Wren and N. F. Fairweather (2008). "Microarray analysis of the transcriptional responses of *Clostridium difficile* to environmental and antibiotic stress." J Med Microbiol **57**(Pt 6): 757-764.

Evans, T. and W. V. Lerberghe. (2008). "Maintenant plus que jamais." Rapport sur la santé dans le monde, from http://www.who.int/whr/2008/08_report_fr.pdf.

Fawley, W. N., J. Freeman, C. Smith, C. Harmanus, R. J. van den Berg, E. J. Kuijper and M. H. Wilcox (2008). "Use of highly discriminatory fingerprinting to analyze clusters of *Clostridium difficile* infection cases due to epidemic ribotype 027 strains." J Clin Microbiol **46**(3): 954-960.

Ferretti, J. J., D. Ajdic and W. M. McShan (2004). "Comparative genomics of streptococcal species." Indian J Med Res **119** **Suppl**: 1-6.

Figuroa-Bossi, N. and L. Bossi (1999). "Inducible prophages contribute to *Salmonella* virulence in mice." Mol Microbiol **33**(1): 167-176.

Fineran, P. C., T. R. Blower, I. J. Foulds, D. P. Humphreys, K. S. Lilley and G. P. Salmond (2009). "The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair." Proc Natl Acad Sci U S A **106**(3): 894-899.

Ford, M. E., G. J. Sarkis, A. E. Belanger, R. W. Hendrix and G. F. Hatfull (1998). "Genome structure of mycobacteriophage D29: implications for phage evolution." J Mol Biol **279**(1): 143-164.

Fortier, L. C. and S. Moineau (2007). "Morphological and genetic diversity of temperate phages in *Clostridium difficile*." Appl Environ Microbiol **73**(22): 7358-7366.

Freeman, V. J. (1951). "Studies on the virulence of bacteriophage-infected strains of *Corynebacterium diphtheriae*." J Bacteriol **61**(6): 675-688.

Gamage, S. D., J. E. Strasser, C. L. Chalk and A. A. Weiss (2003). "Nonpathogenic *Escherichia coli* can contribute to the production of Shiga toxin." Infect Immun **71**(6): 3107-3115.

Garcia-Russell, N., B. Elrod and K. Dominguez (2009). "Stress-induced prophage DNA replication in *Salmonella enterica* serovar *Typhimurium*." Infect Genet Evol **9**(5): 889-895.

Genth, H., S. C. Dreger, J. Huelsenbeck and I. Just (2008). "*Clostridium difficile* toxins: more than mere inhibitors of Rho proteins." Int J Biochem Cell Biol **40**(4): 592-597.

Gerber, M., C. Walch, B. Loffler, K. Tischendorf, U. Reischl and G. Ackermann (2008). "Effect of sub-MIC concentrations of metronidazole, vancomycin, clindamycin and linezolid on toxin gene transcription and production in *Clostridium difficile*." J Med Microbiol **57**(Pt 6): 776-783.

Gerding, D. N., C. A. Muto and R. C. Owens, Jr. (2008). "Treatment of *Clostridium difficile* infection." Clin Infect Dis **46 Suppl 1**: S32-42.

Geric, B., R. J. Carman, M. Rupnik, C. W. Genheimer, S. P. Sambol, D. M. Lyerly, D. N. Gerding and S. Johnson (2006). "Binary toxin-producing, large clostridial toxin-negative *Clostridium difficile* strains are enterotoxic but do not cause disease in hamsters." J Infect Dis **193**(8): 1143-1150.

Goh, S., B. J. Chang and T. V. Riley (2005). "Effect of phage infection on toxin production by *Clostridium difficile*." J Med Microbiol **54**(Pt 2): 129-135.

Goh, S., P. F. Ong, K. P. Song, T. V. Riley and B. J. Chang (2007). "The complete genome sequence of *Clostridium difficile* phage phiC2 and comparisons to phiCD119 and inducible prophages of CD630." Microbiology **153**(Pt 3): 676-685.

Goh, S., T. V. Riley and B. J. Chang (2005). "Isolation and characterization of temperate bacteriophages of *Clostridium difficile*." Appl Environ Microbiol **71**(2): 1079-1083.

Goncalves, C., D. Decre, F. Barbut, B. Burghoffer and J. C. Petit (2004). "Prevalence and characterization of a binary toxin (actin-specific ADP-ribosyltransferase) from *Clostridium difficile*." J Clin Microbiol **42**(5): 1933-1939.

Goshorn, S. C. and P. M. Schlievert (1989). "Bacteriophage association of streptococcal pyrogenic exotoxin type C." J Bacteriol **171**(6): 3068-3073.

Govind, R., J. A. Fralick and R. D. Rolfe (2006). "Genomic organization and molecular characterization of *Clostridium difficile* bacteriophage PhiCD119." J Bacteriol **188**(7): 2568-2577.

Govind, R., G. VEDIYAPPAN, R. D. Rolfe, B. Dupuy and J. A. Fralick (2009). "Bacteriophage-mediated toxin gene regulation in *Clostridium difficile*." J Virol **83**(23): 12037-12045.

Groman, N. B. (1956). "Conversion in *Corynebacterium diphtheriae* with phages originating from nontoxigenic strains." Virology **2**(6): 843-844.

Hacker, J., G. Blum-Oehler, I. Muhldorfer and H. Tschape (1997). "Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution." Mol Microbiol **23**(6): 1089-1097.

Halsey, J. (2008). "Current and future treatment modalities for *Clostridium difficile*-associated disease." Am J Health Syst Pharm **65**(8): 705-715.

Hammerschmidt, S., A. Muller, H. Sillmann, M. Muhlenhoff, R. Borrow, A. Fox, J. van Putten, W. D. Zollinger, R. Gerardy-Schahn and M. Frosch (1996). "Capsule phase variation in *Neisseria meningitidis* serogroup B by slipped-strand mispairing in the polysialyltransferase gene (siaD): correlation with bacterial invasion and the outbreak of meningococcal disease." Mol Microbiol **20**(6): 1211-1220.

Hawkins, C. C., B. P. Buggy, R. Fekety and D. R. Schaberg (1984). "Epidemiology of colitis induced by *Clostridium difficile* in hamsters: application of a bacteriophage and bacteriocin typing system." J Infect Dis **149**(5): 775-780.

Hayashi, T., K. Makino, M. Ohnishi, K. Kurokawa, K. Ishii, K. Yokoyama, C. G. Han, E. Ohtsubo, K. Nakayama, T. Murata, M. Tanaka, T. Tobe, T. Iida, H. Takami, T. Honda, C. Sasakawa, N. Ogasawara, T. Yasunaga, S. Kuhara, T. Shiba, M. Hattori and H. Shinagawa (2001). "Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12." DNA Res **8**(1): 11-22.

He, M., M. Sebaihia, T. D. Lawley, R. A. Stabler, L. F. Dawson, M. J. Martin, K. E. Holt, H. M. Seth-Smith, M. A. Quail, R. Rance, K. Brooks, C. Churcher, D. Harris, S. D. Bentley, C. Burrows, L. Clark, C. Corton, V. Murray, G. Rose, S. Thurston, A. van Tonder, D. Walker, B. W. Wren, G. Dougan and J. Parkhill (2010). "Evolutionary dynamics of *Clostridium difficile* over short and long time scales." Proc Natl Acad Sci U S A **107**(16): 7527-7532.

Hendrix, R. W., G. F. Hatfull and M. C. Smith (2003). "Bacteriophages with tails: chasing their origins and evolution." Res Microbiol **154**(4): 253-257.

Hendrix, R. W., M. C. Smith, R. N. Burns, M. E. Ford and G. F. Hatfull (1999). "Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage." Proc Natl Acad Sci U S A **96**(5): 2192-2197.

Hermoso, J. A., J. L. Garcia and P. Garcia (2007). "Taking aim on bacterial pathogens: from phage therapy to enzybiotics." Curr Opin Microbiol **10**(5): 461-472.

Herold, S., J. Siebert, A. Huber and H. Schmidt (2005). "Global expression of prophage genes in *Escherichia coli* O157:H7 strain EDL933 in response to norfloxacin." Antimicrob Agents Chemother **49**(3): 931-944.

Horgan, M., O. O'Sullivan, A. Coffey, G. F. Fitzgerald, D. van Sinderen, O. McAuliffe and R. P. Ross (2010). "Genome analysis of the *Clostridium difficile* phage PhiCD6356, a temperate phage of the Siphoviridae family." Gene **462**(1-2):34-43.

Huang, H., A. Weintraub, H. Fang and C. E. Nord (2009). "Antimicrobial resistance in *Clostridium difficile*." Int J Antimicrob Agents **34**(6): 516-522.

Ireton, K., N. W. t. Gunther and A. D. Grossman (1994). "spo0J is required for normal chromosome segregation as well as the initiation of sporulation in *Bacillus subtilis*." J Bacteriol **176**(17): 5320-5329.

Ishida, Y., T. Maegawa, T. Kondo, A. Kimura, Y. Iwakura, S. Nakamura and N. Mukaida (2004). "Essential involvement of IFN-gamma in *Clostridium difficile* toxin A-induced enteritis." J Immunol **172**(5): 3018-3025.

Janoir, C., S. Pechine, C. Grosdidier and A. Collignon (2007). "Cwp84, a surface-associated protein of *Clostridium difficile*, is a cysteine protease with degrading activity on extracellular matrix proteins." J Bacteriol **189**(20): 7174-7180.

Johnson, S. and D. N. Gerding (1998). "*Clostridium difficile*--associated diarrhea." Clin Infect Dis **26**(5): 1027-1034; quiz 1035-1026.

Karaolis, D. K., S. Somara, D. R. Maneval, Jr., J. A. Johnson and J. B. Kaper (1999). "A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria." Nature **399**(6734): 375-379.

Karjalainen, T., A. J. Waligora-Dupriet, M. Cerquetti, P. Spigaglia, A. Maggioni, P. Mauri and P. Mastrantonio (2001). "Molecular and genomic analysis of genes

encoding surface-anchored proteins from *Clostridium difficile*." Infect Immun **69**(5): 3442-3446.

Kliem, M. and B. Dreiseikelmann (1989). "The superimmunity gene sim of bacteriophage P1 causes superinfection exclusion." Virology **171**(2): 350-355.

Kuijper, E. J., B. Coignard and P. Tull (2006). "Emergence of *Clostridium difficile*-associated disease in North America and Europe." Clin Microbiol Infect **12 Suppl 6**: 2-18.

Kuijper, E. J., J. T. van Dissel and M. H. Wilcox (2007). "*Clostridium difficile*: changing epidemiology and new treatment options." Curr Opin Infect Dis **20**(4): 376-383.

Labbe, A. C., L. Poirier, D. Maccannell, T. Louie, M. Savoie, C. Beliveau, M. Laverdiere and J. Pepin (2008). "*Clostridium difficile* infections in a Canadian tertiary care hospital before and during a regional epidemic associated with the BI/NAP1/027 strain." Antimicrob Agents Chemother **52**(9): 3180-3187.

Lee, C. Y. and J. J. Iandolo (1985). "Mechanism of bacteriophage conversion of lipase activity in *Staphylococcus aureus*." J Bacteriol **164**(1): 288-293.

Lee, C. Y. and J. J. Iandolo (1986). "Lysogenic conversion of staphylococcal lipase is caused by insertion of the bacteriophage L54a genome into the lipase structural gene." J Bacteriol **166**(2): 385-391.

- Lemee, L., A. Dhalluin, M. Pestel-Caron, J. F. Lemeland and J. L. Pons (2004). "Multilocus sequence typing analysis of human and animal *Clostridium difficile* isolates of various toxigenic types." J Clin Microbiol **42**(6): 2609-2617.
- Lister, P. D., D. J. Wolter and N. D. Hanson (2009). "Antibacterial-resistant *Pseudomonas aeruginosa*: clinical impact and complex regulation of chromosomally encoded resistance mechanisms." Clin Microbiol Rev **22**(4): 582-610.
- Livny, J. and D. I. Friedman (2004). "Characterizing spontaneous induction of Stx encoding phages using a selectable reporter system." Mol Microbiol **51**(6): 1691-1704.
- Llewelyn, M. and J. Cohen (2002). "Superantigens: microbial agents that corrupt immunity." Lancet Infect Dis **2**(3): 156-162.
- Loo, V. G., L. Poirier, M. A. Miller, M. Oughton, M. D. Libman, S. Michaud, A. M. Bourgault, T. Nguyen, C. Frenette, M. Kelly, A. Vibien, P. Brassard, S. Fenn, K. Dewar, T. J. Hudson, R. Horn, P. Rene, Y. Monczak and A. Dascal (2005). "A predominantly clonal multi-institutional outbreak of *Clostridium difficile*-associated diarrhea with high morbidity and mortality." N Engl J Med **353**(23): 2442-2449.
- Lu, M. J. and U. Henning (1994). "Superinfection exclusion by T-even-type coliphages." Trends Microbiol **2**(4): 137-139.
- Lucchini, S., F. Desiere and H. Brussow (1999). "Similarly organized lysogeny modules in temperate Siphoviridae from low GC content gram-positive bacteria." Virology **263**(2): 427-435.

Lunde, M., A. H. Aastveit, J. M. Blatny and I. F. Nes (2005). "Effects of diverse environmental conditions on (ACHESON *et al.*)LC3 prophage stability in *Lactococcus lactis*." Appl Environ Microbiol **71**(2): 721-727.

Lunde, M., J. M. Blatny, D. Lillehaug, A. H. Aastveit and I. F. Nes (2003). "Use of real-time quantitative PCR for the analysis of phiLC3 prophage stability in lactococci." Appl Environ Microbiol **69**(1): 41-48.

Lyras, D., J. R. O'Connor, P. M. Howarth, S. P. Sambol, G. P. Carter, T. Phumoonna, R. Poon, V. Adams, G. Vedantam, S. Johnson, D. N. Gerding and J. I. Rood (2009). "Toxin B is essential for virulence of *Clostridium difficile*." Nature **458**(7242): 1176-1179.

Mahony, D. E., P. D. Bell and K. B. Easterbrook (1985). "Two bacteriophages of *Clostridium difficile*." J Clin Microbiol **21**(2): 251-254.

Mann, S. and Y. P. Chen (2010). "Bacterial genomic G+C composition-eliciting environmental adaptation." Genomics **95**(1): 7-15.

Matamouros, S., P. England and B. Dupuy (2007). "*Clostridium difficile* toxin expression is inhibited by the novel regulator TcdC." Mol Microbiol **64**(5): 1274-1288.

Matsushiro, A., K. Sato, H. Miyamoto, T. Yamamura and T. Honda (1999). "Induction of prophages of enterohemorrhagic *Escherichia coli* O157:H7 with norfloxacin." J Bacteriol **181**(7): 2257-2260.

Matte, I., D. Lane, E. Cote, A. E. Asselin, L. C. Fortier, C. Asselin and A. Piche (2009). "Antiapoptotic proteins Bcl-2 and Bcl-XL inhibit *Clostridium difficile* toxin A-induced cell death in human epithelial cells." Infect Immun **77**(12): 5400-5410.

Mayer, M. J., A. Narbad and M. J. Gasson (2008). "Molecular characterization of a *Clostridium difficile* bacteriophage and its cloned biologically active endolysin." J Bacteriol **190**(20): 6734-6740.

McDonald, L. C., G. E. Killgore, A. Thompson, R. C. Owens, Jr., S. V. Kazakova, S. P. Sambol, S. Johnson and D. N. Gerding (2005). "An epidemic, toxin gene-variant strain of *Clostridium difficile*." N Engl J Med **353**(23): 2433-2441.

McLeod, B. N. and G. B. Spiegelman (2005). "Soj antagonizes Spo0A activation of transcription in *Bacillus subtilis*." J Bacteriol **187**(7): 2532-2536.

Merrell, D. S., S. M. Butler, F. Qadri, N. A. Dolganov, A. Alam, M. B. Cohen, S. B. Calderwood, G. K. Schoolnik and A. Camilli (2002). "Host-induced epidemic spread of the cholera bacterium." Nature **417**(6889): 642-645.

Meyer, G. K., A. Neetz, G. Brandes, D. Tsikas, J. H. Butterfield, I. Just and R. Gerhard (2007). "*Clostridium difficile* toxins A and B directly stimulate human mast cells." Infect Immun **75**(8): 3868-3876.

Moak, M. and I. J. Molineux (2000). "Role of the Gp16 lytic transglycosylase motif in bacteriophage T7 virions at the initiation of infection." Mol Microbiol **37**(2): 345-355.

Murray, R., D. Boyd, P. N. Levett, M. R. Mulvey and M. J. Alfa (2009). "Truncation in the *tdcC* region of the *Clostridium difficile* PathLoc of clinical isolates does not predict increased biological activity of Toxin B or Toxin A." BMC Infect Dis **9**: 103.

Nagy, E. and J. Foldes (1991). "Electron microscopic investigation of lysogeny of *Clostridium difficile* strains isolated from antibiotic-associated diarrhea cases and from healthy carriers." APMIS **99**(4): 321-326.

Nakayama, K., S. Kanaya, M. Ohnishi, Y. Terawaki and T. Hayashi (1999). "The complete nucleotide sequence of phi CTX, a cytotoxin-converting phage of *Pseudomonas aeruginosa*: implications for phage evolution and horizontal gene transfer via bacteriophages." Mol Microbiol **31**(2): 399-419.

O'Connor, J. R., S. Johnson and D. N. Gerding (2009). "*Clostridium difficile* infection caused by the epidemic BI/NAP1/027 strain." Gastroenterology **136**(6): 1913-1924.

Ohnishi, M., K. Kurokawa and T. Hayashi (2001). "Diversification of *Escherichia coli* genomes: are bacteriophages the major contributors?" Trends Microbiol **9**(10): 481-485.

Pedruzzi, I., J. P. Rosenbusch and K. P. Locher (1998). "Inactivation in vitro of the *Escherichia coli* outer membrane protein FhuA by a phage T5-encoded lipoprotein." FEMS Microbiol Lett **168**(1): 119-125.

Pelaez, T., E. Cercenado, L. Alcala, M. Marin, A. Martin-Lopez, J. Martinez-Alarcon, P. Catalan, M. Sanchez-Somolinos and E. Bouza (2008). "Metronidazole resistance in *Clostridium difficile* is heterogeneous." J Clin Microbiol **46**(9): 3028-3032.

Pepin, J., L. Valiquette, M. E. Alary, P. Villemure, A. Pelletier, K. Forget, K. Pepin and D. Chouinard (2004). "*Clostridium difficile*-associated diarrhea in a region of Quebec from 1991 to 2003: a changing pattern of disease severity." CMAJ **171**(5): 466-472.

Quinones, M., H. H. Kimsey and M. K. Waldor (2005). "LexA cleavage is required for CTX prophage induction." Mol Cell **17**(2): 291-300.

Quisel, J. D. and A. D. Grossman (2000). "Control of sporulation gene expression in *Bacillus subtilis* by the chromosome partitioning proteins Soj (ParA) and Spo0J (ParB)." J Bacteriol **182**(12): 3446-3451.

Roberts, R. J., T. Vincze, J. Posfai and D. Macelis (2007). "REBASE--enzymes and genes for DNA restriction and modification." Nucleic Acids Res **35**(Database issue): D269-270.

Rocha, M. F., M. E. Maia, L. R. Bezerra, D. M. Lyerly, R. L. Guerrant, R. A. Ribeiro and A. A. Lima (1997). "*Clostridium difficile* toxin A induces the release of neutrophil chemotactic factors from rat peritoneal macrophages: role of interleukin-1beta, tumor necrosis factor alpha, and leukotrienes." Infect Immun **65**(7): 2740-2746.

Rodriguez-Palacios, A., H. R. Staempfli, T. Duffield and J. S. Weese (2007). "*Clostridium difficile* in retail ground meat, Canada." Emerg Infect Dis **13**(3): 485-487.

Rokney, A., O. Kobiler, A. Amir, D. L. Court, J. Stavans, S. Adhya and A. B. Oppenheim (2008). "Host responses influence on the induction of lambda prophage." Mol Microbiol **68**(1): 29-36.

Samore, M. H., L. Venkataraman, P. C. DeGirolami, M. M. Merrigan, S. Johnson, D. N. Gerding, Y. Carmeli and S. Harbarth (2006). "Genotypic and phenotypic analysis of *Clostridium difficile* correlated with previous antibiotic exposure." Microb Drug Resist **12**(1): 23-28.

Saxton, K., S. D. Baines, J. Freeman, R. O'Connor and M. H. Wilcox (2009). "Effects of exposure of *Clostridium difficile* PCR ribotypes 027 and 001 to fluoroquinolones in a human gut model." Antimicrob Agents Chemother **53**(2): 412-420.

Schmidt, C., B. Löffler and G. Ackermann (2007). "Antimicrobial phenotypes and molecular basis in clinical strains of *Clostridium difficile*." Diagn Microbiol Infect Dis **59**(1): 1-5.

Schwan, C., B. Stecher, T. Tzivelekidis, M. van Ham, M. Rohde, W. D. Hardt, J. Wehland and K. Aktories (2009). "*Clostridium difficile* toxin CDT induces formation of microtubule-based protrusions and increases adherence of bacteria." PLoS Pathog **5**(10): e1000626.

Sebahia, M., B. W. Wren, P. Mullany, N. F. Fairweather, N. Minton, R. Stabler, N. R. Thomson, A. P. Roberts, A. M. Cerdeno-Tarraga, H. Wang, M. T. Holden, A. Wright, C. Churcher, M. A. Quail, S. Baker, N. Bason, K. Brooks, T. Chillingworth, A. Cronin, P. Davis, L. Dowd, A. Fraser, T. Feltwell, Z. Hance, S. Holroyd, K. Jagels, S. Moule, K. Mungall, C. Price, E. Rabinowitsch, S. Sharp, M. Simmonds, K.

Stevens, L. Unwin, S. Whithead, B. Dupuy, G. Dougan, B. Barrell and J. Parkhill (2006). "The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome." Nat Genet **38**(7): 779-786.

Seddon, S. V., I. Hemingway and S. P. Borriello (1990). "Hydrolytic enzyme production by *Clostridium difficile* and its relationship to toxin production and virulence in the hamster model." J Med Microbiol **31**(3): 169-174.

Sell, T. L., D. R. Schaberg and F. R. Fekety (1983). "Bacteriophage and bacteriocin typing scheme for *Clostridium difficile*." J Clin Microbiol **17**(6): 1148-1152.

Skurnik, M. and E. Strauch (2006). "Phage therapy: facts and fiction." Int J Med Microbiol **296**(1): 5-14.

Spigaglia, P., R. Cardines, S. Rossi, M. G. Menozzi and P. Mastrantonio (2001). "Molecular typing and long-term comparison of *Clostridium difficile* strains by pulsed-field gel electrophoresis and PCR-ribotyping." J Med Microbiol **50**(5): 407-414.

Spigaglia, P. and P. Mastrantonio (2004). "Comparative analysis of *Clostridium difficile* clinical isolates belonging to different genetic lineages and time periods." J Med Microbiol **53**(Pt 11): 1129-1136.

Stabler, R. A., D. N. Gerding, J. G. Songer, D. Drudy, J. S. Brazier, H. T. Trinh, A. A. Witney, J. Hinds and B. W. Wren (2006). "Comparative phylogenomics of *Clostridium difficile* reveals clade specificity and microevolution of hypervirulent strains." J Bacteriol **188**(20): 7297-7305.

Stabler, R. A., M. He, L. Dawson, M. Martin, E. Valiente, C. Corton, T. D. Lawley, M. Sebahia, M. A. Quail, G. Rose, D. N. Gerding, M. Gibert, M. R. Popoff, J. Parkhill, G. Dougan and B. W. Wren (2009). "Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the evolution of a hypervirulent bacterium." Genome Biol **10**(9): R102.

Tan, K. S., B. Y. Wee and K. P. Song (2001). "Evidence for holin function of *tcdE* gene in the pathogenicity of *Clostridium difficile*." J Med Microbiol **50**(7): 613-619.

Tanner, H. E., K. J. Hardy and P. M. Hawkey (2010). "Coexistence of multiple multilocus variable-number tandem-repeat analysis subtypes of *Clostridium difficile* PCR ribotype 027 strains within fecal specimens." J Clin Microbiol **48**(3): 985-987.

Tasteyre, A., M. C. Barc, A. Collignon, H. Boureau and T. Karjalainen (2001). "Role of FliC and FliD flagellar proteins of *Clostridium difficile* in adherence and gut colonization." Infect Immun **69**(12): 7937-7940.

van den Berg, R. J., I. Schaap, K. E. Templeton, C. H. Klaassen and E. J. Kuijper (2007). "Typing and subtyping of *Clostridium difficile* isolates by using multiple-locus variable-number tandem-repeat analysis." J Clin Microbiol **45**(3): 1024-1028.

Vander Byl, C. and A. M. Kropinski (2000). "Sequence of the genome of *Salmonella* bacteriophage P22." J Bacteriol **182**(22): 6472-6481.

Varvio, S. L., K. Auranen, E. Arjas and P. H. Makela (2009). "Evolution of the capsular regulatory genes in *Streptococcus pneumoniae*." J Infect Dis **200**(7): 1144-1151.

Ventura, M., C. Canchaya, D. Pridmore, B. Berger and H. Brussow (2003). "Integration and distribution of *Lactobacillus johnsonii* prophages." J Bacteriol **185**(15): 4603-4608.

Verdoorn, B. P., R. Orenstein, J. E. Rosenblatt, L. M. Sloan, C. D. Schleck, W. S. Harmsen, L. M. Nyre and R. Patel (2010). "High prevalence of tcdC deletion-carrying *Clostridium difficile* and lack of association with disease severity." Diagn Microbiol Infect Dis **66**(1): 24-28.

Voth, D. E. and J. D. Ballard (2005). "*Clostridium difficile* toxins: mechanism of action and role in disease." Clin Microbiol Rev **18**(2): 247-263.

Wagner, P. L., D. W. Acheson and M. K. Waldor (2001). "Human neutrophils and their products induce Shiga toxin production by enterohemorrhagic *Escherichia coli*." Infect Immun **69**(3): 1934-1937.

Waldor, M. K. and D. I. Friedman (2005). "Phage regulatory circuits and virulence gene expression." Curr Opin Microbiol **8**(4): 459-465.

Waldor, M. K. and J. J. Mekalanos (1996). "Lysogenic conversion by a filamentous phage encoding cholera toxin." Science **272**(5270): 1910-1914.

Waligora, A. J., C. Hennequin, P. Mullany, P. Bourlioux, A. Collignon and T. Karjalainen (2001). "Characterization of a cell surface protein of *Clostridium difficile* with adhesive properties." Infect Immun **69**(4): 2144-2153.

Warny, M., J. Pepin, A. Fang, G. Killgore, A. Thompson, J. Brazier, E. Frost and L. C. McDonald (2005). "Toxin production by an emerging strain of *Clostridium difficile*

associated with outbreaks of severe disease in North America and Europe." Lancet **366**(9491): 1079-1084.

Weeks, C. R. and J. J. Ferretti (1984). "The gene for type A streptococcal exotoxin (erythrogenic toxin) is located in bacteriophage T12." Infect Immun **46**(2): 531-536.

Weiss, K., L. Bergeron, H. Bernatchez, M. Goyette, M. Savoie and D. Thirion (2007). "*Clostridium difficile*-associated diarrhoea rates and global antibiotic consumption in five Quebec institutions from 2001 to 2004." Int J Antimicrob Agents **30**(4): 309-314.

Wershil, B. K., I. Castagliuolo and C. Pothoulakis (1998). "Direct evidence of mast cell involvement in *Clostridium difficile* toxin A-induced enteritis in mice." Gastroenterology **114**(5): 956-964.

Wilcox, M. H. and W. N. Fawley (2000). "Hospital disinfectants and spore formation by *Clostridium difficile*." Lancet **356**(9238): 1324.

Wright, A. (1971). "Mechanism of conversion of the salmonella O antigen by bacteriophage epsilon 34." J Bacteriol **105**(3): 927-936.