



HAL
open science

Rôle des ARNs non codants dans la pathogénèse des infections liées à un enteropathogène émergent humain, *Clostridium difficile*

Victor Kreis

► To cite this version:

Victor Kreis. Rôle des ARNs non codants dans la pathogénèse des infections liées à un enteropathogène émergent humain, *Clostridium difficile*. Bactériologie. Université Paris-Saclay, 2022. Français. NNT : 2022UPASL005 . tel-04114467

HAL Id: tel-04114467

<https://theses.hal.science/tel-04114467>

Submitted on 2 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Rôle des ARNs non codants dans la
pathogénèse des infections liées à un
entéropathogène émergent humain,
Clostridioides difficile

*Role of non-coding RNAs in the pathogenesis of infections related to an
emerging human enteropathogen, Clostridioides difficile*

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 577, Structure et dynamique des systèmes vivants (SDSV)
Spécialité de doctorat : Microbiologie
Graduate School : Life Science and Health. Référent : Faculté des sciences d'Orsay.

Thèse préparée dans l'unité de recherche **Institute for Integrative Biology of the
Cell (I2BC) (Université Paris-Saclay, CEA, CNRS)**, sous la direction de **Olga
SOUTOURINA**, Professeure, et la co-direction de **Claire JANOIR-JOUVESHOMME**,
Professeure.

Thèse soutenue à Paris-Saclay, le 1^{er} février 2022, par

Victor KREIS

Composition du Jury

Fabienne MALAGNAC Professeur, Université Paris-Saclay, I2BC	Présidente
Laurent DEBARBIEUX DR HDR, Institut Pasteur Paris, Université de Paris	Rapporteur & Examineur
Florence HOMMAIS MCU HDR, Université de Lyon, INSA, UCBL	Rapporteur & Examinatrice
Frédéric BARBUT PU-PH, Université de Paris, Hôpital Saint- Antoine, Paris	Examineur
Jean-Marc GHIGO Professeur, Institut Pasteur, Paris, Université de Paris	Examineur
Olga SOUTOURINA Professeur, Université Paris-Saclay, I2BC	Directrice de thèse

Titre : Rôle des ARNs non codants dans la pathogénèse des infections liées à un entéropathogène émergent humain, *Clostridioides difficile*

Mots clés : Régulation de l'expression des gènes, ARNs régulateurs, Bactéries pathogènes, ARN antisens, Protéine chaperon à l'ARN Hfq, *Clostridioides difficile*

Résumé :

Clostridioides difficile est une bactérie pathogène à Gram positif retrouvée principalement dans les sols et les intestins des mammifères. Il est la principale cause de diarrhées associées aux soins chez les adultes dans les pays industrialisés. L'incidence de ces infections continue à augmenter et cette tendance est accentuée par le vieillissement général de la population. *C. difficile* représente aujourd'hui un réel danger pour la santé humaine et animale. Mieux comprendre la régulation du processus de colonisation du tube digestif semble donc indispensable pour l'étude de ce pathogène émergent. Chez la plupart des bactéries, les ARN non codants (ARNnc) interviennent dans la régulation de nombreux processus physiologiques, dont les processus de virulence. Des expériences récentes de transcriptomique menées par mon laboratoire ont permis d'identifier un grand nombre d'ARNs régulateurs chez *C. difficile*. L'objectif de cette thèse est de caractériser le rôle des ARNs non codant dans le contrôle du cycle infectieux chez *C. difficile*, mais également de mieux comprendre les réseaux de régulation contrôlant les processus indispensables au développement de *C. difficile* au contact de son hôte. Pour ce faire, une approche de transcriptomique permettant de visualiser simultanément les ARNs induits ou réprimés chez l'hôte et le pathogène durant l'infection a été mise en œuvre, ce qui a permis d'identifier certains ARNs différenciellement exprimés au cours de

l'infection par rapport à une condition contrôle. D'autre part, plusieurs ARNs régulateurs spécifiques d'une souche hypervirulente de PCR-ribotype 027 avaient déjà été identifiés lors d'une expérience de transcriptomique comparative entre cette même souche et une souche de référence 630. L'étude de l'un de ces ARN non codant localisée dans une région prophagique a permis de caractériser un nouveau *riboswitch*, situé en amont d'un gène intervenant dans un processus de défense contre l'infection phagique (Abi). Finalement, un dernier aspect original du contrôle de la virulence par *C. difficile* est également exploré par l'étude du rôle d'une protéine ARN chaperonne (Hfq) connue pour être impliquées dans de nombreux réseaux de régulations de la physiologie bactérienne, mais également dans le contrôle des processus de virulence chez la plupart des bactéries à Gram négatif. La caractérisation d'un mutant de délétion pour cette protéine a permis d'étudier le rôle de cette protéine chez *C. difficile* dans la régulation de la virulence via la caractérisation de divers phénotypes liés à la virulence et durant des tests *in vivo* chez la souris. Ce projet a ainsi permis une meilleure compréhension des mécanismes de régulation de certains aspects originaux du contrôle de la virulence par ARN non codant chez *C. difficile*.

Title : Role of non-coding RNAs in the pathogenesis of infections related to an emerging human enteropathogen, *Clostridioides difficile*

Keywords : Regulation of gene expression, Regulatory RNAs, Pathogenic bacteria, Antisense RNA, RNA chaperone protein Hfq, *Clostridioides difficile*

Abstract :

Clostridioides difficile is a Gram-positive pathogenic bacterium found mainly in the soils and intestines of mammals. It is the leading cause of adult health care associated diarrhea in industrialized countries. The incidence of these infections continues to increase and this trend is accentuated by the general aging of the population. *C. difficile* now represents a real danger to human and animal health. A better understanding of the regulation of the colonization process of the digestive tract therefore seems essential for the study of this emerging pathogen. In most bacteria, non-coding RNAs (ncRNAs) govern the regulation of many physiological processes, including virulence processes. Recent transcriptomic experiments initiated by my laboratory have identified a large number of regulatory RNAs in *C. difficile*. The objective of this thesis is to characterize the role of non-coding RNAs in the control of the infectious cycle in *C. difficile*, but also to better understand the regulatory networks controlling the processes essential for the development of *C. difficile* in contact with its host. To do this, a transcriptomic approach making it possible to simultaneously visualize the RNAs induced or repressed in the host and the pathogen during the infection was implemented, which made it possible to identify certain RNAs differentially expressed

during the infection versus a controlled condition. On the other hand, several regulatory RNAs specific to a hypervirulent ribotype 027 strain had already been identified during a comparative transcriptomic experiment between this same strain and a reference strain 630. The study of one of these non-coding RNAs located in a prophagic region made it possible to characterize a new *riboswitch*, located upstream of a gene involved in a process of defense against phage infection (Abi). Finally, a last original aspect of virulence control by *C. difficile* is also explored by studying the role of an RNA chaperone protein (Hfq) known to be involved in numerous regulatory networks of bacterial physiology, but also in the control of virulence processes in most Gram-negative bacteria. The characterization of a deletion mutant for this protein made it possible to study the role of this protein in *C. difficile* in the regulation of virulence through the characterization of various phenotypes related to virulence and during *in vivo* tests in mice. This project has thus enabled a better understanding of the regulatory mechanisms of certain original aspects of virulence control by non-coding RNA in *C. difficile*.

Remerciements

Je remercie vivement les rapporteurs de cette thèse, Madame le Docteur Florence Hommais et Monsieur le Docteur Laurent Debarbieux pour le temps qu'ils ont consacré à la lecture de mon manuscrit et à l'évaluation de mon travail. Je tiens également à remercier Madame le Professeur Fabienne Malagnac, Monsieur le Professeur Frédéric Barbut et Monsieur le Professeur Jean-Marc Ghigo pour l'intérêt qu'ils ont porté à mes travaux, et pour avoir accepté de les examiner. Je joins mes remerciements à mes deux directrices de thèse, Madame le Professeur Olga Soutourina et Madame le Professeur Claire Janoir-Jouveshomme, pour m'avoir accueilli dans leur laboratoire ainsi que pour le soutien dont elles ont fait preuve lors de ces trois années de thèse.

Aux membres de l'équipe ARNCLO d'Orsay, dans laquelle j'ai réalisé la majeure partie de mes travaux de recherche, merci pour votre soutien à la fois scientifique et moral. Anaïs, ma chère amie, merci pour tous ces merveilleux moments de rire et de joie passés en ta compagnie. Tu as été un pilier de mon doctorat, et travailler avec toi fut un réel plaisir. Emilie, merci pour ta gentillesse et ton soutien moral lors des moments difficiles à la fin de ma thèse. Claire, tu as fait preuve d'une patience et d'une pédagogie remarquables à mon égard. Merci pour tous tes conseils et pour le temps que tu as passé à me former. Merci aussi pour tout le travail d'analyse que tu as fourni, et enfin merci d'avoir toujours su te rendre disponible pour discuter de mon projet quand j'en avais besoin. Un très grand merci également à Julian, de l'institut Pasteur, pour son enthousiasme et l'aide qu'il a apporté lors des analyses des données de RNA-seq. Je remercie également tous les membres de l'unité BaPs de la Faculté de pharmacie de Chatenay Malabry, sans qui je n'aurais pas pu réaliser les expérimentations sur modèle animal. Cécile, merci pour ta patience, ton aide et ta gentillesse durant ces longues et pénibles journées passées à disséquer des souris. Merci pour avoir su me remonter le moral et rester positive en toutes circonstances. Imad et Jean-Christophe, je vous remercie également pour l'aide et le soutien que vous m'avez apporté lors de la mise au point de l'expérience de dual RNA-seq et pour l'analyse des données obtenues. Sandra, un grand merci pour ton aide précieuse lors de ces longues séries de numération. Je remercie aussi la plateforme de séquençage de l'I2BC, qui a réalisé le séquençage des ARN, pour leur aide lors de la mise en place de l'expérience de RNA-seq. Merci à toutes les personnes qui ont participé de près ou de loin à ce projet et qui m'ont permis d'arriver jusqu'au bout ce doctorat.

Enfin, mes remerciements les plus chaleureux vont à mes proches, qui ont dû essuyer les pires moments de ces trois dernières années. Merci à ma mère, mon père et à ma sœur que j'aime du fond de mon cœur, et à qui je dois tout. Comme vous l'avez toujours fait par le passé, vous avez su m'apporter tout le soutien dont j'avais besoin durant les moments les plus pénibles de ma thèse. Merci Brenna pour ton soutien et ton aide précieuse à la préparation de ma soutenance. Je souhaite également te remercier Sabrina pour le dévouement et le soutien dont tu as fait preuve lorsque ma famille et moi-même étions au plus mal. Merci également à toi, mon cher Alain. Comme toujours, tu as fait preuve d'une amitié sans faille et tu as su me prêter main forte lorsque j'en avais le plus besoin. A Polina, ma chère et tendre, спасибо тебе огромное ! Merci d'avoir illuminé mes journées et supporté mes sauts d'humeur lors des moments les plus éprouvant. Merci à vous tous, sans qui ma vie ne serait pas ce qu'elle est aujourd'hui et qui me donnez la force d'avancer et de pouvoir faire face à toutes les difficultés.

Table des matières

Introduction.....	8
Chapitre 1 : Clostridioides difficile, un entéropathogène majeur	9
I. Infections à Clostridioides difficile	9
1. Origines et caractéristiques bactériologiques	9
2. Épidémiologie des ICD	10
3. Cas particulier de la souche épidémique NAP1/BI/027.....	13
4. Diagnostic des ICD	14
5. Traitement des ICD.....	17
II. Pathogenèse et réponse de l'hôte.....	20
1. Cycle infectieux et physiopathologie.....	20
2. Les facteurs de colonisation	23
3. Les toxines TcdA et TcdB	29
4. La toxine binaire (CDT) : expression, mode d'action et importance	36
5. Sporulation et germination	37
6. Interaction hôte-pathogène dans les ICDs	42
III. C. difficile dans l'univers des bactériophages	45
1. Taxonomie, structure et cycle infectieux des phages	45
2. Les mécanismes de défense contre les phages	47
3. Phages et virulence chez les bactéries	53
4. Les phages chez C. difficile	55
Chapitre 2 : Les riborégulateurs	58
I. Les ARN non codants bactériens : présentation générale.....	58
1. La découverte des ARN non codants (ARNnc).....	58
2. Caractéristiques principales des ARN non codants bactériens	59
3. Rôles physiologiques des ARN non codants bactériens	60
II. Les différents modes d'action des ARN non codants bactériens.....	65
1. Mode d'action par complémentarité de base.....	66
2. Les ARNnc senseurs : thermosenseurs, T-boxes et riboswitch.....	76
3. Les ARNnc interférant avec l'activité d'une protéine.....	79
4. Le système CRISPR-Cas.....	81
III. Les ARN régulateurs chez les bactéries pathogènes	84
1. Bactéries à Gram négatif	86
2. Bactéries à Gram positif	91
3. Le cas de Clostridioides difficile	97
IV. Les ARN régulateurs chez les eucaryotes	100
1. Aperçu des différentes classes et de leurs modes d'action.....	100
2. Comparaison des rôles des ARNnc eucaryotes et procaryotes	104

Projet scientifique.....	109
Résultats.....	111
Partie 1 : Dynamique d'expression des ARNnc chez l'hôte et le pathogène au cours d'une ICD.....	112
I. Principe et mise au point de la technique pour <i>C. difficile</i>	112
1. Principe de la technique de dual RNA-seq	112
2. Expérience de suivi clinique de souris	114
3. Mise au point de l'expérience de dual RNA-seq et de purification des ARNs	116
II. Analyse des données obtenues	120
1. Contrôle qualité et alignement des <i>reads</i>	120
2. Quantification et analyse différentielle des données	123
Partie 2 : Rôle de la protéine Hfq dans la virulence de <i>C. difficile</i>.....	132
I. Construction du mutant de délétion Δhfq	132
II. Étude phénotypique du mutant de délétion <i>hfq</i>	133
1. Étude de la sporulation	134
2. Étude de la formation de biofilms.....	135
3. Étude de la réponse au stress	137
III. Étude du rôle d'Hfq dans la virulence de <i>C. difficile</i>	141
Partie 3 : Caractérisation d'un ARNnc spécifique des souches de ribotype 027.....	144
I. Identification et caractérisation de l'IGR-1461	144
II. Construction et caractérisation du mutant de délétion.....	147
1. Identification des extrémités par 5'-3' RACE.....	147
2. Stratégie de mutagenèse	148
3. Mécanisme de régulation par le riboswitch	150
III. Impact de l'expression de la protéine <i>Abi</i> sur la croissance	154
Discussion et perspectives.....	156
I. Identification des ARN exprimés lors de l'infection.....	157
II. Rôle de la protéine Hfq dans la virulence	164
III. Caractérisation du <i>riboswitch</i> IGR-1461	167
IV. Conclusion	171
Annexes.....	173
Publication 1.....	185
Publication 2.....	186

Liste des abréviations

AMM	Autorisation de mise sur le marché	ICD	Infection à <i>C. difficile</i>	PNP	Para-nitrophénylphosphate
ARNi	ARN interférent	IDSA	<i>Infectious Diseases Society of America</i>	PVRL3	<i>Poliovirus receptor-like protein 3</i>
ARNnc	ARN non codant	IGR	<i>Intergenic Region</i>	QS	<i>Quorum sensing</i>
ARNP	ARN polymérase	IP	Intra-péritonéale	RACE	<i>Rapid Amplification of cDNA-ends</i>
ARNr	ARN ribosomal	ITU	Infections du tractus urinaire	RBS	<i>Ribosome Binding Site</i>
ARNt	ARN de transfert	lncRNA	<i>Long non coding RNA</i>	REA	<i>Restriction Endonuclease Analysis</i>
BREX	<i>Bacteriophage Exclusion</i>	miRNA	<i>Micro RNA</i>	RM	Restriction Modification
CAMPs	Peptides cationiques antimicrobiens	MAPS	<i>MS2-affinity purification</i>	ROI	<i>Reactive Oxygen Intermediate</i>
CCCNA	<i>Cell Culture Cytotoxicity Neutralisation Assay</i>	MLST	<i>Multi Locus Sequence Typing</i>	RT	Ribotype
CDT	<i>C. difficile</i> transférase	MLVA	<i>Multiple Loci VNTR Analysis</i>	SD	Shine-Dalgarno
CFU	<i>Colony Forming Unit</i>	MSCRAMMs	<i>Microbial Surface Components Recognizing Adhesive Matrix Molecules</i>	SHEA	<i>Society for Healthcare Epidemiology of America</i>
CPM	Colite pseudomembraneuse	NAATs	<i>Nucleic Acid Amplification tests</i>	siRNA	<i>Small interfering RNA</i>
CRISPR	<i>Clustered Regularly Interspaced Short Palindromic Repeats</i>	NOD1	<i>Nucleotide Binding Oligomerization Domain 1</i>	SLPs	<i>S-Layer Proteins</i>
crRNA	ARN CRISPR	ORF	<i>Open Reading Frame</i>	TA	Toxine-Antitoxine
CSPG4	<i>Chondroitin sulfate proteoglycan 4</i>	PA	Phosphatase Alcaline	TCD	Toxines Clostridiales
DISARM	<i>Defence Islands System Associated with RM</i>	PAI	Peptide auto-inducteur	TCdA/B	Toxines A et B de <i>C. difficile</i>
dNTP	<i>deoxyribonucleoside triphosphate</i>	PaLoc	<i>Pathogenicity Locus</i>	TCS	<i>Two Component System</i>
EDG	Expression différentielle des gènes	PAM	<i>Protospacer Adjacent Motif</i>	TLRs	<i>Toll Like Receptors</i>
EGM	Eléments génétiques mobiles	PAMPs	<i>Pathogen Associated Molecular Patterns</i>	tracrRNA	<i>Trans-activating crRNA</i>
EHEC	<i>E. coli</i> entérohémorragique	PCR	<i>Polymerase Chain Reaction</i>	TSS	<i>Transcription Start Site</i>
ESCMID	<i>European Society of Clinical Microbiology and Infectious Diseases</i>	PG	Peptidoglycane	UE	Unité Enzymatique
GDH	Glutamate déshydrogénase	piRNA	<i>Piwi RNA</i>	UPEC	<i>E. coli</i> uropathogène
GTD	<i>Glycosyltransferase Domain</i>	PRRs	<i>Pattern Recognition receptors</i>	VNTR	<i>Variable Number of Tandem Repeats</i>

Introduction

I. Infections à *Clostridioides difficile*

1. Origines et caractéristiques bactériologiques

Clostridioides difficile, également connu sous le nom de *Clostridium difficile*¹, est un bacille (de 3 à 15 µm de long) à Gram positif appartenant au phylum des Firmicutes. Il a la capacité de produire des spores et ne peut être cultivé qu'en anaérobiose stricte. En culture en milieu solide, les colonies sont opaques, non hémolytiques, irrégulières et légèrement rhizoïdes (**Figure 1**). On le retrouve de manière ubiquitaire dans l'environnement, essentiellement dans les sols sous forme de spores, mais également dans la partie basse du tube digestif des mammifères².

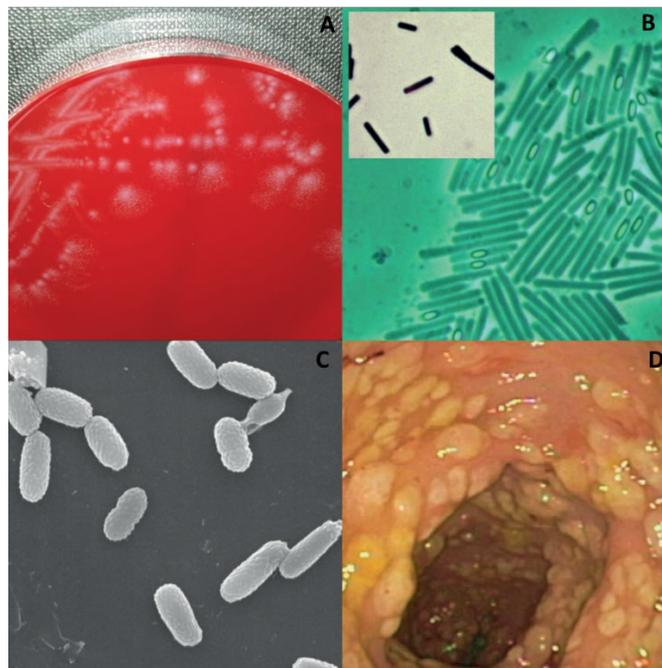


Figure 1 - *Clostridioides (Clostridium) difficile*. (A) Image typique de colonies de *C. difficile* sur une boîte de gélose au sang. (B) Image en microscopie à contraste de phase d'une culture de *C. difficile* montrant des cellules végétatives, des spores à phase sombre (taches sombres subterminales) et des spores à phase lumineuse (ellipsoïdes brillants). Dans l'encadré supérieur gauche est représentée une coloration de Gram de la culture. (C) Microscopie électronique à balayage des spores de *C. difficile*. (D) Endoscopie montrant une colite pseudomembraneuse (CPM) causée par *C. difficile*. Le tissu du côlon sain est rose, les pseudomembranes résultant d'une infection à *C. difficile* sont jaunes.²

Il fut décrit pour la première fois en 1935 comme faisant partie de la flore intestinale des nouveau-nés³ et est alors baptisé *Bacillus difficilis* en raison de sa difficulté à être isolé et de sa croissance lente lors des cultures en laboratoire. Son nom fut cependant changé en

Clostridium difficile dans les années 1970s. Si les auteurs de cette étude avaient déjà constaté une toxicité de la bactérie chez les cochons d'inde et les lapins, sa virulence complète ne fut décrite que bien plus tard. Plusieurs études avaient fait état de colites pseudomembraneuses (CPM) létales associées à un traitement antibiotique et à la présence d'une toxine dans les fèces de sujets ayant reçu un traitement antibiotique ⁴⁻⁶, mais il faudra attendre la fin des années 1970 avant qu'un lien ne soit établi entre *C. difficile*, la toxine et les lésions observées lors de ces études ⁷⁻⁹. Aujourd'hui, les infections à *C. difficile* (ICD) sont reconnues comme étant les causes principales de diarrhées associées aux soins chez l'adulte dans les pays industrialisés ¹⁰⁻¹².

2. Épidémiologie des ICD

C. difficile est un agent pathogène opportuniste associé aux soins, entraînant des désordres intestinaux plus ou moins sévères. Il est retrouvé en portage asymptomatique chez 66% des nourrissons et enfants en bas âge. Ce paradoxe infantile pourrait être dû au fait que tous les récepteurs coliques ne soient pas encore arrivés à maturité ¹³. La fréquence de portage asymptomatique au sein de la population adulte est comprise entre 4% et 15% et augmente considérablement au sein des établissements d'hébergement de personnes âgées, et lors des périodes prolongées d'hospitalisation ¹³⁻¹⁵. La raison pour laquelle certains patients ne sont pas affectés par la présence de *C. difficile* alors que d'autres développent des formes graves d'ICD reste encore mal comprise ¹⁴. Une étude a cependant mis en évidence le lien entre la concentration sérique élevée d'IgG contre la toxine A et l'absence de symptômes chez des patients colonisés par *C. difficile* ¹⁶. Le risque de développer une ICD est fortement augmenté chez les patients ayant un système immunitaire affaibli et également lors de la prise d'un traitement antibiotique, ou à la suite d'une période d'hospitalisation prolongée. Un âge supérieur à 65 ans est également associé à un risque accru de développer une ICD ¹². Bien que la plupart des ICD soient classiquement associées à une exposition au système de santé, des études récentes ont mis en évidence une augmentation du nombre d'ICD communautaires, pouvant représenter jusqu'à 30% des cas d'ICD ¹⁷⁻¹⁹. De plus, une catégorie de souches hypervirulentes de PCR-ribotype 078 semble être associée le plus souvent à des populations jeunes et à des ICD communautaires ²⁰.

Depuis le début des années 1980, l'épidémiologie de *C. difficile* est étudiée grâce à de nombreux outils de typage moléculaire ^{21,22}. Les techniques d'analyse moléculaire classiquement utilisées pour la caractérisation des souches de *C. difficile* sont notamment le toxinotypage qui se base sur le polymorphisme des séquences des toxines A et B, le PCR-ribotypage utilisant la variation de taille d'espacement intergénique entre les gènes codant pour les ARN ribosomiaux (ARNr) 16S et 23S, l'analyse du polymorphisme de longueur des fragments de restriction (ou REA de l'anglais *Restriction Endonuclease Analysis*), le typage par séquençage MLST (*Multi Locus Sequence Typing*) qui consiste à séquencer 5 à 7 gènes de ménage, le MLVA (pour *Multiple Loci VNTR Analysis*) qui analyse le polymorphisme des motifs VNTR (*Variable Number of Tandem Repeats*). *C. difficile* présente de nombreux PCR-ribotypes (RTs) répartis dans différents pays, régions et aussi hôpitaux et qui sont dominés par une petite partie d'entre eux ¹⁴. Par exemple, une étude de l'*European Society of Clinical Microbiology and infectious Diseases* (ESCMID) menée en 2014 sur la répartition des RTs de 953 isolats de *C. difficile* isolés dans 22 pays européens a permis d'identifier 99 PCR-ribotypes, dont les RTs majoritaires 027, 014, 001/072, et 078 ²³. Une étude européenne plus large menée en 2012-2013 impliquant la participation de 482 hôpitaux de 19 pays différents et 1196 isolats de *C. difficile* avait identifié 125 RTs répartis en Europe, avec les mêmes RTs majoritaires que l'étude précédemment citée (**Figure 2**) ²⁴. Une étude réalisée en 2013 à l'échelle nationale a identifié les RTs 014/020/077 (18,7% – 21,9 %) et 078/126 (9,5% – 12,1 %) comme les RTs prédominants sur le territoire français alors que la prévalence du RT 027 était faible (3 %), principalement isolé au Nord de la France ²⁵.

Au début du 21^e siècle, l'épidémiologie de *C. difficile* a considérablement changé ^{26,27} avec une forte augmentation de l'incidence et de la sévérité des ICD, notamment en Europe ²⁸ et en Amérique du Nord ^{29,30}, accompagnée de complications et de taux de mortalité élevés ³¹⁻³³. La mortalité associée aux ICDs (toutes causes confondues) a été estimée entre 15% et 25% dans le mois suivant la survenue de l'infection ³⁰ et une forte augmentation de l'incidence des ICD dans les hôpitaux aux États-Unis, de presque un facteur deux, a été observée entre 1996 et 2003 ³⁴. Une autre étude faite au Canada a révélé que la fréquence des diarrhées et CPM en 2003 au sein de la population globale avait été multipliée

par 4 depuis 1991 et par 10 depuis 1938 ³⁵. Cette dernière étude montre également une augmentation de la survenue des symptômes liés à une ICD chez les patients hospitalisés, passant de 3 pour 1000 patients en 1991 à 43 pour 1000 patients en 2004, mais également une aggravation de ces symptômes. Une étude européenne menée en 2008 dans 34 pays sur 455 patients a rapporté un taux de mortalité de 22% dont 40% sont morts de l'infection ³⁶. En France en 2009, le taux de mortalité associé aux ICDs a été estimé à 4% ²⁵. L'augmentation du nombre et de la sévérité des ICDs a également des répercussions économiques importantes. En effet l'augmentation de la durée des hospitalisations, les réadmissions liées aux rechutes, les tests en laboratoire et la prescription accrue de médicaments ont pour conséquence une augmentation du coût global associé aux ICDs ²⁴. Aux États-Unis par exemple, le coût global directement lié aux ICDs a été estimé à 4,8 milliards de dollars en 2008 ³⁷ avec un coût d'hospitalisation moyen de 24 400\$ ¹⁷. La durée d'hospitalisation impliquant un diagnostic secondaire d'ICD est plus de deux fois plus longue qu'une hospitalisation pour une ICD en diagnostic initial et est environ trois fois plus coûteuse ¹⁷.

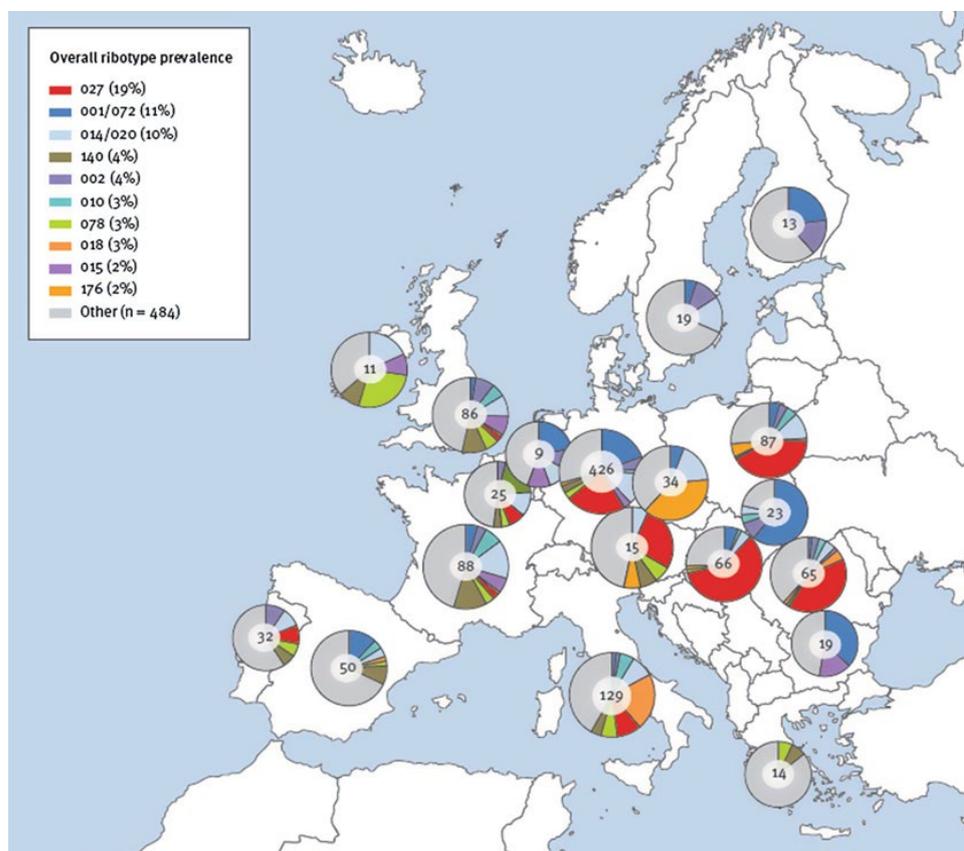


Figure 2 - Répartition géographique des PCR-ribotypes de *C. difficile*, par pays européen en 2013. ²⁴

3. Cas particulier de la souche épidémique NAP1/BI/027

Comme indiqué précédemment, depuis le début des années 2000, l'incidence des ICDs a considérablement augmenté avec notamment une proportion de formes sévères accrue en Europe et Amérique du Nord, due à l'émergence puis la dissémination de souches dites hypervirulentes ou épidémiques^{28,38,39} résistantes aux fluoroquinolones, notamment les souches de PCR-ribotype 027. Néanmoins, les bases moléculaires de l'hypervirulence de ces souches restent largement inconnues³⁹. Ces souches épidémiques sont habituellement référencées dans la littérature en tant que souches BI/NAP1/027 pour PCR-ribotype 027 / PFGE (*Pulse Field Gel Electrophoresis*) type NAP1 (*North American pulsotype 1*) / toxinotype III / REA (*Restriction Rendonuclease Analysis*) groupe B1^{10,40}. La première souche RT 027 a été isolée en 1985⁴¹. La recrudescence épidémique de cette souche a débuté au Canada et aux États-Unis^{29,30}. Elle s'est ensuite propagée en Europe en 2005 d'abord en Grande Bretagne⁴², puis dans 16 autres pays européens et a fini par atteindre l'Asie⁴³⁻⁴⁵ et l'Australie⁴⁶ vers les années 2010⁴⁷. En France, les premiers isolats de la souche RT 027 ont été identifiés en 2006 dans le Nord Pas-de-Calais⁴⁸. Aujourd'hui la prévalence des souches NAP1/BI/027 dans les ICDs a récemment été estimée entre 18% et 24%^{49,50}.

Bien que globalement conservés, les génomes des souches hypervirulentes 027 présentent des différences considérables par rapport aux souches non épidémiques. Par exemple, la souche épidémique R20291, isolée en Angleterre, possède 234 gènes de plus que la souche 630, qui pourraient contribuer aux phénotypes différents observés entre ces deux souches comme la motilité, la résistance aux antibiotiques et la toxicité⁵¹. La souche épidémique 027 présente également 5 *loci* génétiques uniques qui ne sont pas retrouvés chez les souches non épidémiques comme la 630, incluant un prophage qui pourrait être lié à l'émergence de cette bactérie hypervirulente⁵¹. Outre les gains de régions observés lors des études comparatives, des pertes de fonction de certains gènes par insertion ont également été décrites, ainsi que des mutations spécifiques et des inversions localisées en amont de régions codantes pouvant ainsi impacter leur expression^{51,52}. Les souches NAP1/BI/027 produisent également une toxine supplémentaire appelée toxine binaire dont le rôle dans la pathogenèse est encore mal compris, et présentent une version délétée de

18 pb du gène régulateur négatif de l'expression des toxines *tcdC* avec en plus une délétion ponctuelle à la position 117 qui induit la production d'une forme tronquée de la protéine TcdC qui ne peut plus exercer son rôle répresseur sur la production des deux toxines majeures TcdA et TcdB⁵³. Toutefois, le lien entre ces modifications du facteur TcdC chez les souches épidémiques de RT 027 et l'augmentation des toxines est encore discuté⁵⁴⁻⁵⁶. La différence dans l'efficacité et la cinétique de sporulation des souches épidémiques par rapport aux souches non épidémiques est également débattue. Des études ont mis en évidence une augmentation de la production de spores chez des souches épidémiques par rapport à des souches non épidémiques contribuant ainsi à l'augmentation de la persistance de ces souches dans l'intestin de l'hôte^{57,58}. Cependant une autre étude utilisant un plus grand nombre d'échantillons n'a pas constaté d'augmentation de sporulation chez les souches de RT 027 par rapport aux souches n'appartenant pas à ce RT⁵⁹. Enfin, une étude a montré que la souche R20291 (RT 027) présente une motilité plus importante que la souche de référence 630 et qu'elle s'auto-agglutine moins⁵¹.

4. Diagnostic des ICD

Comme pour toute pathologie, la réalisation d'un diagnostic correct d'une ICD repose sur un contexte clinique adapté, un historique de la prise de traitement antibiotique par le patient et la réalisation de tests appropriés. La recherche de *C. difficile* chez des patients présentant les symptômes d'une ICD se fait grâce à différents tests en laboratoire ou moins fréquemment par endoscopie (sigmoïdoscopie et colonoscopie). Même si cette dernière méthode est requise pour un diagnostic spécifique de CPM, elle est très peu utilisée pour un diagnostic initial dans la mesure où elle est très invasive, inconfortable pour le patient, chère et peu sensible⁶⁰. Il existe deux techniques de laboratoires majeures pour l'établissement d'un diagnostic d'ICD, mais il n'y a pas de consensus quant à une méthode *gold standard*^{61,62}.

La première de ces deux techniques, appelée CCCNA (pour *Cell Culture Cytotoxicity Neutralisation Assay*) permet de détecter la présence des toxines clostridiales (TCD) dans les selles du patient. La technique consiste à inoculer le filtrât d'une suspension de selles dans une culture de lignée cellulaire (type Vero, HeLa ou Hep-2), et d'observer 24h à 48h après

(voir 4h à 6h dans les cas les plus sévères) un éventuel effet cytotoxique causé par la perturbation de la structure du cytosquelette qui entraîne un arrondissement de la morphologie cellulaire⁶³. Cet effet est principalement dû à la toxine B qui est 1000 fois plus efficace que la toxine A⁶⁴. Si un effet cytotoxique est constaté alors un test de neutralisation est effectué afin de s'assurer que l'effet observé est lié à des TCD. Cette méthode a longtemps été considérée comme le *gold standard* grâce à sa forte spécificité⁶⁵. Cependant, la non détection des souches TcdA+/B- par cette technique⁶⁶, sa faible sensibilité (60% à 90% selon les études) et la nécessité d'avoir recours à des lignées cellulaires pouvant être onéreuses l'ont peu à peu écartée des normes diagnostiques⁶³. La seconde technique utilisée dans le diagnostic des ICD permet d'isoler *C. difficile* directement à partir d'une culture d'un échantillon de selles, et de déterminer par la suite si l'isolat récupéré est une souche toxigène ou non. Après ensemencement des selles sur un milieu sélectif et incubation durant 48h, les colonies présentant des caractéristiques microbiologiques semblables à celles de *C. difficile* sont repiquées. Une fois que *C. difficile* a été isolé, sa capacité à produire des toxines doit être testée *in vitro*. Cela peut être effectué en utilisant la méthode CCCNA décrite précédemment, soit par des méthodes immunologiques comme le test ELISA permettant de détecter la présence de la toxine A seule ou des deux toxines A et B, soit par des méthodes d'amplification d'acide nucléique (NAAT pour *Nucleic Acids Amplification Tests*) permettant la détection des gènes codant pour les TCD ou d'autres gènes spécifiques à *C. difficile*⁶¹. Cette dernière méthode peut cependant être longue (peut aller jusqu'à 7 jours) et ne permet pas de différencier le portage asymptomatique des souches toxigènes de *C. difficile* des ICD symptomatiques⁶⁷.

Les deux techniques décrites précédemment peuvent donc être coûteuses, difficiles à mettre en place et nécessitent du personnel qualifié pour leur réalisation. Des méthodes plus accessibles ont donc été développées et commercialisées sous forme de kit pour un diagnostic rapide des ICD. Chacune de ces méthodes présente cependant des avantages et des inconvénients. La détection d'une souche de *C. difficile* peut se faire rapidement par la détection d'antigènes (grâce à l'utilisation d'anticorps) dans des échantillons de selles de patients en utilisant différentes approches. Les kits EIA (*Enzyme Immunoassays*) pour la

détection des TCD ont été très utilisés pour le diagnostic des ICD en raison de leur rapidité (résultats en quelques heures), de leur faible coût et de leur facilité d'utilisation. Les kits ne détectant que la toxine A qui ne permettent pas la détection des souches ne produisant que la toxine B, ont été pour la majorité d'entre eux remplacés par des kits détectant les deux toxines⁶⁸. Ces tests présentent une excellente spécificité allant de 95% à 100% mais présentent en revanche une faible sensibilité, de 60% à 80% (variant selon les kits et les études)⁶⁹⁻⁷¹ ce qui a pour conséquence une proportion non négligeable de faux négatifs^{13,72}. Compte tenu de cette faible sensibilité, plusieurs laboratoires ont utilisé la glutamate déshydrogénase (GDH) comme marqueur, qui, combinée à un test de détection des toxines, permet d'obtenir la sensibilité supérieure à 90%^{73,74}.

Les NAATs (pour *Nucleic Acid Amplification tests*) sont des méthodes de diagnostic plus récentes basées sur l'amplification d'acides nucléique comme la PCR (*Polymerase Chain Reaction*), la PCR en temps réel (RT-PCR) ou la LAMP⁷⁵ (pour *Loop-Mediated Isothermal Amplification*) qui permettent d'identifier les gènes codant les TCD. Bien que plus onéreuses que les EIA, ce sont des méthodes très rapides (1 à 2 h) présentant une sensibilité supérieure à 90%^{61,73,76}. De plus, la possibilité de réaliser un génotypage simultané confère à ces méthodes un avantage considérable face aux tests immunologiques, permettant une identification rapide des souches épidémiques afin de contrôler leur éventuelle dissémination. Cependant, les NAATs ne permettent que la détection des gènes des TCD ce qui empêche la détection du portage asymptomatique de souches non toxigènes⁷⁷. De plus, l'utilisation de ces seules méthodes ne permet pas de faire la différence entre l'infection et la colonisation, ce qui peut avoir pour conséquence un diagnostic par excès (faux positif) et un traitement inapproprié⁷⁸.

Compte tenu de ce constat, l'ESCMID et les *Society for Healthcare Epidemiology of America/The Infectious Diseases Society of America* (SHEA/IDSA) recommandent l'utilisation d'algorithmes combinant plusieurs tests à la fois afin de maximiser la précision des diagnostics d'ICD (**Figure 3**)^{68,79}. Le premier test doit être un test à valeur prédictive négative élevée (il peut s'agir d'une GDH EIA ou d'un NAAT). Le deuxième test doit avoir une valeur prédictive positive élevée comme un test EIA TCDA/B. Si le premier test est négatif

alors la CDI est exclue. En revanche, s'il est positif, un deuxième test EIA TCDA/B doit être effectué. Si ce dernier est positif alors la CDI est confirmée et s'il est négatif, le cas doit être évalué cliniquement afin de trancher entre les trois cas suivants possibles : (1) une CDI avec un niveau de toxine inférieur au seuil de détection, (2) un faux négatif ou (3) un portage asymptomatique de *C. difficile*.

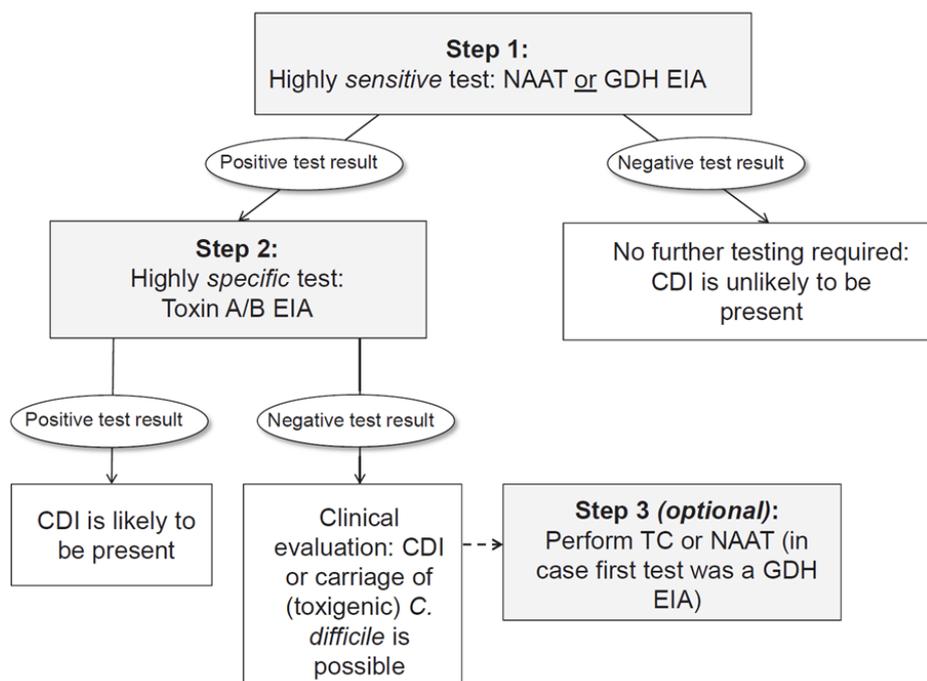


Figure 3 – Algorithme décisionnel recommandé par l’ESCMID pour le diagnostic d’une infection à *C. difficile*.⁶⁸

5. Traitement des ICD

D’une manière générale, la prise en charge des ICD consiste dans un premier temps à stopper le traitement antibiotique inducteur et à le remplacer, lorsque cela est possible, par une famille d’antibiotiques moins à risque de favoriser une ICD (aminoglycosides, macrolides, tétracycline, vancomycine), ainsi que les inhibiteurs de la pompe à protons favorisant les réinfections par le biais de l’augmentation du pH gastrique^{80,81}. Ce premier geste thérapeutique ainsi qu’une prise en charge des troubles électrolytiques, peuvent conduire chez une minorité de patient à une résolution spontanée de l’infection⁸². En revanche, la non interruption du traitement antibiotique responsable de l’ICD peut non seulement entraîner l’aggravation de la condition du patient mais également favoriser les rechutes par la suite⁸³. Une fois que le diagnostic d’une ICD a été établi chez un patient hospitalisé, il est indispensable de mettre rapidement en place les mesures de prévention

de propagation du germe dans l'établissement de santé. Cela inclut l'isolement géographique du patient infecté, le lavage systématique des mains à l'eau et au savon, la désinfection appropriée quotidienne de la chambre et après la sortie du patient, la gestion des excréta, l'interruption des traitements ralentisseurs du transit et la prise en charge des troubles hydroélectrolytiques associés à la diarrhée ^{27,84}. Les mesures de prévention et de contrôle du risque sanitaire ont été publiées en détail par la SHEA/IDSA ⁷⁹ et la ESCMID ^{85,86}.

Chez les patients symptomatiques (plus de 3 selles non moulées par jour) ayant une infection à *C. difficile* toxigène confirmée microbiologiquement, un traitement antibiotique approprié doit être initié ⁸⁷. Cependant, les patients porteurs ne présentant aucun signe clinique d'une ICD ne doivent pas recevoir de traitement antibiotique dans la mesure où un portage de souche non toxigène peut protéger contre le développement de formes symptomatiques ⁸⁸⁻⁹⁰. Il est important de noter que le traitement des ICD doit être adapté à la sévérité de la maladie. Il faut donc identifier les variables cliniques permettant d'estimer la gravité de l'ICD et l'éventuelle survenue de complications par le biais d'un système de *scoring*. Une étude a montré qu'une numération leucocytaire > 30 000/mm³ et une augmentation de la créatinémie au-delà de 50% du niveau normal permettaient de prédire des complications de l'ICD ⁹¹. Une autre étude a identifié trois facteurs permettant de prédire les formes graves : (1) une distension abdominale, (2) une numération leucocytaire > 20 000/mm³ et (3) une hypoalbuminémie (<3g/dl) ⁹².

Les dernières recommandations européennes faites par l'ESCMID proposent un algorithme décisionnel de la prise en charge des ICD en fonction de la gravité de l'infection et des risques de rechutes ⁸⁵. Dans les formes d'ICD simples à modérées, le métronidazole qui est un traitement peu onéreux et relativement efficace peut être recommandé en première intention, cependant, si les symptômes ne disparaissent pas après 5 à 7 jours, un traitement par la vancomycine est indiqué ⁸⁰. Cependant, en présence de certains facteurs pouvant favoriser le développement de formes sévères, le traitement par la vancomycine doit être privilégié. En effet, bien qu'il ait été montré que la vancomycine et le métronidazole ont la même efficacité dans le traitement de formes simples, plusieurs études font état d'une efficacité accrue de la vancomycine par rapport au métronidazole dans le cas des formes

sévères⁹³⁻⁹⁵. Pour les formes sévères d'ICD avec complication il est recommandé d'utiliser la vancomycine (250 mg à 500 mg 4 fois par jour) par administration PO ou nasogastrique en combinaison avec du métronidazole par voie intraveineuse (IV)⁷⁹. Depuis 2011, la fidaxomicine a obtenu une autorisation de mise sur le marché (AMM). C'est un antibiotique macrocyclique administrable par voie orale, faiblement absorbé et présentant une concentration minimale inhibitrice basse (CMI₉₀ = 0,25 mg/L) pour *C. difficile*⁸⁴. Cette molécule est également caractérisée par une faible absorption systémique et un spectre d'activité étroit agissant spécifiquement sur les bactéries à Gram positif et sur les *Bacteroides*, ce qui a pour avantage de ne pas perturber de manière significative le microbiote intestinal^{19,84}. Elle ne présente pas de différence d'efficacité significative avec la vancomycine, et est associée à un plus faible taux de récurrence que cette dernière^{96,97}. De plus, contrairement à la vancomycine elle sélectionne moins d'entérocoques résistants (7% versus 30% pour la vancomycine)⁹⁸.

Après un premier épisode d'ICD, 15% à 20% des patients ayant guéri connaîtront un nouvel épisode infectieux à *C. difficile* dans les 3 à 4 semaines suivant le traitement, et ce risque est augmenté avec l'utilisation d'antibiotiques pour le traitement des éventuelles infections récurrentes^{19,27,83,99}. Les ICD récurrentes sont majoritairement des rechutes (*i.e.* persistance de la même souche de *C. difficile*) mais peuvent également être liées à une réinfection par l'acquisition d'une nouvelle souche de *C. difficile*⁹⁹. L'utilisation de traitement standard comme la vancomycine entraîne un déséquilibre du microbiote intestinal ce qui peut permettre à *C. difficile* de conserver un développement normal. De plus, les spores résistantes aux antibiotiques peuvent germer une fois le traitement antibiotique arrêté entraînant ainsi une rechute de l'ICD. La transplantation de microbiote fécal (TMF) a été étudiée comme alternative à un traitement antibiotique standard dans le traitement des multi-récidives d'ICD, dans l'objectif de restaurer un microbiote intestinal sain avec l'utilisation de microorganismes provenant d'un donneur sain. Une revue systématique de 27 études retrouvait de manière globale une efficacité de 92 % sur un effectif global total de 317 patients traités¹⁰⁰. L'utilisation de probiotiques (par exemple *Saccharomyces boulardii*) n'est actuellement pas recommandée en raison du manque de données cliniques sur les

bénéfiques des probiotiques dans le traitement des ICD ^{19,79}. Enfin, dans de rares cas (moins de 5%) la chirurgie (colectomie) peut être indiquée chez des patients ne répondant pas au traitement initial et ayant développé une colite fulminante qui est associée à un taux de mortalité élevée (jusqu'à 40%).

II. Pathogenèse et réponse de l'hôte

C. difficile est tenu pour principal responsable des diarrhées associées au soins et tient une place importante dans les épidémies de diarrhées acquises dans les communautés. L'émergence et la dissémination de certaines souches épidémiques référencées comme hypervirulentes (comme les PCR-ribotypes 027 et 078) lors des deux dernières décennies a nécessité une compréhension plus avancée des mécanismes de virulence afin de pouvoir faire face à cette nouvelle menace épidémiologique. Bien que la plupart des facteurs intervenant dans l'expression de la virulence de *C. difficile* soient aujourd'hui connus, le tableau physiopathologique des ICDs dévoile un caractère multifactoriel de la virulence de *C. difficile* impliquant de nombreux acteurs agissant de concert (production de toxines, sporulation, formation des biofilms...) et responsable de phénotypes variables ¹⁰¹.

1. Cycle infectieux et physiopathologie

Les ICD sont acquises majoritairement par transmission orale des spores de *C. difficile*, présentes en faible densité dans l'environnement et en concentration plus importante au sein des instituts de santé (**Figure 4**). Contrairement à la forme végétative plus fragile, cette forme confère à *C. difficile* de multiples résistances aux stress environnementaux (comme des températures élevées, des pH acides, des composés antimicrobiens, l'oxygène), ce qui lui permet ainsi de subsister pendant de longues périodes à l'extérieur de son hôte ^{102,103}. Les spores sont importantes non seulement pour la survie du pathogène hors de son hôte, mais elles sont également essentielles pour la transmission et la persistance dans l'hôte. Une fois ingérées, les propriétés physiques des spores ¹⁰⁴ permettent à *C. difficile* de franchir la barrière gastrique et d'atteindre l'intestin, où la germination est induite par la présence d'acides biliaires primaires produits par le foie (comme le taurocholate) ¹⁰². Avant la germination, les spores de *C. difficile* sont également capables d'adhérer à la paroi du colon où le processus de germination est inhibé par la présence de certains acides biliaires

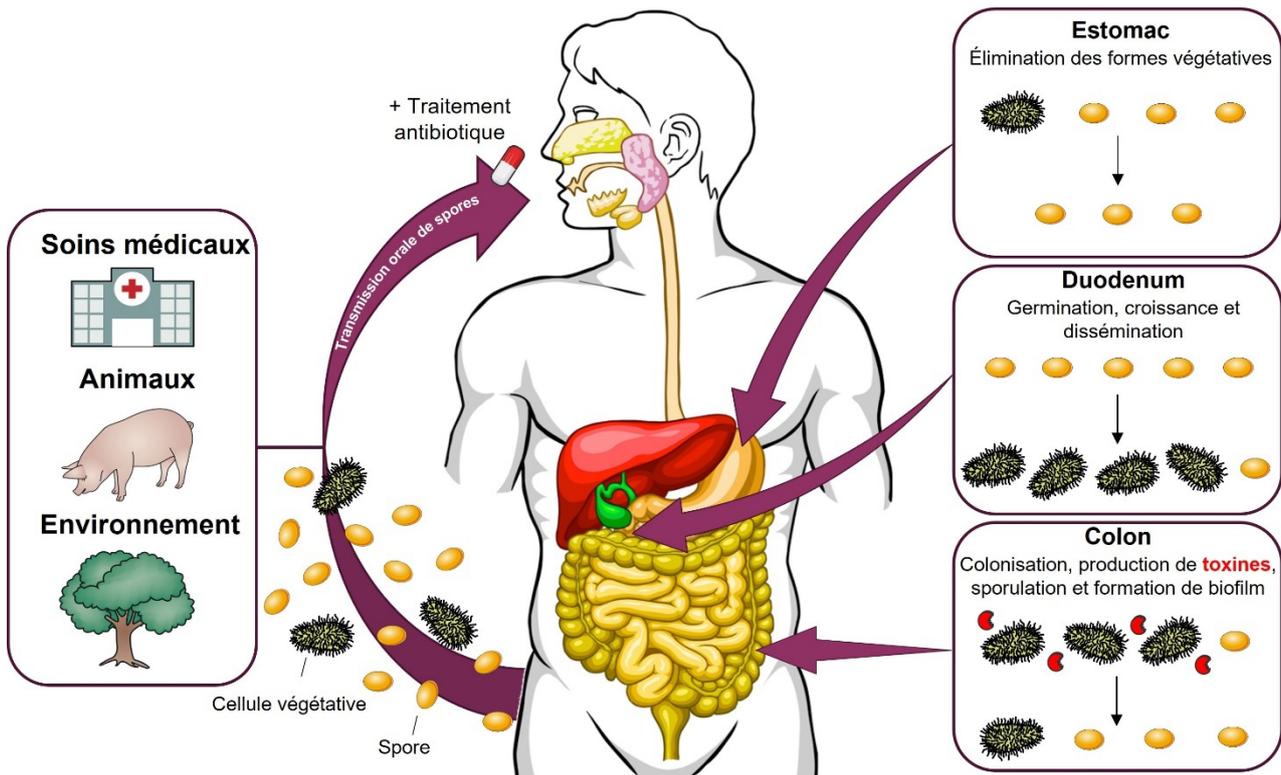


Figure 4 - Représentation du cycle infectieux de *C. difficile* dans le tractus gastro-intestinal humain. Les sources d'infections (soins médicaux, alimentation, animaux et environnement) sont indiquées dans l'encadré de gauche et les différentes étapes du cycle infectieux à droite. Les toxines sont représentées en rouge dans l'encadré du colon.²

C. difficile est un pathogène opportuniste, cela signifie qu'il ne peut s'implanter et se développer dans le tractus intestinal de son hôte qu'en condition de dysbiose, c'est-à-dire en cas de déséquilibre du microbiote intestinal¹⁰⁶⁻¹⁰⁸. Les microbiotes sont des ensembles complexes de microorganismes colonisant les tissus ou les cavités naturelles d'un être vivant ou d'un environnement. Le microbiote intestinal est constitué de l'ensemble des microorganismes peuplant le système digestif. Il permet notamment l'assimilation des nutriments, la modulation du système immunitaire¹⁰⁹ et la lutte contre le développement d'espèces pathogènes^{110,111}. Dans les ICD, la propension des spores à germer et coloniser l'épithélium intestinal est étroitement liée au microbiote intestinal et à son métabolome^{106,112}. Il a été montré que la capacité d'un individu à résister à la colonisation par *C. difficile* dépend de plusieurs processus directement liés au microbiote intestinal qui interviennent de manière concomitante et/ou synergique afin d'empêcher l'implantation et le développement du pathogène dans le tractus intestinal^{106,113,114}. Il peut s'agir par exemple

de la conversion d'acides biliaires primaires en acides biliaires secondaires ¹¹⁵, de la production de peptides antimicrobiens par certaines espèces bactériennes du microbiote ¹¹⁶, ou encore d'une exclusion compétitive de l'environnement microbien ^{117,118}. Les mécanismes de compétition nutritionnelle entre les bactéries du microbiote et *C. difficile* conditionnent également l'implantation et le développement du pathogène. Des études récentes ont mis en évidence la modification de l'environnement nutritionnel par l'inflammation induite par *C. difficile* afin d'exclure les compétiteurs du microbiote ^{119,120}. Ainsi, l'altération de la richesse et de la diversité microbienne du microbiote intestinal entraîne une perte de la résistance à la colonisation de *C. difficile* et favorise son implantation ^{121,122}. L'utilisation des antibiotiques à large spectre constitue donc un risque majeur de développer une ICD, notamment l'utilisation de certains antibiotiques comme la clindamycine, les céphalosporines et les fluoroquinolones ^{123,124}. Cependant, d'autres facteurs peuvent engendrer une dysbiose et favoriser le développement de *C. difficile* comme un âge avancé (supérieur à 65 ans) ¹², ou encore l'utilisation d'agents diminuant l'acidité gastrique comme les inhibiteurs de la pompe à protons ou l'affaiblissement du système immunitaire/immunodépression.

Une fois l'équilibre du microbiote rompu, *C. difficile* est capable d'adhérer à l'épithélium intestinal et de coloniser le colon. Bien que l'ensemble du colon puisse être affecté par une ICD, il semblerait que la partie distale soit la plus communément envahie ¹², et c'est à ce niveau du tractus intestinal que *C. difficile* exprime ses facteurs de virulence majeurs avant son éjection hors de l'hôte ^{2,113}. Ses principaux facteurs de virulence sont les deux exotoxines TcdA et TcdB qui ciblent les cellules de l'épithélium intestinal en inactivant par glycosylation les GTPases Rho, Rac et Cdc42, entraînant ainsi une dénaturation du cytosquelette, la mort cellulaire puis la rupture de la barrière intestinale ^{113,125}. Outre leur action cytotoxique directe, elles ont pour conséquence une réponse inflammatoire engendrant des dommages tissulaires et pouvant entraîner des séquelles cliniques majeurs. Il a été montré qu'une augmentation significative de la quantité de toxine retrouvée dans les fèces est associée à une aggravation des symptômes et à la dégradation de l'état général des patients ⁵⁷. Les manifestations cliniques d'une ICD sont très hétérogènes, et peuvent

varier de diarrhées faibles à modérées, jusqu'à des formes inflammatoires graves comme les CPM ou les mégacôlons toxiques, pouvant entraîner des perforation coliques et aboutissant à la mort dans certains cas ^{126,127}. Les manifestations cliniques les plus fréquentes sont les diarrhées aqueuses modérées, accompagnée parfois de crampes abdominales. La diarrhée débute peu de temps après le début de l'antibiothérapie et disparaît généralement avec l'arrêt du traitement ¹²⁸. La colite sans formation de pseudomembranes est une forme clinique d'une ICD également courante mais plus sévère caractérisée par de la fièvre, des malaises, de fortes diarrhées aqueuses légèrement sanglantes, des nausées, une anorexie et de fortes douleurs abdominales ^{128,129}.

2. Les facteurs de colonisation

Avant la production de ses principaux facteurs de virulence, les toxines A et B, *C. difficile* doit d'abord entreprendre une étape de colonisation du colon durant laquelle plusieurs facteurs sont impliqués. Ce phénomène nécessite la régulation de l'expression des gènes contrôlant l'adhérence et la motilité qui constituent un facteur crucial pour l'efficacité de colonisation ; les protéines de surface (*S-Layer*) et la paroi bactérienne sont également impliquées dans les interactions avec l'hôte ^{113,130,131}.

a. Les adhésines

Pour beaucoup de pathogènes, l'adhérence aux cellules de l'hôte est une première étape majeure dans le processus de colonisation et de développement de l'infection. Ce processus requiert des adhésines bactériennes qui sont essentiellement des protéines de surfaces. La couche S ou *S-Layer* est une couche externe présente chez des centaines d'espèces procaryotes, incluant bactéries et archées. Elle est généralement composée de monomères de glycoprotéines identiques régulièrement structurées en réseau cristallin formant la couche la plus externe de la cellule et représente jusqu'à 15% des protéines totales de la cellule ¹³². Elle peut également présenter à sa surface d'autres types de protéines ancrées dans la paroi cellulaire mais reste de très loin le composant majoritaire de la surface de la cellule ¹³³. Ses fonctions varient entre les différentes espèces bactériennes et comprennent la protection contre les phages, l'adhérence et la résistance au complément ¹³². Contrairement à la majorité des souches pour lesquelles la couche S est constituée de

plusieurs monomères identiques, toutes les souches de *C. difficile* possèdent une couche S majoritairement formée de deux sous-unités protéiques de poids moléculaires différents : la sous unité HMW-SLP (*High-Molecular Weight-Surface Layer Protein*) de poids moléculaire élevé (47 kDa) formant la couche inférieure, et la sous unité LMW-SLP (*Low-Molecular Weight-Surface Layer Protein*) de faible poids moléculaire (36 kDa) formant la couche de surface (**Figure 5A**)^{134,135}. Ces deux sous-unités dérivent du clivage du même précurseur polypeptidique SlpA par la protéase à cystéine Cwp84^{136,137} et forment un hétérodimère lié de manière non covalente à la surface de la cellule jouant un rôle dans l'adhésion aux cellules hôtes¹³⁸. Bien que les protéines de la *S-Layer* (SLPs) de *C. difficile* ne semblent pas être glycosylées¹³⁹, un cluster de glycosylation a pu être identifié dans certaines souches¹⁴⁰. Les formes natives et recombinantes de la protéine SlpA sont capables de se fixer *in vitro* aux cellules Hep-2 et aux cellules gastro-intestinales humaines¹⁴¹ et un prétraitement de cellules Caco-2 avec les sous-unités de la SlpA purifiées, ou une incubation de la bactérie avec un anticorps anti-SlpA réduit significativement l'adhérence de *C. difficile* à ces cellules¹⁴². Le gène *slpA* est localisé dans une région du génome de *C. difficile* de 36,6 kb (souche 630) appelé *Cell wall protein (Cwp) cluster* dont les gènes sont très conservés entre les souches de *C. difficile* à l'exception du gène *slpA*^{131,135 143}. Ce *locus* contient 11 paralogues de *slpA* (comme *cwp66* et *cwp84*) et 17 autres sont répartis dans le génome^{135,144}. Tous ces gènes codent pour des protéines appartenant à la famille des Cwp qui sont caractérisées par la présence de trois domaines répétés en tandem permettant leur ancrage à la surface de la bactérie *via* le polysaccharide II¹⁴⁵, et pour certaines d'entre elles d'autres domaines de fonctions spécifiques connues ou inconnues (**Figure 5B**)¹³⁵. Bien que certaines Cwp aient été identifiées, beaucoup d'autres doivent être caractérisées et leur rôle dans la pathogenèse des ICDs doit encore être élucidé.

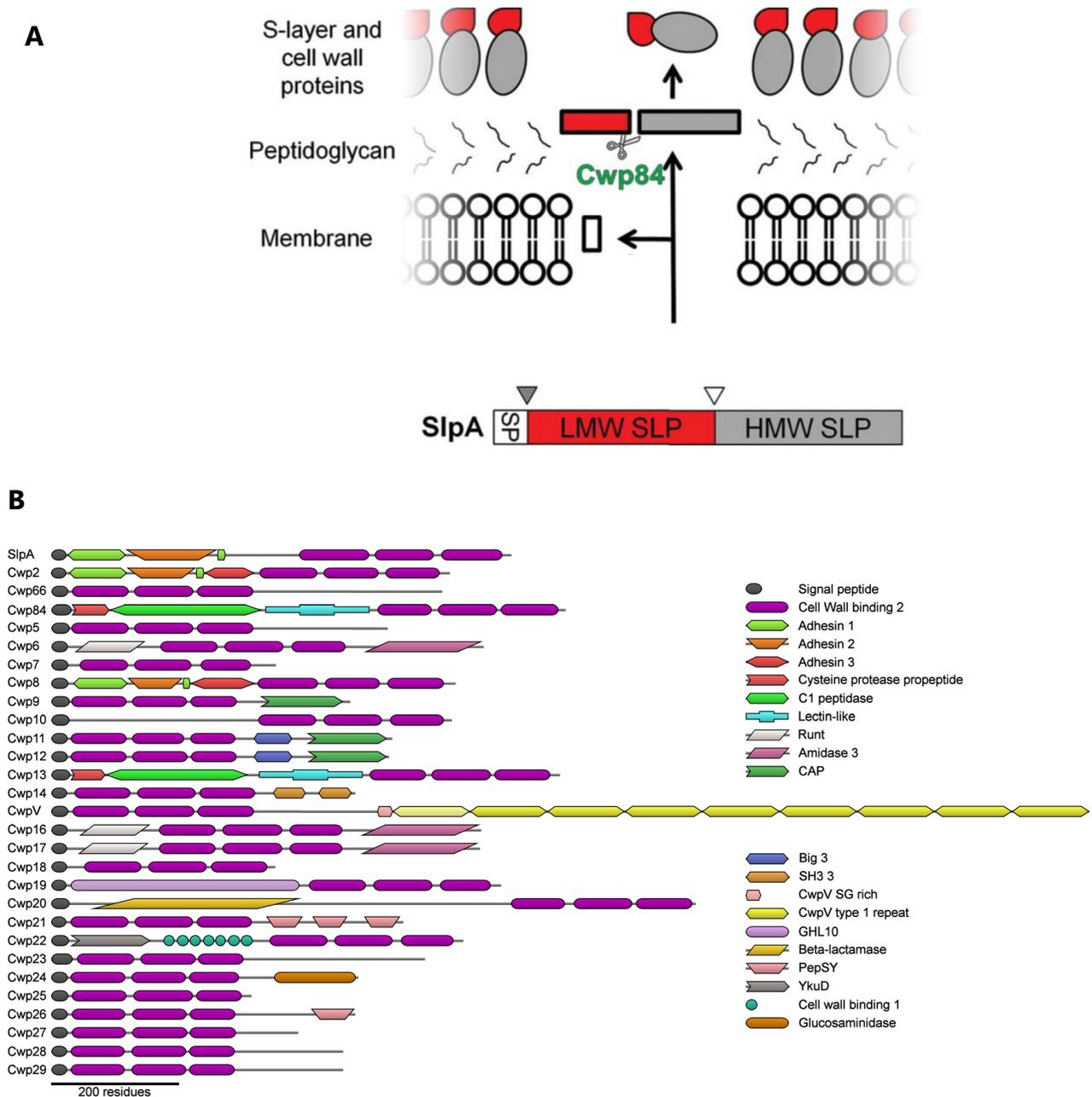


Figure 5 - Protéines Cwp de *C. difficile*. (A) Représentation des étapes de maturation des SLP de *C. difficile*¹³⁰. (B) Représentation des domaines putatifs des 29 protéines Cwp de la souche 630 de *C. difficile*¹³⁵.

Les protéases jouent un rôle important dans les processus d'adhésion. *C. difficile* possède une zinc-métalloprotéase (CD630_28300) appelée Zmp1 qui est capable de cliver plusieurs protéines de l'hôte comme la protéine de choc thermique HSP90 β , les anticorps IgA2, le fibrinogène ou la fibronectine^{146,147}. La protéase Zmp1 est également capable de cliver la protéine CD630_28310 fixant le collagène dont le gène est localisé à proximité de celui de Zmp1. L'expression de ces deux protéines est régulée par le second messager c-di-GMP (*cyclic dimeric guanosine monophosphate*). Dans des conditions de faible concentration de c-di-GMP, la protéase Zmp1 est produite et clive la protéine CD630_28310

réduisant ainsi l'adhérence de la bactérie à la paroi intestinale de l'hôte. A l'inverse, lorsque le taux de c-di-GMP est élevé, la faible production de protéase Zmp1 ne permet plus de cliver l'adhésine augmentant ainsi la capacité de la bactérie à adhérer à son hôte ¹⁴⁷. Une lipoprotéine associée à la surface (CD630_08730) présente également des propriétés adhésives. Les mutants de délétion pour cette protéine présentent une diminution d'adhérence aux cellules Caco-2 et la préincubation de la bactérie sauvage avec des anticorps spécifiques réduit significativement sa capacité à adhérer à ces cellules ¹⁴⁸. Parmi les facteurs d'adhésion, on peut également citer la protéine de choc thermique (Hsp60) GroEL de 58 kDa qui est relarguée dans le milieu extracellulaire après divers stress (choc thermique, choc acide ou osmotique, carence en fer, présence de concentration sub-inhibitrice d'ampicilline) et pourrait permettre à la bactérie d'adhérer à la surface par un mécanisme inconnu ^{131,149}.

b. Protéines de liaison à la matrice extracellulaire

C. difficile, comme beaucoup d'autres bactéries pathogènes, présente à sa surface des protéines appelées MSCRAMMs (*Microbial Surface Components Recognizing Adhesive Matrix Molecules*) qui lui confèrent la capacité d'adhérer à son hôte *via* diverses protéines de la matrice extracellulaire (MEC) comme la fibronectine, la laminine et le collagène ^{131,150}. *C. difficile* possède une métalloprotéine de 68kDa (Fbp68 ou FbpA) capable de se fixer à divers constituants de la MEC incluant la fibronectine, le fibrinogène et la vitronectine, et qui est codée par un gène très conservé entre ses différentes souches ^{151,152}. Même s'il a été montré que des anticorps anti-Fbp68 permettent une inhibition partielle de la fixation de *C. difficile* aux cellules, son rôle dans le développement de la pathogenèse reste encore mal compris ¹⁵². Une autre protéine permettant la fixation aux collagènes de type I et V (Cbpa) a été identifiée chez *C. difficile* ¹⁵³.

c. Pili de type IV

Les *pili* de type IV sont des appendices impliqués dans de nombreux phénotypes chez les bactéries à Gram positif dont l'adhérence, l'agrégation cellulaire et la formation de biofilm. Le génome de *C. difficile* présente un cluster de gènes (CD630_35030 à CD630_35130) codant pour les protéines nécessaires à l'assemblage du *pilus* de type IV et

qui semble être présent chez toutes les souches ^{154,155}. Les *pili* de type IV semblent être exprimés *in vitro* et *in vivo* ¹³¹. La protéine PilA1 (CD630_35130) constitue la composante principale de la piline et son gène présente une variabilité génétique permettant une diversification des épitopes en réponse à la pression du système immunitaire de l'hôte ^{156,157}. Le rôle du *pilus* de type IV dans la pathogenèse de *C. difficile* est cependant toujours hypothétique. L'expression des gènes impliqués dans la formation du *pilus* de type IV est induite en présence de concentrations élevées de c-di-GMP par un *riboswitch* de type II localisé en amont du gène *pilA1* ce qui a pour conséquence d'induire l'auto-agrégation de la bactérie et la formation de biofilm en présence de c-di-GMP ¹⁵⁸⁻¹⁶⁰. De cette manière, le c-di-GMP agit chez *C. difficile* comme un signal clé pour le passage d'un état motile et toxigénique à un état sessile fortement adhérent à la surface et formant un biofilm.

d. Les flagelles

Les flagelles sont d'importants facteurs de colonisation pour bon nombre d'entéropathogènes motiles comme *Escherichia coli*, *Salmonella sp* et *Listeria monocytogenes* et dont l'expression est régulée par des facteurs environnementaux. Chez *C. difficile*, les gènes impliqués dans sa motilité sont exprimés à un niveau plus élevé durant la phase de croissance exponentielle et sont sous le contrôle négatif du facteur sigma de la phase de transition SigH ¹⁶¹. Les gènes contrôlant l'assemblage flagellaire de *C. difficile* sont organisés en trois opérons ^{131,154}. Le *locus* F3 contient les gènes de stade précoce de l'assemblage, dont le facteur sigma flagellaire spécifique FliA (appelé également SigD) qui contrôle l'expression des gènes tardifs du *locus* F1 notamment les gènes codant pour la flagelline FliC ¹⁶² et la protéine FliD ¹⁶³. Le régulon F2 est responsable des modifications post-transcriptionnelles du flagelle qui sont essentielles pour son assemblage fonctionnel et la motilité de *C. difficile* ^{164,165}. L'absence du *locus* F3 dans les souches émergentes de PCR-ribotype 078 suggère que le flagelle et la motilité ne sont pas indispensables à la virulence de *C. difficile*.

L'expression du flagelle est très variable entre les différentes souches de *C. difficile* et a été associée à l'adhérence de *C. difficile* aux cellules de l'épithélium intestinal ^{166,167}. Cependant son rôle dans les processus de colonisation reste encore mal compris et peut

varier en fonction de la souche. Par exemple, dans la souche 630, des mutants de FliC et FliD montrent une capacité accrue à adhérer aux cellules Caco-2^{168,169}, suggérant une absence de corrélation entre flagelles et adhésion ou bien une interférence entre la présence des flagelles et l'adhérence de *C. difficile in vitro*; en revanche dans la souche épidémique 027 (R20291), l'adhérence aux cellules Caco-2 est significativement diminuée pour ces mutants¹⁶⁹. De plus, dans la même souche R20291, le mutant FliC montre une perte de *fitness* de colonisation en comparaison à la souche parentale dans un modèle souris dixénique, suggérant que la présence du flagelle apporterait un avantage dans le *fitness* général de la bactérie¹⁶⁹. Il s'avère en revanche que les souches mutantes non flagellées produisent une plus grande quantité de toxines ce qui entraîne une virulence accrue *in vivo*^{169,170}. Une étude récente a mis en évidence un mécanisme de régulation génétique original de l'expression des toxines et du flagelle faisant intervenir un *switch* d'une séquence inversable localisée en amont de l'opéron F3 qui selon son orientation permet à la bactérie de produire ou non un flagelle et des toxines¹⁷¹. Chez *C. difficile*, l'expression du flagelle est régulée par le second messenger c-di-GMP qui agit comme ligand répresseur en se fixant sur un *riboswitch* de type I en amont de l'opéron F3¹⁷²⁻¹⁷⁴. Un niveau élevé de c-di-GMP intracellulaire réprime donc l'expression du flagelle et la motilité¹⁷⁵, mais réprime également indirectement la synthèse de TcdA et TcdB en réprimant l'expression du gène *sigD* présent au sein de l'opéron F1¹⁷⁶. SigD régule positivement l'expression des toxines *via* le contrôle direct de la transcription de *tcdR*¹⁷⁷.

e. Le rôle des biofilms dans la colonisation

C. difficile est capable de former des biofilms sur des surfaces abiotiques, et la robustesse du biofilm varie entre les différentes souches¹⁷⁸⁻¹⁸¹. Au sein du biofilm, la bactérie est enchevêtrée dans une matrice constituée d'ADN extracellulaire, de polysaccharides et protéines dont certaines Cwp et des toxines (majoritairement la toxine A)^{182,183}. La formation de biofilms chez *C. difficile* est contrôlée par des régulateurs centraux comme Spo0A et LuxS, mais aussi par des composants de surface comme le flagelle et la protéase à cystéine Cwp84¹⁸³⁻¹⁸⁵. Le second messenger c-di-GMP régule positivement la formation de biofilm et l'augmentation du c-di-GMP intracellulaire dans la souche 630 résulte en un biofilm plus

robuste ¹⁷⁴. Bien que la capacité de *C. difficile* à former des biofilms *in vitro* n'est plus à prouver, la formation de biofilms *in vivo* demeure peu claire. Peu d'études testant cette hypothèse ont été réalisées. Dans une étude sur un modèle murin, des agrégats de *C. difficile* ont été observés en association avec des tissus endommagés, ce qui pourrait suggérer la formation de micro-colonies ¹⁸⁶. Une étude plus récente dans un modèle monoxénique murin fait également état d'agrégats de *C. difficile* ¹⁸⁷. *C. difficile* a également été retrouvé dans des communautés bactériennes multi-espèces en association avec la muqueuse du colon ¹⁸⁸. Une étude très récente a mis en évidence la régulation des interactions entre *C. difficile* et les autres espèces bactériennes au sein de ces communautés par le régulateur LuxS ¹⁸⁹. Les biofilms ont également été corrélés à une résistance aux antibiotiques et au stress oxydatif ^{182,184} et favorisent la persistance de *C. difficile* sous forme de spores ¹⁹⁰⁻¹⁹².

3. Les toxines TcdA et TcdB

Bien que les colites associées aux ICDs soient la conséquence de la synergie de plusieurs facteurs de virulence exprimés par *C. difficile* lors de l'infection, les deux toxines TcdA et TcdB sont considérées comme les facteurs de virulence majeurs. Elles ciblent principalement les cellules épithéliales intestinales et sont produites en réponse à une carence nutritionnelle. Après leur endocytose et leur activation dans le cytosol, les cellules épithéliales vont subir une nécrose entraînant la perte de l'intégrité de la barrière intestinale et l'activation d'une réponse inflammatoire rapide.

a. Organisation du locus de pathogénicité

La pathogénicité de *C. difficile* est étroitement liée à la production de ses deux toxines : l'entérotoxine A (TcdA) et la puissante cytotoxine B (TcdB) ¹⁹³. Les gènes codant respectivement pour ces deux toxines (*tcdA* et *tcdB*) font partie d'un locus de 19,6 kb appelé PaLoc (pour *Pathogenicity Locus*), considéré comme indispensable au caractère virulent de *C. difficile* ^{194,195}. Le PaLoc contient 3 autres gènes : *tcdR* et *tcdC* qui codent pour des protéines impliquées dans la régulation transcriptionnelle des gènes des toxines, et *tcdE* dont le produit est une protéine *holin-like* requis pour la sécrétion des toxines A et B dans le milieu extracellulaire ^{196,197} (**Figure 6**). Le PaLoc est remplacé par une région non codante de 115 pb dans les souches non toxigènes ^{194,195,198}. Bien que le PaLoc ne présente pas les

caractéristiques d'un élément génétique mobile, son taux G/C différent du reste du génome¹⁹⁹ et sa capacité à être transféré vers une souche non toxigène²⁰⁰ suggèrent une acquisition par transfert horizontal.

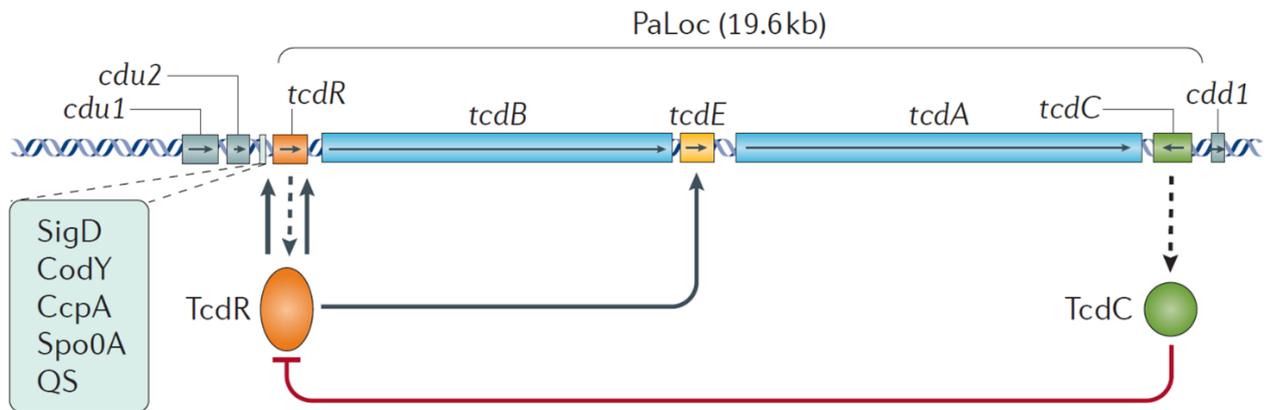


Figure 6 - Régulation de l'expression des toxines TcdA et TcdB. Représentation schématique du *locus* de pathogénicité (PaLoc) et de ses régions adjacentes avec ses éléments régulateurs².

b. Régulation de l'expression des gènes *tcdA* et *tcdB*

La transcription des gènes des toxines A et B est régulée par plusieurs facteurs transcriptionnels. Le premier gène du PaLoc, *tcdR* (anciennement nommé *tcdD* ou *txeR*) code pour une petite protéine de 22 kDa présentant un domaine C-terminal hélice-boucle-hélice de liaison à l'ADN^{195,198}. Il est requis dans l'induction de l'expression des gènes des toxines A et B en tant que facteur sigma spécifique de l'ARN polymérase (ARNP)^{198,201}. *tcdR* présente en amont de sa séquence deux promoteurs TcdR-dépendant signifiant qu'il est capable également d'induire sa propre expression²⁰². A l'autre extrémité du PaLoc et sur le brin négatif, se trouve le gène *tcdC* qui code pour une petite protéine acide de 26 kDa associée à la membrane présentant un domaine transmembranaire prédit en N-terminal^{195,203}. TcdC a la capacité de se dimériser, ce qui est cohérent avec la présence d'un motif en super hélice (ou *coiled coil*) au milieu de sa séquence protéique^{204,205}. *tcdC* présente un profil d'expression inversé à ceux des autres gènes du PaLoc, puisqu'il est transcrit durant la phase exponentielle de croissance tandis que les gènes *tcdA*, *tcdB*, *tcdE* et *tcdR* sont transcrits en entrée de phase stationnaire²⁰⁶. TcdC régule négativement l'expression des toxines A et B de *C. difficile* en empêchant le complexe ARNP/TcdR de reconnaître les promoteurs des gènes *tcdA* et *tcdB*^{53,204}. En fait, les formes libre ou couplée à l'holoenzyme de TcdR sont sensibles à TcdC. En revanche, une fois qu'un complexe ouvert stable est formé avec le

promoteur, TcdC ne peut plus perturber la liaison de l'ARNP à ce même promoteur, suggérant une action très précoce de TcdC au moment de l'initiation de la transcription ²⁰⁴. TcdC agit donc comme un facteur anti-sigma classique sur TcdR en le séquestrant, mais est également capable d'interagir avec le *core-enzyme* de l'ARNP ce qui suggère l'existence de plusieurs mécanismes d'interférence sur la transcription dépendante de TcdR. Cependant, la manière dont TcdC inactive ou déstabilise l'holoenzyme ARNP/TcdR reste toujours en suspens.

La synthèse de TcdR et l'induction subséquente des gènes *tcdA* et *tcdB* est également régulée par divers éléments de l'environnement de *C. difficile*. Un facteur physique comme la température a pour conséquence une diminution de la production de toxines lorsqu'elle s'éloigne de 37°C ²⁰⁷. La quantité et le type de nutriments présents dans l'environnement joue un rôle important pour la régulation de l'expression des toxines. Par exemple, la présence de proline et de cystéine ont pour conséquence une forte diminution du niveau de production des toxines ^{208,209} et la présence d'acides gras à chaîne courte comme l'acide butyrique induit une augmentation de la production de toxines ²¹⁰. La production des toxines en condition de carence nutritionnelle est liée à la répression de l'ensemble des gènes du PaLoc par le régulateur transcriptionnel pléiotrope sensible au GTP, CodY. Ce régulateur CodY est impliqué dans la réponse adaptative à la quantité de nutriments dans le milieu ^{211,212}. L'affinité de CodY à ses gènes cibles est augmentée en présence d'acides aminés à chaîne ramifiée (BCAA ; isoleucine, leucine et valine) et de GTP, séparés ou simultanément, qui signale une disponibilité des nutriments dans le milieu ²¹³. Ainsi, en condition de limitation nutritionnelle, les niveaux de BCAA et de GTP diminuent entraînant une dérégulation des gènes ciblés par CodY afin de pouvoir s'adapter aux conditions de carence. Un autre signal nutritionnel important pour la production des toxines est la présence de glucose et d'autres sources de carbone rapidement métabolisées qui ont pour conséquence une répression de l'expression des toxines médiée par le régulateur pléiotrope CcpA (*Carbon catabolite control protein A*) ²¹⁴⁻²¹⁶.

La production des toxines est également contrôlée par un réseau de régulation complexe impliqué dans la transition vers la phase stationnaire et l'initiation de la

sporulation. Ainsi, SigH et Spo0A, qui régulent chez *C. difficile* les gènes de sporulation mais aussi plusieurs autres gènes impliqués dans diverses fonctions (biosynthèse du butyrate, métabolisme de la paroi cellulaire, motilité ou production des protéines de surface), contrôlent également l'expression des gènes *tcdA* et *tcdB*^{161,217,218}. Cependant, Spo0A réprime l'expression des gènes des toxines A et B seulement dans certaines souches de PCR-ribotype 027 et n'a qu'un effet négligeable dans d'autres souches comme la CD630E ou les souches de PCR-ribotype 078^{217,219}. Le facteur sigma flagellaire SigD joue également un rôle dans la régulation de l'expression des gènes des toxines. Il induit indirectement l'expression des gènes *tcdA* et *tcdB* en se fixant à la séquence promotrice du gène *tcdR* afin d'induire sa transcription¹⁷⁷. Le facteur RstA, impliqué dans la régulation de la sporulation et de la motilité, est également lié indirectement à la répression de l'expression des gènes des toxines en réprimant la transcription de SigD²²⁰. Il a également été montré qu'une forte concentration de c-di-GMP (*cyclic di-guanosyl-51monophosphate*), second messenger bactérien impliqué dans le contrôle du passage entre les modes de vie planctonique et sessile, inhibe l'expression des toxines A et B de manière indirecte en inhibant l'expression du facteur sigma SigD¹⁷⁶. Le *quorum sensing* (QS) est la capacité qu'on des populations bactériennes de coordonner l'expression de leurs gènes en fonction d'un signal moléculaire ou de la densité bactérienne et qui permet de déclencher une réponse synchronisée des bactéries présentes dans cette population²²¹. Ce système implique la production, la détection et la réponse au niveau de molécules signaux extracellulaires appelées autoinducteurs (AI). Il a été montré que le *locus agr* impliqué dans la détection du QS pourrait jouer un rôle important dans la régulation de l'expression des toxines^{222,223}. D'un point de vue global, la régulation de l'expression des toxines de *C. difficile* est étroitement liée à l'état métabolique de la bactérie et à son environnement²²⁴ (**Figure 6**).

c. Mécanisme d'action et structure des toxines A et B

Les symptômes diarrhéiques et inflammatoires caractéristiques des ICDs sont principalement dus aux toxines TcdA et TcdB sécrétées par *C. difficile* au niveau du colon. Initialement, les études sur les toxines purifiées dans les modèles animaux indiquaient que TcdA était le principal facteur de virulence^{225,226}, à tel point que les tests de détection de *C.*

difficile en laboratoire ne permettait que de détecter la toxine A, occultant les ICDs impliquant des souches TcdA-/B+. Cependant, des études ultérieures utilisant des souches mutantes isogéniques pour *tcdA* et *tcdB* sur des modèles animaux ont révélé que la toxine B est suffisante pour engendrer la maladie à elle seule, et ce selon divers mécanismes²²⁷⁻²²⁹. De plus, une autre étude a mis en évidence une atténuation de la virulence d'un mutant *tcdA+/B-* dans différents modèles animaux, ainsi qu'une action locale et systémique de la toxine B seule engendrant de nombreux dommages accompagnés d'une réponse inflammatoire rapide²³⁰. De plus, des études *in vitro* sur des cellules de colon humain ont montré un potentiel inflammatoire important de la toxine B et plus faible pour la toxine A^{231,232}. Cependant, bien que la toxine TcdB soit suffisante pour engendrer la pathologie, les deux toxines jouent un rôle important dans le développement de ICD²³³.

Les toxines TcdA et TcdB présentent environ 66% de similarité de séquence²³⁴ et sont structurellement organisées en quatre domaines fonctionnels (ABCD)^{235,236} : un domaine N-terminal A (*Activity*) responsable de l'activité glycosyltransférase (GTD) de la toxine ; un domaine C (*Cutting*) de protéase à cystéine (CPD) qui permet l'auto-clivage de la toxine et le relargage du domaine catalytique dans le cytosol ; un domaine D (*Delivery*) hydrophobe interne permettant la formation de pores et le relargage du domaine A ; et le domaine B (*Binding*) de fixation aux récepteurs cellulaires localisé en C-terminal et composé de séquences répétées d'oligopeptides combinés (CROPS) (**Figure 7A**). Des études structurales de microscopie électronique et de cristallographie aux rayons X ont révélé une organisation similaire de ces domaines entre les deux toxines^{236,237}. L'action des toxines A et B se divise en quatre phases. Dans un premier temps, les toxines vont se fixer à un récepteur *via* le domaine de fixation pour être ensuite internalisées par endocytose²³⁸. Le domaine CROPS C-terminal de TcdA a longtemps été considéré comme étant le motif permettant la fixation de la toxine sur un glucide (α -Gal-(1,3)- β -Gal-(1,4)- β -GlcNAcO(CH₂)₈CO₂CH₃) associé à la membrane conduisant à une endocytose clathrine dépendante^{237,239}. Les domaines CROPS des deux toxines divergent dans leur séquence et leur taille, avec un domaine CROPS de la toxine B considérablement plus court que celui de la toxine A suggérant donc des récepteurs différents pour les deux toxines¹³¹. TcdA se fixe à une grande variété de glucides mais le

récepteur spécifiquement reconnu par TcdA pour sa fixation aux cellules épithéliales humaine reste inconnu²³⁷. En revanche, deux récepteurs ont pu être identifiés pour la toxine B : le récepteur du poliovirus 3 ou PVLR3 (*poliovirus receptor-like protein 3*) appelé également nectine 3²⁴⁰, et le CSPG4 (*chondroitin sulfate proteoglycan 4*)²⁴¹. Le PVLR3 est fortement exprimé à la surface des cellules humaines épithéliales du colon et colocalise avec TcdB dans des tissus récupérés de patients infectés par *C. difficile*, suggérant une fixation de TcdB à ce récepteur initial lors de l'infection²⁴⁰. Quant à CSPG4, il est fortement exprimé dans les myofibroblastes sub-épithéliaux intestinaux humains et murins, suggérant une fixation de TcdB plus tardive qui ne pourra se faire qu'après la détérioration de l'épithélium du colon^{241,242}. De plus, il a été montré que les toxines TcdA et TcdB dépourvues de leur domaine CROPS sont toujours capables d'infecter les cellules, suggérant donc l'implication d'autres domaines dans la fixation des toxines aux récepteurs^{243,244}. Cette capacité de se fixer à différents récepteurs pourrait également permettre aux toxines d'affecter un répertoire plus large de types cellulaires pouvant expliquer les variations de pathogénèse dans les ICDs¹³¹. Une fois les toxines dans la vésicule d'endocytose, le domaine hydrophobe interne s'insère dans la membrane de la vésicule ce qui permet la libération du domaine catalytique dans le cytosol après son clivage autoprotéolytique *via* le domaine CPD, induit par l'inositol hexakisphosphate présent dans le milieu intracellulaire²³⁷. Une fois dans le cytosol, le domaine catalytique GTD inactive par glycosylation les GTPases des familles Ras et Rho (**Figure 7B**), qui jouent un rôle capital dans de nombreux processus cellulaires comme l'adhésion intercellulaire et le maintien de l'intégrité du cytosquelette. L'inactivation de ces protéines a pour conséquence une altération du cytosquelette qui entraîne un arrondissement de la cellule, une rupture des jonctions serrées et une déstabilisation de la barrière intestinale, ainsi que la mort cellulaire par apoptose ou nécrose. La perte de l'intégrité de la barrière de l'épithélium intestinal a pour conséquence une infiltration et une accumulation de fluides ainsi que de sérieux dommages intestinaux. L'action des toxines a également pour conséquence la sécrétion de cytokines et chimiokines pro-inflammatoires par les cellules épithéliales et les cellules immunitaires résidentes du tissu épithélial, entraînant un influx de neutrophiles et la destruction du tissu intestinal^{235,245} (**Figure 8**).

A



B

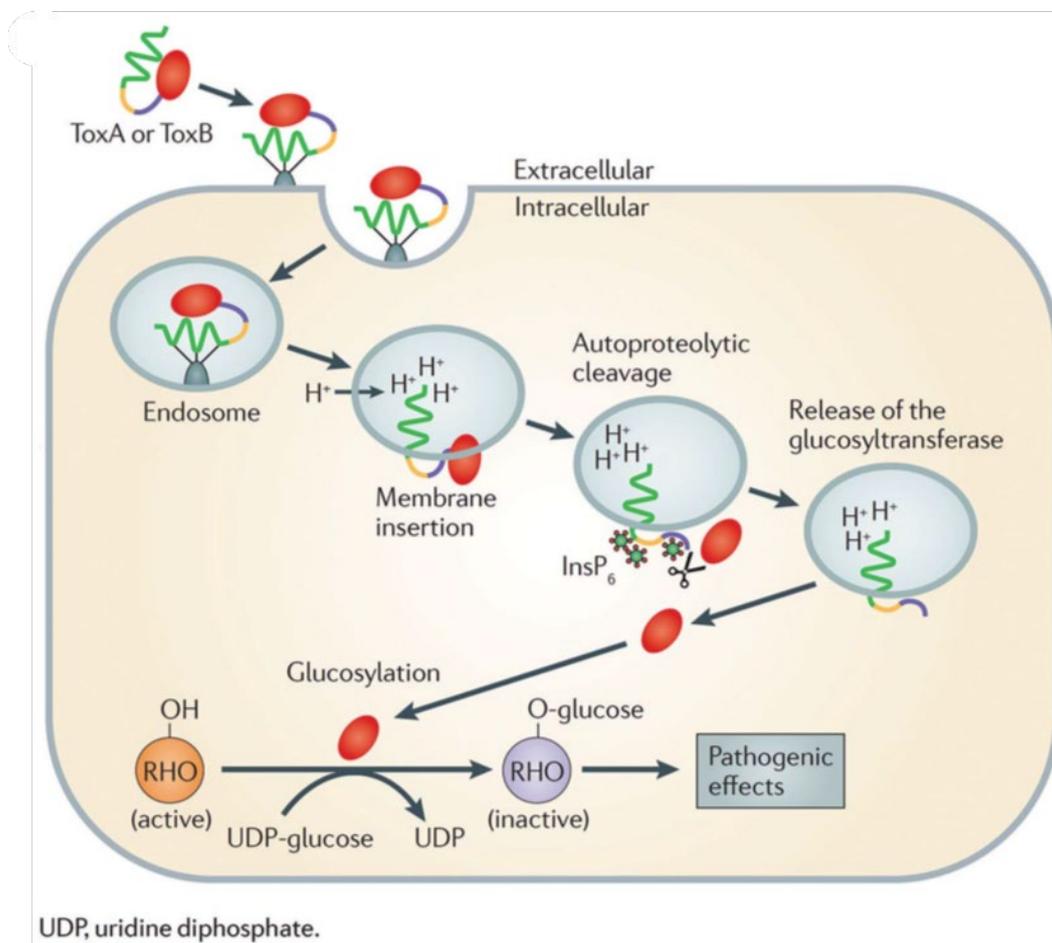


Figure 7 – Structure primaire et mécanisme d’action des toxines TcdA et TcdB. (A) Représentation des 4 domaines fonctionnels des toxines homologues TcdA et TcdB²³⁷. (B) Mécanisme d’action de TcdA et TcdB. Le code couleur des domaines fonctionnels est le même entre les figures (A) et (B)¹³⁰.

4. La toxine binaire (CDT) : expression, mode d'action et importance

La toxine binaire CDT (ou *C. difficile* transférase) est une troisième toxine produite par un nombre limité de souches de *C. difficile*, dont les souches épidémiques de PCR-ribotype 027 et 078²⁴⁶. Elle est codée par deux gènes, *cdtA* et *cdtB*, qui forment un opéron localisé dans une région appelée *locus* de la toxine binaire ou CdtLoc (**Figure 8A**)^{247,248}. Les souches qui ne produisent pas de CDT présentent une délétion d'environ 2kb dans leur génome²⁴⁹. Le CdtLoc contient également un gène localisé en amont de l'opéron *cdtAB* codant pour le régulateur transcriptionnel positif de ce même opéron : CdtR (**Figure 8A**)²⁵⁰. Cependant, une délétion partielle de la séquence de CdtR ne supprime pas l'expression de la toxine CDT, suggérant donc que CdtR n'est pas indispensable à son expression²⁵¹. Contrairement au PaLoc, les voies de signalisation régulant l'expression des gènes du CdtLoc ne sont pas connues. La toxine CDT appartient à la famille des toxines binaires à ADP-ribosylation et est constituée de deux unités : une unité portant l'activité ADP-ribosyltransférase de la toxine qui permet la destruction du cytosquelette d'actine (CDTa), et une seconde unité (CDTb) impliquée dans la fixation de la toxine à la cellule cible et qui est chargée de réaliser la translocation de l'unité enzymatique dans le cytosol (**Figure 8B**)²⁴⁶. En dépit d'une compréhension approfondie des mécanismes d'action de la CDT sur les cellules infectées, le rôle de cette toxine dans la pathogenèse des ICD demeure incertain. Des données expérimentales ont suggéré que l'action de la CDT a pour conséquence la formation de protrusion micro-tubulaires sur les cellules épithéliales qui pourraient accroître l'adhérence et la colonisation de *C. difficile*²⁵². Cependant, la présence de cette toxine dans des souches cliniques fréquemment associées avec des formes sévères d'ICD, comme les souches des de PCR-ribotypes 027 et 078, ainsi que l'isolement de quelques souches cliniques TcdA-TcdB-CDT+ suggèrent bien que cette toxine joue un rôle dans le développement de la maladie²⁵³⁻²⁵⁵. La récente identification d'une nouvelle souche associée à des formes sévères d'ICD en Chine qui n'est ni RT 027 ni 078 mais qui est capable de produire la toxine binaire renforce également les hypothèses concernant son rôle dans la virulence²⁵⁶.

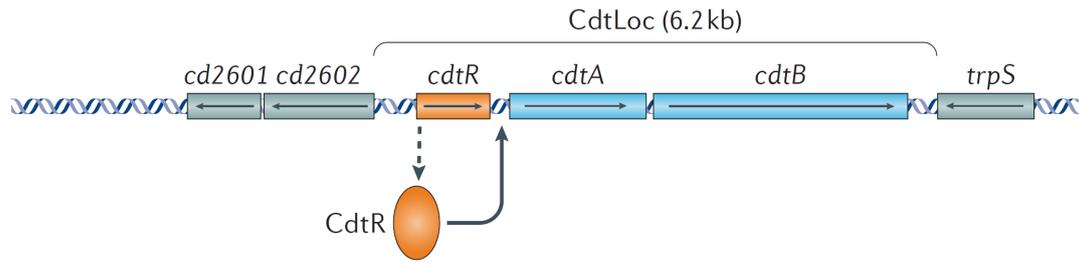
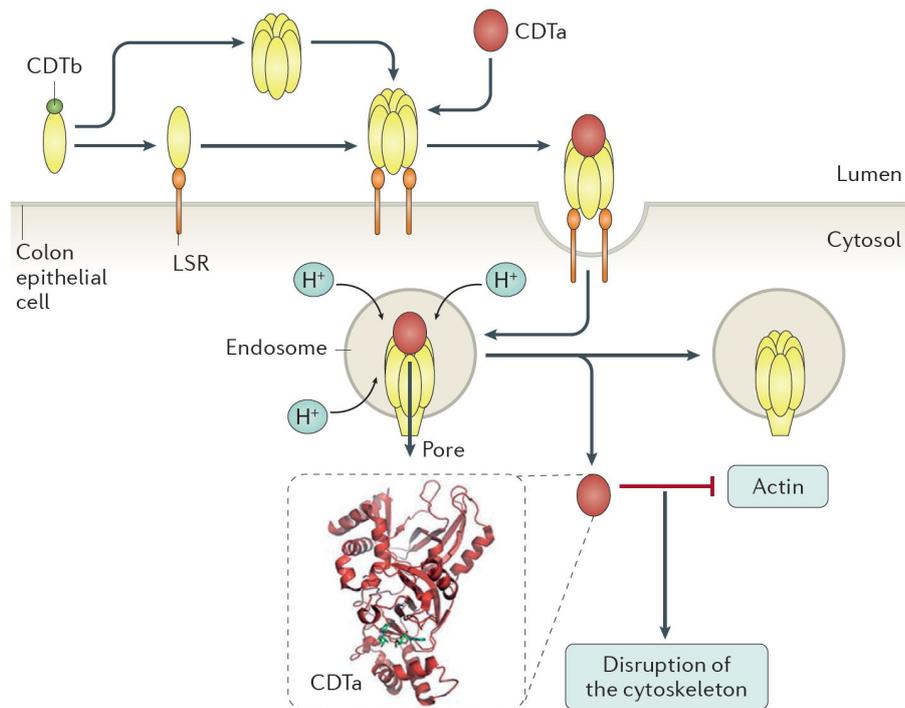
A**B**

Figure 8 – Organisation du CdtLoc et mécanisme d'action de la toxine binaire. (A) Représentation du *locus* de la toxine binaire CDT (CdtLoc) et de ses régions adjacentes. (B) Mécanisme d'action de la toxine binaire.²

5. Sporulation et germination

La lumière du côlon humain est un environnement anoxique permettant aux bactéries anaérobies strictes telles que *C. difficile* de survivre et, si les conditions s'y prêtent, de proliférer, de produire des toxines et d'endommager l'épithélium intestinal. Durant le cycle infectieux, *C. difficile* a la capacité de former des spores lui permettant de subsister à l'extérieur de l'hôte et de résister à de nombreux stress environnementaux comme la présence de l'oxygène, la chaleur, le pH acide, et les traitements antimicrobiens favorisant ainsi sa persistance et sa propagation chez les patients hospitalisés²⁵⁷⁻²⁵⁹. La structure globale de la spore chez *C. difficile* est semblable à celle des spores des autres bactéries sporulantes et est organisée en plusieurs couches qui contribuent à ses propriétés résistantes particulières.

Dans la partie la plus interne de la spore se trouve le noyau qui contient le chromosome bactérien, de l'ARN et des complexes protéiques. Il est recouvert par une membrane interne très peu perméable aux petites molécules (dont l'eau) qui protège ainsi l'ADN des molécules pouvant le dégrader. Cette membrane interne est recouverte par une paroi qui deviendra la paroi de la cellule bactérienne après germination de la spore. La paroi est entourée par une épaisse couche de peptidoglycane (PG) appelée cortex qui est lui-même recouvert par une membrane externe dérivant de la cellule mère, essentielle à la formation de la spore mais ne qui ne confère aucune propriété de résistance. Un manteau protéique (appelé également tunique) entourant la membrane externe constitue la couche la plus externe de la spore. Chez *C. difficile* comme chez d'autres espèces, une couche protéique supplémentaire, appelée exosporium, est présente (**Figure 9A**)¹⁰². Les signaux qui permettent à *C. difficile* de déclencher le processus de sporulation *in vivo* n'ont pas encore été identifiés, mais il s'agirait très probablement, comme pour la plupart des bactéries sporulantes, de *stimuli* environnementaux tels que la carence nutritionnelle ou le quorum sensing¹⁰². Cependant, les réseaux de régulation moléculaire contrôlant la formation de spores sont aujourd'hui bien décrits. L'initiation du processus de sporulation chez *C. difficile* se fait *via* des histidine kinases (CD1352, CD1492, CD1579, CD1949 et CD2492 chez la souche 630) entraînant l'activation par phosphorylation du facteur de transcription Spo0A^{260,261}. Lorsqu'il est activé, Spo0A va contrôler le réseau de régulation de la sporulation en régulant directement une douzaine de gènes^{218,219}. Spo0A est conservé parmi les organismes sporulant et sa présence est souvent utilisée comme un indicateur de sporulation (ou de capacité à sporuler)^{262,263}. Lorsqu'il est inactivé chez *C. difficile*, les cellules deviennent incapables de produire des spores. Dans un modèle souris, l'incapacité à sporuler diminue largement la capacité de la bactérie à persister chez l'hôte et à se propager entre différents individus, suggérant donc un rôle clé de Spo0A dans la persistance, les récurrences et la transmission des ICDs²⁶⁴. Spo0A initie le processus de division asymétrique engendrant la formation de deux compartiments de tailles inégales contenant chacun une copie identique de chromosomes mais exprimant des gènes différents : la préspore et la cellule mère dans laquelle la future spore est assemblée avant sa libération dans le milieu extérieur (**Figure 9B**)

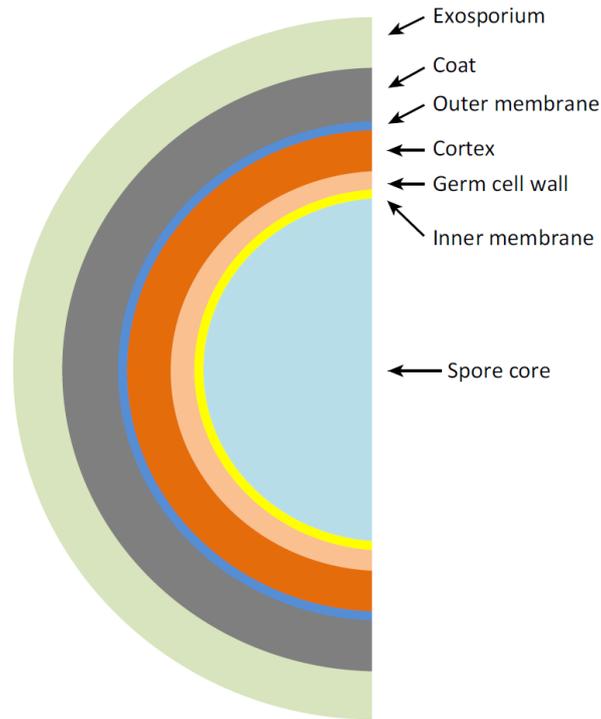
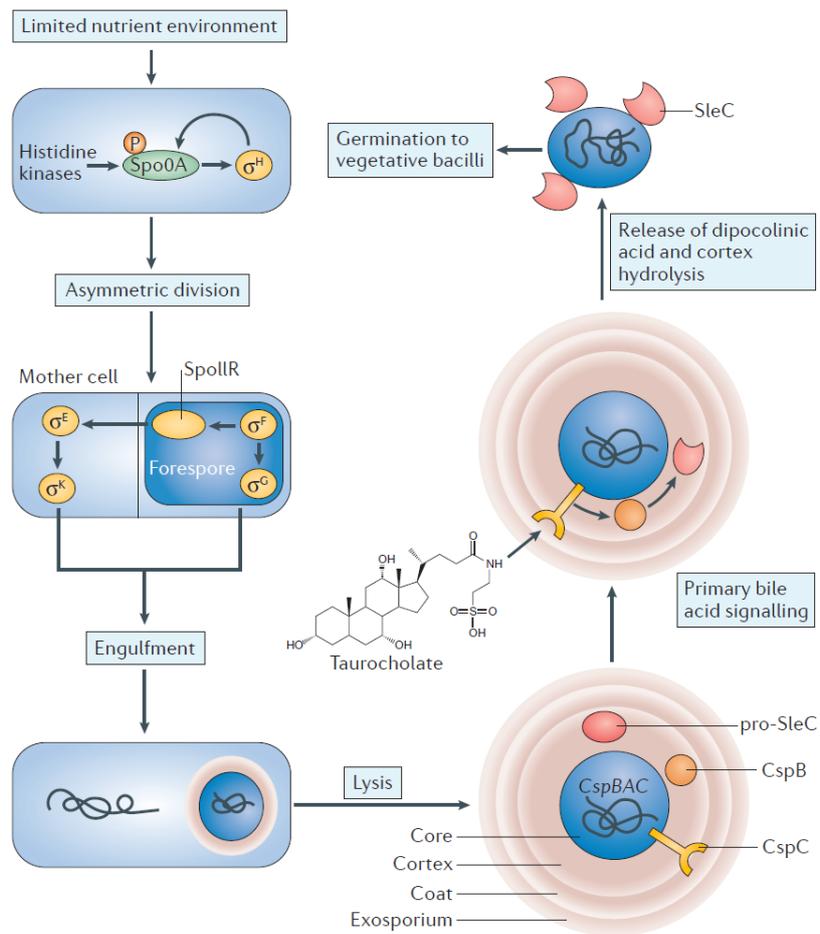
A**B**

Figure 9 - Sporulation et germination chez *C. difficile*. (A) Structure de la spore de *C. difficile*¹⁰². (B) Description des processus de sporulation et de germination¹⁶⁶.

Dans un premier temps, l'expression de Spo0A permettant l'initiation du processus de sporulation est rétro-contrôlée positivement par le facteur sigma σ^H (SigH) qui exerce également un contrôle positif sur l'expression de plusieurs autres gènes de la sporulation¹⁶¹. L'expression de Spo0A est également contrôlée par les facteurs de transcription CodY et CcpA en fonction de l'état nutritionnel. De plus, la délétion de deux perméases à oligopeptides (Opp et App), dont la fonction première est l'importation de petits peptides non spécifiques (de 3 à 20 acides aminés) comme sources de nutriment, induit l'augmentation de l'expression de Spo0A et la production de spores renforçant donc l'hypothèse de corrélation entre carence nutritionnelle et sporulation²⁶⁰. Après la division asymétrique, l'expression des gènes dans la pré-spore et la cellule mère est activée par quatre facteurs sigma de la sporulation agissant de manière compartiment-spécifique : σ^F , σ^E , σ^G et σ^K ^{265,266}. Bien que ces quatre facteurs de sporulation soient également présents chez *Bacillus subtilis*, leurs mécanismes d'action chez *C. difficile* sont différents^{267,268}. Chez *C. difficile* et *B. subtilis*, le facteur σ^F n'est activé que dans la préspore, entraînant la production de la protéine de signalisation SpoIIR et subséquemment celle du facteur σ^E de la cellule mère^{266,267}. La délétion individuelle de chacun de ces facteurs sigma a mis en évidence l'induction de l'activation de σ^G par σ^F dans la préspore, et celle de σ^K par σ^E dans la cellule mère. Cependant, contrairement à ce qui est observé chez *B. subtilis*, l'activité de σ^G n'est pas dépendante de σ^E et σ^K , ce qui souligne l'absence de conservation de la signalisation inter-compartimentale entre *C. difficile* et *B. subtilis*²⁶⁸.

La germination des spores de *C. difficile* et la croissance des formes végétatives a lieu dans la partie basse du tractus gastrointestinal où les concentrations d'oxygène sont négligeables²⁶⁹. Certaines substances présents dans l'intestin, notamment les acides biliaires (taurocholate, glycocholate, cholate, et déoxycholate), permettent l'induction de la germination¹⁰². Il est important de noter que les acides biliaires n'ont pas tous la même capacité à induire la germination et certains d'entre eux peuvent avoir des conséquences sur la croissance végétative²⁷⁰. Par exemple, le taurocholate en combinaison avec la glycine induit la germination des spores sans affecter la croissance végétative, quant au contraire le sel biliaire secondaire déoxycholate induit la germination mais réduit fortement la croissance

végétative ²⁷¹. En condition normale, le microbiote colique convertit les acides biliaires primaires en acides biliaires secondaires et autres dérivés qui inhibent la croissance végétative de *C. difficile*. Une dysbiose induite par traitement antibiotique supprime les organismes responsables de cette conversion ce qui permet à *C. difficile* de germer et se multiplier (**Figure 10**) ¹⁶⁶. La germination de la spore passe par différentes étapes ^{272,273} (dégradation du cortex, libération de calcium et d'acide dipicolinique, réhydratation de la spore) qui sont sous le contrôle des gènes du *locus cspBAC* ¹⁰². Parmi les gènes de ce *locus* se trouve le gène codant pour le récepteur aux acides biliaires CspC. En présence de taurocholate, le récepteur CspC va activer la sérine protéase CspB qui est requise pour l'activation protéolytique du propeptide pro-SleC dont la forme mature SleC permet la dégradation du cortex de la spore durant la germination (**Figure 9B**) ²⁷⁴⁻²⁷⁶. La germination est un processus complexe faisant intervenir plusieurs centaines de gènes ¹⁶⁶ et dont l'efficacité varie entre les différents isolats cliniques de *C. difficile*. Des études font état d'un lien de corrélation entre la sévérité des ICD et l'efficacité de germination de certaines souches ^{277,278}.

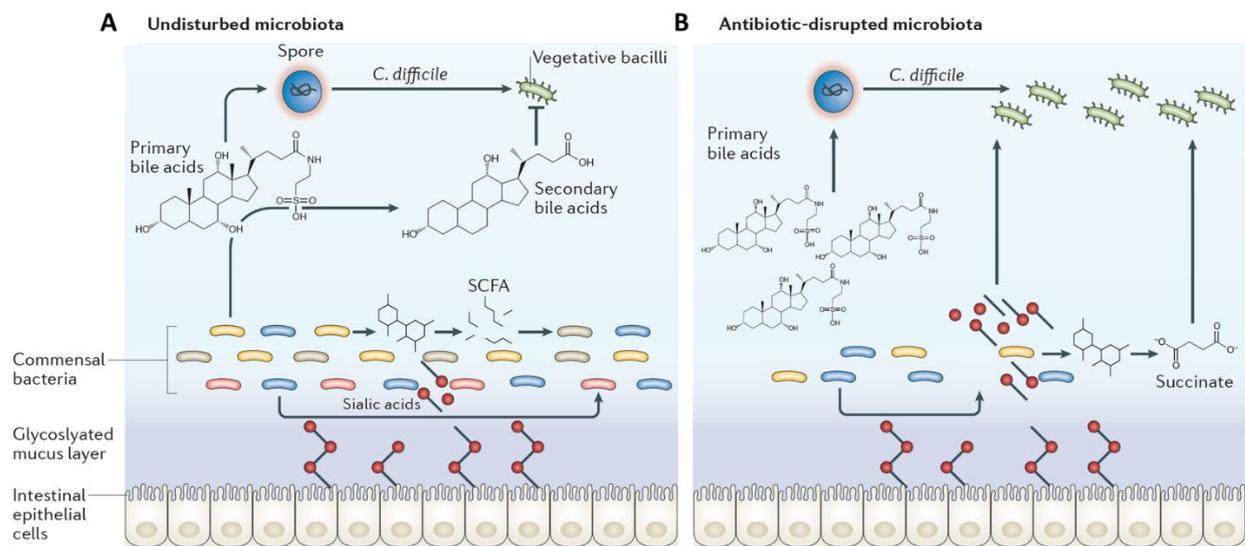


Figure 10 - Mécanismes de défense liés au microbiote vis-à-vis de *C. difficile*. (A) Lorsqu'il est intact, le microbiote convertit les acides biliaires primaires en acides biliaires secondaires ce qui a pour conséquence d'inhiber la croissance des cellules végétatives de *C. difficile*. Les bactéries du microbiote peuvent également convertir les sucres attachés à la surface des cellules épithéliales en acide sialique, et les glucides en acides gras à chaîne courte (SCFA) comme le succinate afin de les utiliser comme source de nutriment. (B) Une dysbiose induite par un traitement antibiotique supprime la conversion des acides biliaires primaires en acides biliaires secondaires, ce qui permet aux cellules végétatives de *C. difficile* de se diviser. Les antibiotiques éliminent également les bactéries entrant en compétition avec *C. difficile* pour l'assimilation d'acide sialique et de succinate. ¹⁶⁶

6. Interaction hôte-pathogène dans les ICDs

a. La réponse immunitaire innée

L'implantation, la multiplication et la production des facteurs de virulence de *C. difficile* dans le tractus intestinal va induire une réponse rapide du système immunitaire inné de l'hôte avec notamment l'induction de cytokine pro-inflammatoire et le recrutement de polynucléaires neutrophiles (**Figure 11**).

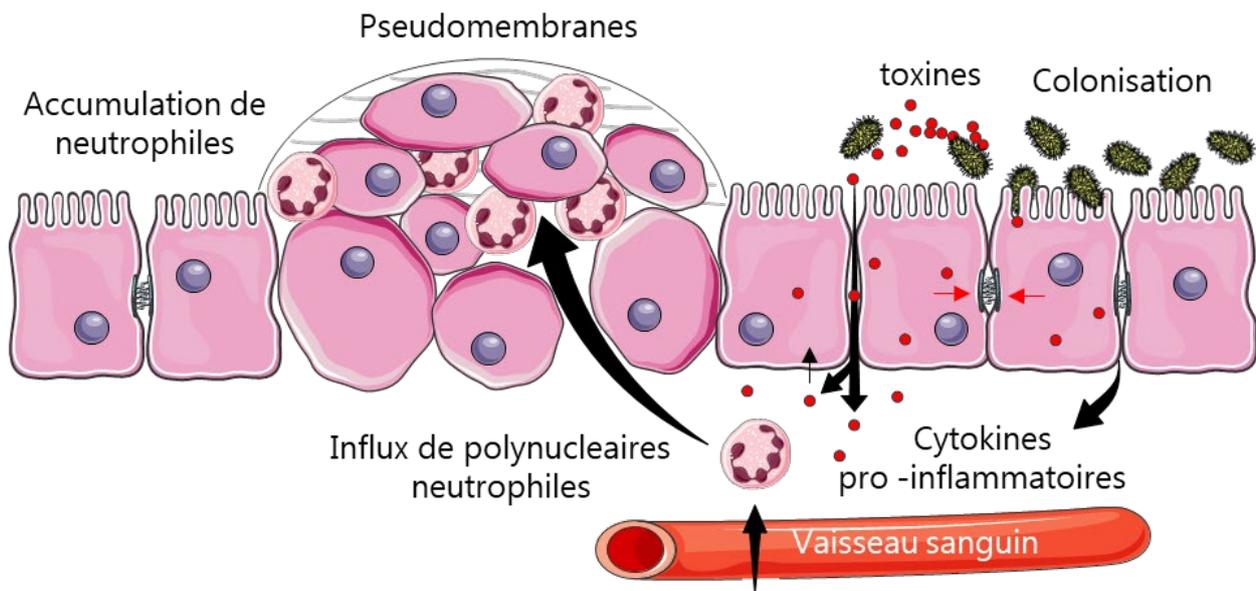


Figure 11 – Réponse de l'hôte lors de l'infection à *C. difficile*.

Les mécanismes de défense innée contre *C. difficile* incluent le microbiote intestinal, la barrière de la muqueuse intestinale, les cellules de l'épithélium intestinal et les cellules immunitaires intestinales (macrophages, monocytes, mastocytes, cellules lymphoïdes innées et cellules dendritiques) ²⁷⁹. Les ICDs sont caractérisées par une réponse inflammatoire intestinale aiguë accompagnée d'une infiltration importante de neutrophiles entraînant des dommages aux tissus intestinaux. Il a été montré dans une étude en modèle souris que l'inhibition de la réponse inflammatoire aiguë pourrait protéger des dommages intestinaux causés par une ICD ²⁸⁰. L'activation de cette réponse immunitaire rapide se fait par le biais de récepteurs de reconnaissance appelés PRRs (*Pattern Recognition receptors*), comme les TLRs (*Toll Like Receptors*) et les NLRs (*Nod Like Receptors*), qui reconnaissent des signatures moléculaires spécifiques aux microorganismes appelés PAMPs (*Pathogen Associated Molecular Patterns*), parmi lesquelles se trouve les composants de la paroi bactérienne (lipopolysaccharide, peptidoglycane) et des flagelles (flagelline), mais aussi des acides

nucléiques bactériens et des toxines bactériennes. Les interactions entre les PRRs et les PAMPs entraîne l'activation rapide de voies de signalisation aboutissant à la production de médiateurs pro-inflammatoires (comme les cytokines et les chimiokines) afin de lutter contre le pathogène, mais qui peut s'avérer délétère dans le cas d'une réponse excessive.

L'action des toxines clostridiales A et B sur les cellules épithéliales du colon induit une sécrétion de fluide, augmente la production de ROI (*Reactive Oxygen Intermediate*)²⁸¹ et diminue la production de mucine²⁸². Les toxines A et B interagissent avec les récepteurs TLR4, TLR5 et NOD1 (*Nucleotide Binding Oligomerization Domain 1*)²⁸³⁻²⁸⁵ ce qui a pour conséquence d'induire la production de nombreuses cytokines pro-inflammatoires et de chimiokines (incluant l'IL-12, l'IL-18, l'IFN- γ , l'IL-1 β , le TNF- α , le MIP-1 α , le MIP-2, l'IL-8 et la leptine²⁸⁶⁻²⁸⁹) responsables de l'inflammation, des dommages intestinaux ainsi que des caractéristiques histopathologiques des ICDs comme une infiltration locale ou systémique de neutrophiles²⁹⁰. Les deux toxines activent également l'inflammasome dépendant de l'IL- β ²⁹¹. L'inflammasome est un complexe multiprotéique constitué de protéines senseurs dont la formation est induite par des *stimuli* spécifiques et conduit à la formation de caspases actives qui clivent la pro-IL-1 β et la pro-IL-18 sous leurs formes bioactives. Il a été récemment montré que la toxine binaire CDT est capable d'augmenter la sévérité de la maladie en agissant de concert avec les toxines A et B en augmentant la réponse inflammatoire de l'hôte *via* une voie dépendante du TLR2 qui est requis pour la suppression de la réponse protectrice éosinophile durant l'infection²⁹².

Les protéines de la *S-Layer*, ou SLPs (*Surface Layer Proteins*), de *C. difficile* sont des antigènes immunodominants et les anticorps anti-SLPs sont présents en grande quantité dans le sérum des patients infectés¹¹³. Des études *ex vivo* ont mis en évidence l'induction de la production de cytokines inflammatoires (IL-1 β , IL-6 et IL-10) par les SLPs de *C. difficile*, ainsi que la maturation et l'activation des cellules dendritiques^{283,293}. Il a été montré que les SLPs de *C. difficile* sont reconnues par le TLR4 entraînant l'activation de voies de signalisation du facteur NF- κ B (*nuclear factor-kappa B*) mais pas de l'IRF3 (*Interferon Regulatory Factor 3*), indiquant une voie de signalisation dépendante de MyD88 (*Myeloid Differentiation Primary Response 88*)²⁸³. La flagelline, unité structurale du flagelle, est également un

puissant activateur de la réponse immunitaire innée de l'hôte. Les récepteurs TLR5 présent à la surface des cellules épithéliales intestinales et des cellules dendritiques reconnaissent les monomères de flagelline qui peuvent être sécrétés lors de l'assemblage flagellaire, ou provenant de filaments flagellaires désagrégés. L'interaction de la flagelline avec le TLR5 entraîne l'activation de NF- κ B et la production de cytokines pro-inflammatoires ²⁹⁴.

b. Résistance de C. difficile aux défenses de l'hôte

Durant son cycle infectieux, *C. difficile* doit faire face à une grande variété de défenses de l'hôte incluant la production de composés antimicrobiens par les tissus intestinaux, comme le lysozyme qui est abondamment sécrété par les cellules de l'épithélium intestinal. Le lysozyme a une activité peptidoglycane hydrolase en clivant les liaisons $\beta(1-4)$ entre un MurNac (acide N-Acétilmuramique) et un GlcNac (N-Acétylglucosamine). De nombreuses bactéries pathogènes ont acquis divers mécanismes de résistance au lysozyme, comme la N-désacétylation du GlcNac du peptidoglycane ²⁹⁵. *C. difficile* présente une résistance naturelle élevée *in vitro* au lysozyme liée à un fort taux de désacétylation de son peptidoglycane ²⁹⁶. Il a été montré qu'un niveau encore plus élevé de désacétylation du résidu GlcNac peut être induite par des concentrations sub-inhibitrices de lysozyme augmentant ainsi le niveau de résistance au lysozyme ²⁹⁷.

Les peptides cationiques antimicrobiens (CAMPs, incluant les cathélicidines et les défensines comme les -défensines 1 et 5) sont des composés dont la charge globale positive facilite les interactions avec la membrane des bactéries chargée globalement négativement. Différentes stratégies d'évasion aux CAMPs utilisées par les bactéries pathogènes ont été identifiés, comme une estérification de la D-alanine des acides téichoïques de la paroi médiée par l'opéron *dlt*, qui a pour conséquence une augmentation de la charge cationique globale de la surface bactérienne réduisant ainsi son affinité avec les CAMPs ²⁹⁸. Il existe également d'autres mécanismes de défenses adoptés par *C. difficile* pour lutter contre les CAMPs impliquant des transporteurs ABC permettant l'efflux des CAMPs ²⁹⁸. Il est intéressant de noter que la souche épidémique de RT 027 présente une résistance naturelle élevée à la cathélicidine LL-37 mais est plus sensible aux défensines que les autres souches ^{299,300}. Enfin, il a été montré que les cathélicidines et les défensines humaines réduisent significativement

les dommages et l'inflammation causés par les toxines A et B, suggérant donc que la résistance aux CAMPs aurait un rôle dans les étapes de colonisation et de production de toxines lors des ICDs ^{301,302}.

III. *C. difficile* dans l'univers des bactériophages

A cours de son cycle infectieux chez l'hôte *C. difficile* doit survivre dans les environnements de l'intestin riche en bactériophages. Les bactériophages (appelés aussi phages) sont les organismes les plus importants de la biosphère et sont retrouvés depuis les fonds des océans jusqu'aux intestins de mammifères ³⁰³. Ce sont des virus capables d'infecter les bactéries et leur impact sur les populations bactériennes varie en fonction de leur stratégie de réplication. Les phages sont d'important vecteurs évolutifs qui, avec d'autres éléments génétiques mobiles (EGM), participent au transfert horizontal de gènes qui contribue à l'acquisition de nouveaux traits adaptatifs importants pour la survie, la virulence et la résistance aux antibiotiques des bactéries. Bactéries et phages ont dû mettre au point différentes stratégies au cours de l'évolution afin de pouvoir co-exister dans de nombreux environnements. Dans cette course à l'armement permanente, les bactéries ont dû développer des mécanismes de défenses innés et adaptatifs efficaces permettant de lutter contre la pression phagique, et les phages de leur côté ont dû mettre au point des stratégies leur permettant d'échapper aux défenses immunitaires bactériennes durant leur cycle de développement.

1. Taxonomie, structure et cycle infectieux des phages

Le génome viral est constitué d'une molécule d'acides nucléiques (ADN ou ARN) bicaténaire (double brin, db) ou monocaténaire (simple brin, sb), segmentée ou non, linéaire ou circulaire. Pour la grande majorité des phages connus, la nature de leur génome est de l'ADN double brin. La taille du génome est directement liée à la complexité morphologique du phage et peut varier du plus simple (par exemple un génome ARN sb de $\approx 3,5$ kb pour le phage MS2) au plus complexe (par exemple un génome ADN db de ≈ 500 kb pour le phage G) ³⁰⁴. Une nucléocapside protéique entoure le génome afin de le protéger et peut être dans certains cas recouverte d'une enveloppe lipoprotéique ³⁰⁵. Les phages présentent une grande diversité de structure morphologique et peuvent présenter une queue (phages

caudés), une symétrie polyédrique, une forme filamenteuse ou pléomorphe. Ces variations de forme, de taille et de symétrie dans leur morphologie permettent d'établir une classification qui évolue constamment et doit être régulièrement mise à jour. Ces modifications sont publiées sur le site de l'ICTV (*International Committee on Taxonomy of Viruses*) qui classifie aujourd'hui plus de 6 000 phages et permet de suivre l'évolution de leur taxonomie en temps réel^{306,307}. La grande majorité des phages caractérisés appartiennent à l'ordre des *Caudovirales* (ADN double brin à morphologie caudée) qui peut être divisé en trois familles : les *Myoviridae* (par exemple le phage T4), les *Siphoviridae* (par exemple le phage λ) and les *Podoviridae* (par exemple le phage T7). Les phages peuvent infecter leur hôte par un cycle lytique ou lysogénique³⁰⁴ (**Figure 12**).

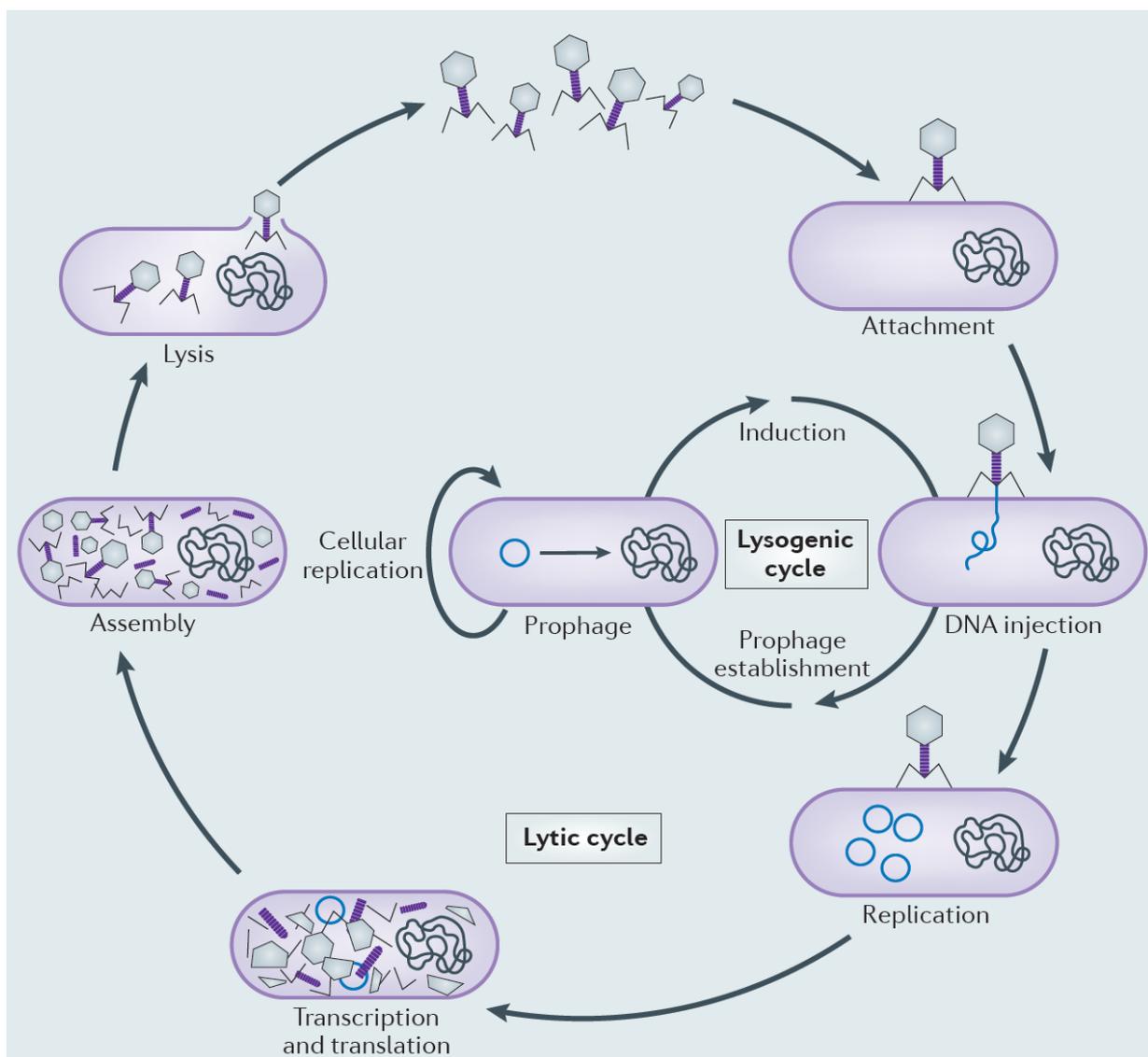


Figure 12 - Cycle de développement des bactériophages³⁰⁴.

Dans un premier temps, le phage doit reconnaître et se fixer spécifiquement à un récepteur présent à la surface de la bactérie à infecter pour ensuite pouvoir injecter son matériel génétique à l'intérieur de la cellule. Les mécanismes d'injection du génome viral sont de mieux en mieux compris et peuvent varier en fonction de leur classe³⁰⁸. La stratégie de réplication qui s'en suit va dépendre de la nature virulente ou tempérée du phage. Les phages virulents (comme le phage T4) ne sont capables de se répliquer que par un cycle lytique qui a comme seule finalité la production et la libération de nouvelles particules virales par la cellule infectée. Les phages tempérés quant à eux (comme le phage λ) peuvent soit entrer dans un cycle lytique, soit former une association stable avec l'hôte appelée lysogénie. Durant la lysogénie, le génome viral prend le nom de prophage et se réplique de concert avec l'ADN de la cellule hôte, soit en tant qu'épisome (comme pour le phage P1) ou intégré dans le chromosome bactérien (comme pour le phage λ). Lors de conditions de stress, le prophage peut sortir de son état lysogénique pour produire des virions qui seront relargués par la bactérie, ce qui entrainera classiquement la lyse de la bactérie³⁰⁹ ; dans le cas des phages filamenteux, les virions sont sécrétés par la bactérie ce qui se traduit par un impact sur la croissance de cette dernière³¹⁰.

2. Les mécanismes de défense contre les phages

Les bactériophages sont présents dans tous les environnements, avec une estimation de 10^{30} particules virales sur Terre et environ 10^{25} infections virales survenant par seconde^{303,311}. A peu près 10 fois plus abondant, ils imposent aux bactéries avec lesquelles ils coexistent une très forte pression de sélection résultant, la plupart du temps, en une lyse cellulaire et un effondrement de la population bactérienne³¹². Dans le modèle de la Reine Rouge de Leigh Van Vallen³¹³ il est postulé que toute adaptation bénéfique d'un organisme entraîne une baisse de *fitness* de toutes les espèces avec lesquelles il interagit, créant ainsi un cycle naturel co-évolutif d'adaptation et contre-adaptation des espèces interagissant dans un même environnement. Ce modèle s'applique parfaitement aux interactions entre phages et bactéries qui ont chacun dû développer des mécanismes de défenses assurant leur survie^{314,315}. Les mécanismes des défenses contre les phages peuvent agir à tous les niveaux du cycle infectieux phagique et on peut grossièrement distinguer deux grandes

catégories de mode d'action : ceux ciblant les étapes précoces du cycle infectieux (adsorption, injection), et ceux intervenant après l'injection du génome phagique dans la cellule (réplication, transcription, traduction des facteurs viraux et assemblage des particules précédant la lyse)³¹¹.

L'adsorption des phages à la cellule hôte est l'étape initiale du cycle infectieux phagique et ne peut avoir lieu qu'après la reconnaissance d'un récepteur spécifique présent à la surface de la bactérie, tel que le LPS (lipopolysaccharide), des protéines membranaires (comme les porines), les pili ou les flagelles. Il existe plusieurs stratégies connues permettant aux bactéries de bloquer l'adsorption des phages, et celles-ci peuvent être divisées en au moins trois catégories : la modification des récepteurs phagiques, la production d'une matrice extracellulaire créant une barrière physique entre les phages et leur récepteurs, et la production d'un inhibiteur compétitif spécifique aux récepteurs phagiques rendant impossible la fixation pour les phages (**Figure 13**)³¹⁴.

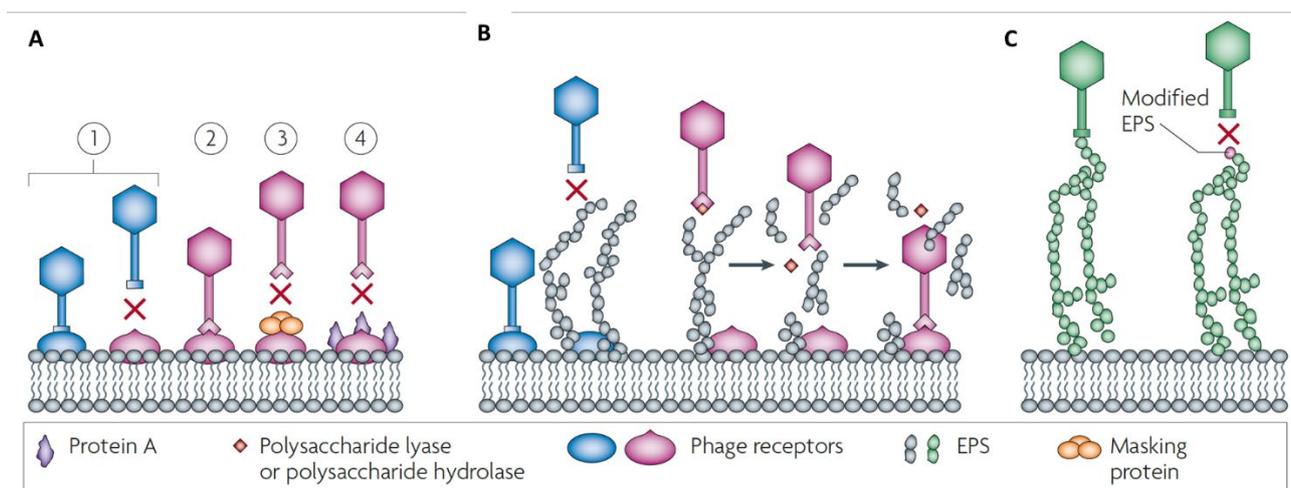


Figure 13 – Stratégies utilisées par les bactéries pour bloquer l'adsorption des phages. (A) Les bactéries peuvent empêcher les phages de se fixer à leur surface (1) en modifiant leurs récepteurs ; (2) les phages peuvent alors s'adapter afin de reconnaître ces nouveaux récepteurs. (3, 4) Les bactéries peuvent alors produire des protéines masquant le récepteur du phage (comme la protéine A chez *Staphylococcus aureus*). (B) L'adsorption des phages peut également être bloquée par la production d'exopolysaccharides (EPS) pouvant être contournée par la production de polysaccharide lysases ou polysaccharide hydrolases par les phages. (C) Les phages ont également évolué afin de reconnaître des polysaccharides comme l'antigène O et l'antigène K.³¹⁴

Après son attachement à son récepteur, le phage va pouvoir, à l'aide d'hydrolase, digérer la paroi bactérienne afin d'injecter son ADN³¹¹. Des mécanismes d'exclusion de surinfection, le plus souvent codés par des prophages, permettent de bloquer l'entrée de l'ADN phagique dans la cellule. Les mécanismes sous-jacents de ce dernier mode d'action

ne sont pas encore très bien compris, bien que certains aient déjà été décrits notamment chez divers Lactocoques et *Streptococcus thermophilus*^{316,317} ou plus récemment chez *Salmonella enterica*³¹⁸. Si le phage parvient à se fixer à son récepteur et à injecter son ADN dans la cellule, d'autres mécanismes de défenses vont se mettre en place afin de lutter contre l'infection. Parmi ces systèmes de défenses, on peut distinguer ceux ciblant les acides nucléiques viraux (les systèmes de restriction-modification et les systèmes CRISPR-Cas), et ceux conduisant à la mort cellulaire induite par la bactérie évitant la propagation de l'infection aux autres membres de la communauté (**Figure 14**)³¹⁹. Parmi ces mécanismes, les plus sophistiqués, les plus abondants et les mieux caractérisés sont ceux ciblant les acides nucléiques³²⁰⁻³²². En tête de liste viennent les systèmes de restriction-modification (RM) qui permettent le clivage ou la dégradation d'ADN *via* la reconnaissance de motifs spécifiques sur le génome viral (**Figure 14A**). Ces motifs sont modifiés dans le génome de la cellule hôte, généralement par méthylation, afin d'éviter que la bactérie ne cible son propre ADN (exception faite des systèmes RM de type IV qui fonctionnent à l'inverse). Ils peuvent être classés en quatre types et sont présents chez plus de 74% des procaryotes^{322,323}. Le nombre moyen de systèmes RM est à peu près constant et égal à deux par génome³²². Les modifications de l'ADN permettant de discriminer le *soi* du *non soi* ne se limitent pas à la méthylation et peuvent impliquer d'autres marquages comme la modification des résidus guanine en 7-déazaguanine³²⁴. Les systèmes BREX (*Bacteriophage Exclusion*)³²⁵ et DISARM (*Defence Islands System Associated with R-M*)³²⁶ utilisent également des mécanismes de protection du génome par méthylation de l'ADN, mais les mécanismes permettant la reconnaissance du phage sont toujours inconnus. Tous ces mécanismes constituent l'immunité innée bactérienne. Une grande partie des bactéries et des archées contiennent dans leur génome le système adaptatif CRISPR-Cas (*Clustered Regularly Interspaced Short Palindromic Repeats*) qui permet la reconnaissance et la dégradation de l'ADN viral (**Figures 14B**)³²⁰. La reconnaissance de l'ADN viral se fait par l'acquisition de courts fragments de séquences dérivées d'ADN viral lors d'évènement infectieux qui sont incorporés dans le génome de la bactérie et appelées *spacers*³²⁷. Ces séquences sont ensuite transcrites et maturées en CRISPR-RNA qui permettent de guider le complexe CRISPR-Cas vers sa séquence complémentaire localisée sur le génome viral cible³²⁰. Il existe plusieurs types de

systèmes CRISPR-Cas comprenant 2 classes, 6 types et plus de 20 sous-types qui diffèrent dans leur composition de la machinerie d'interférence, leur mécanisme de ciblage et la nature des acides nucléiques ciblés qui peut être soit de l'ARN soit de l'ADN^{328,329}. Dans la plupart des cas, l'acquisition du *spacer* et l'interférence nécessite la présence d'une courte séquence située à proximité de la séquence complémentaire du *spacer*, sur la molécule virale cible³³⁰. Les mécanismes moléculaires des systèmes CRISPR/Cas seront plus amplement détaillés dans les différents modes d'actions des ARNnc du chapitre 2.

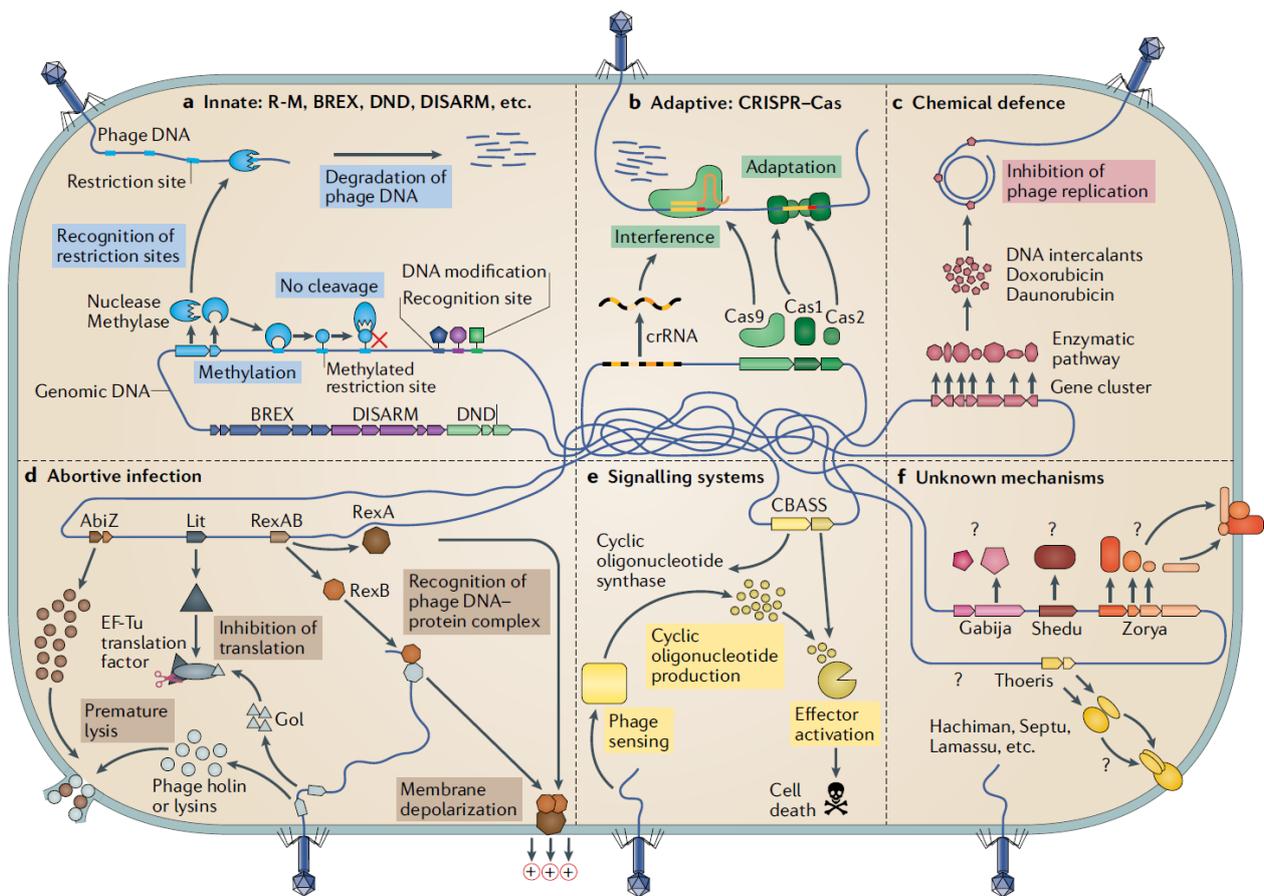


Figure 14 – Systèmes de défenses bactériens contre l'infection phagique. Les systèmes de défense qui ciblent les acides nucléiques englobent à la fois l'immunité innée et adaptative.³¹⁹

Les mécanismes de type *Abi* (*Abortive infection*) ne sont pas des systèmes de défense à proprement parler, mais plutôt une stratégie immunitaire qui se manifeste de différentes manières chez les bactéries pour lutter contre l'infection phagique (**Figures 14D**)^{331,332}. Le principe de cette stratégie repose sur la mort induite de la bactérie après la détection d'une infection afin d'empêcher le phage de compléter son cycle de réplication et d'infecter d'autres bactéries. Le système *Abi* peut ainsi être considéré comme altruiste dans la mesure

où une cellule se sacrifie pour le bien de la communauté. Les systèmes *Abi* ont été identifiés dans de nombreux organismes (dont *E. coli*, *B. subtilis*, *Bacillus licheniformis*, *Shigella dysenteriae*, *Streptococcus pyogenes*, *Vibrio cholerae*, *Lactococcus lactis*)³²¹. Ils sont composés de deux modules fonctionnels : un module senseur de l'infection et un module effecteur responsable de la mort de la cellule ou de la réduction de l'activité métabolique³³¹. Les systèmes *Abi* peuvent détecter différents *stimuli* liés à l'infection phagique comme la réplication du phage, la production des protéines de phage dans la cellule, la transcription à partir de l'ADN du phage ou encore la répression de gènes de l'hôte par un phage³³¹. Une fois le phage détecté, le module effecteur entraînant la mort cellulaire s'active. Ce module doit être fermement régulé en dehors de l'infection afin qu'il n'impact pas la croissance cellulaire et, une fois activé, il doit agir relativement rapidement afin d'entraîner la mort de la cellule ou l'arrêt de son activité métabolique avant que le phage n'ait eu le temps de se répliquer. Les systèmes *Abi* peuvent agir selon différentes méthodes (**Figure 15**) : en dégradant ou dépolarisant la membrane, en inhibant la synthèse protéique par la dégradation des ARNt, en clivant les protéines de la machinerie traductionnelle ou en dégradant sans distinction les ARNm de l'hôte et du phage³³¹.

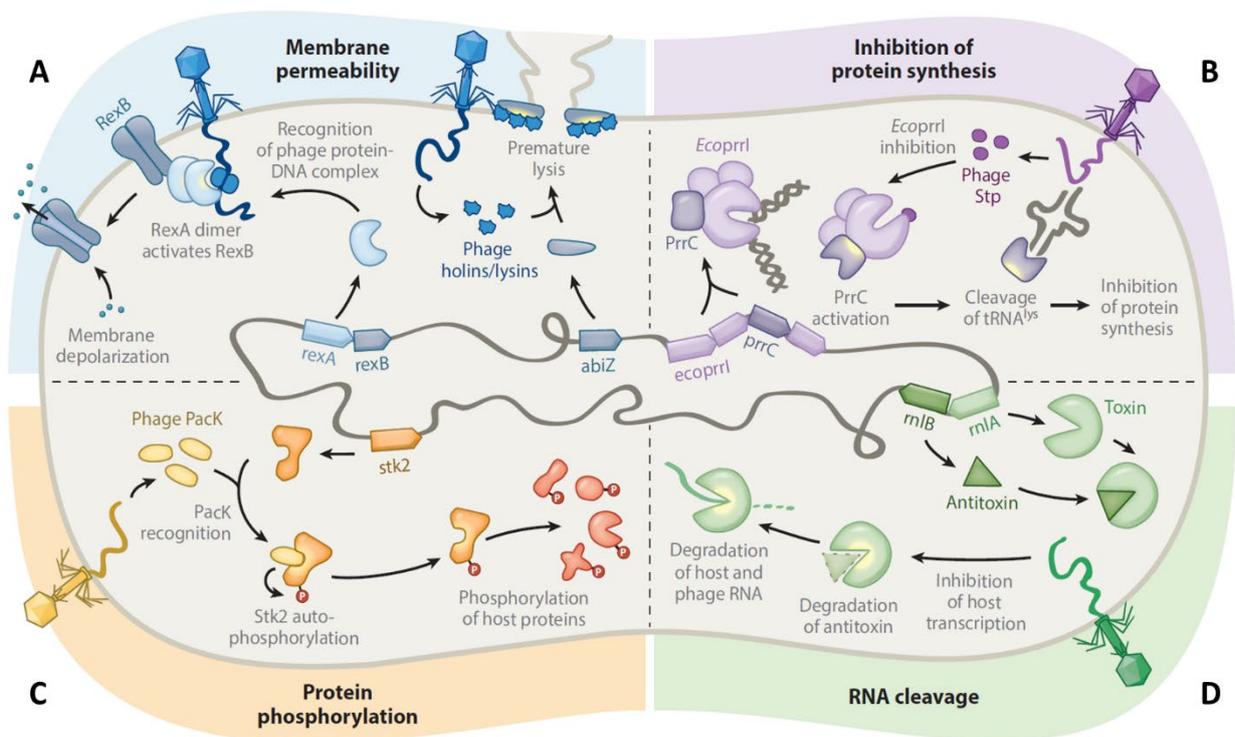


Figure 15 – Mécanismes d'action de différents systèmes *Abi*.³³¹

Dans la plupart des cas, l'activation des systèmes Abi entraîne un arrêt immédiat de la croissance pouvant aboutir à la mort de la cellule, mais parfois leur effet peut être réversible. Il a d'ailleurs été suggéré qu'un arrêt temporaire de la croissance permettrait à la cellule de gagner du temps pour la mise en place d'autres systèmes de défenses comme CRISPR-Cas ou des systèmes RM ^{333,334}.

Le système Rex du phage lysogénique λ d'*E. coli* est l'un des systèmes Abi les mieux caractérisés ³¹⁴. C'est un système à deux composants codés dans le prophage qui nécessite l'intervention des deux protéines RexA et RexB pour une protection contre l'infection par d'autres phages ³³⁵ (**Figure 15A**). La forme inactive de RexA est activée par la détection d'un complexe protéine-ADN du phage qui se forme lors de l'infection en tant qu'intermédiaire de réplication ou de recombinaison ³³⁶. Deux unités de protéine RexA sont requises afin d'activer la protéine ancrée à la membrane RexB qui agit comme un canal ionique permettant le passage des cations monovalents au travers de la membrane interne de la bactérie entraînant une réduction du potentiel membranaire, une chute du niveau d'ATP, une diminution de la synthèse des macromolécules, un arrêt de la multiplication cellulaire puis une lyse de la cellule, stoppant par conséquent la multiplication du phage ^{335,336}. Il existe un mécanisme d'exclusion du phage λ permettant de le protéger du système RexAB lors de son entrée en phase lytique. La surexpression de la protéine RexB permettrait de favoriser l'association d'un monomère RexA avec RexB empêchant ainsi l'ouverture du canal et protégeant le phage λ de son système de défense Abi ^{335,336}. Deux autres systèmes Abi d'*E. coli*, Lit et la nucléase anticodon PrrC, empêchent l'infection phagique d'aboutir en inhibant la machinerie de traduction, ce qui entraîne la mort de la cellule (**Figure 15B**) ³³⁶.

Les mécanismes Abi décrits précédemment sont actifs chez *E. coli* mais beaucoup d'autres de ces systèmes ont été découverts chez les bactéries à Gram positif, en particulier chez *L. lactis*. Aujourd'hui, il existe plus de 20 systèmes Abi (de AbiA à AbiZ) différents qui ont été décrits chez les lactocoques ^{331,332}. Presque toutes les protéines Abi lactococcales sont codées sur des plasmides et une même protéine Abi peut être active contre des phages d'un, de deux ou de trois différents groupes ³³². Les différentes protéines Abi des lactocoques peuvent agir selon différents mécanismes. Par exemple, AbiA, AbiF, AbiK, AbiP et AbiT

agissent sur la réplication de l'ADN, tandis que AbiB, AbiG et AbiU affectent la transcription, AbiC limite la production de la protéine de la capsid du phage, et AbiE, AbiL et AbiQ affecte l'empaquetage de l'ADN. Ou encore AbiD1 qui utilise un mécanisme d'interférence avec une endonucléase de phage RuvC-like et AbiZ qui entraîne la lyse prématurée de la cellule infectée ^{314,331,332}.

Un système récemment découvert appelé CBASS (*Cyclic Oligonucleotide Based Anti-phage Signalling System*) utilise une forme particulière d'Abi (**Figure 14E**) ³³⁷. Il est retrouvé dans plus de 10% des séquences des génomes bactériens, et utilise une voie de signalisation dépendante d'oligonucléotides cyclique (comme le c-GMP-AMP) qui va activer une molécule effectrice entraînant ainsi la mort de la cellule. Une étude récente a permis de découvrir neuf nouveaux systèmes de défense contre les phages (dont Hachiman, Thoreris, Zorya, Gabija et Shedu), très répandus dans les génomes de bactéries et d'archées et dont les mécanismes moléculaires d'action restent encore à déterminer (**Figure 14F**) ³³⁸. Enfin, des espèces de *Streptomyces* produisent de petites molécules appelées doxorubicine et daunorubicine qui agissent comme agents intercalant de l'ADN et qui bloquent spécifiquement la réplication de l'ADN phagique sans interférer avec la réplication de la bactérie hôte (**Figure 14C**) ³³⁹.

3. Phages et virulence chez les bactéries

Selon les conditions, les phages tempérés peuvent établir un cycle lytique ou lysogénique et établir ainsi une relation stable avec leur hôte pouvant avoir de profondes répercussions sur leur mode de vie, leur *fitness* et leur virulence. Les prophages constituent une des principales sources de diversité génétique et plusieurs d'entre eux ont été associés à la virulence de bactéries pathogènes comme *E. coli*, *Clostridium botulinum*, *Streptococcus pyogenes*, *Staphylococcus aureus*, *Corynebacterium diphtheriae* et *Vibrio cholerae* ³⁴⁰. Beaucoup de ces phages codent pour des facteurs de virulence comme des toxines, des protéines effectrices permettant l'invasion, des super-antigènes ou encore des facteurs d'adhésion ³⁴¹. Le phénomène de conversion lysogénique, par lequel un phage converti une bactérie non virulente en bactérie virulente, a été associée à l'émergence de nouveaux clones épidémiques comme la souche de *E. coli* 0157:H7 qui a fait par ce biais l'acquisition de deux

Shigatoxines codées par plusieurs prophages (Sp5 et Sp15)³⁴², ou encore *V. cholerae* qui a acquis sa toxine cholérique grâce au le phage filamenteux CTX ϕ ³⁴³. L'induction des phages, souvent augmentée par certains stress comme l'effet des antibiotiques, peut façonner les communautés bactériennes et favoriser la dissémination de gènes de résistance aux antibiotiques et d'autres EGM, joue donc un rôle clé dans l'évolution des bactéries au sein des communautés³⁴⁴. La relation symbiotique entre prophages et bactéries signifie donc qu'il y a des avantages pour les phages à coder pour des gènes qui confèrent un meilleur *fitness* à leur hôte. Ces gènes ne sont cependant pas requis pour le cycle de développement du phage et sont extrêmement variables d'un phage à l'autre. Ce sont par exemple des éléments génétiques autonomes et non conservés qui ont été appelés *morons* pour la contraction de *more* et *on* puisqu'ils apportent un élément supplémentaire à la bactérie lorsqu'ils sont présents^{345,346}. Ils sont flanqués d'un côté par une région promotrice de type σ^{70} et de l'autre côté par un terminateur de transcription Rho-indépendant. Ce type d'organisation minimise les interférences avec les gènes adjacents du prophage, tout en permettant une expression optimale du ou des gènes de virulence contenu(s) dans ces éléments au cours du cycle lysogène où le prophage est latent et la plupart de ses gènes sont réprimés³⁴⁷. Cette structure favorise également la facilité de transfert des *morons* sous forme de cassette fonctionnelle d'un phage à un autre par recombinaison génétique, sans qu'il n'y ait de conséquence sur la régulation de l'expression des gènes.

L'acquisition de gènes codant des toxines par le biais des phages constitue la majorité des conversions lysogéniques. Parmi celles-ci on peut citer la neurotoxine botulique C1 codée par le phage CE β , les shigatoxines de *E. coli* codées par les phages 933W et H-19B, *Corynebacterium diphtheriae* et sa toxine diphtérique portée par le phage β , *S. pyogenes* produisant plusieurs toxines codée sur un phage, ou encore *S. aureus* dont les différents prophages codent pour les super-antigènes (entérotoxines A et P), la toxine exfoliative A et la Leucocidine de Pantou et Valentine (LPV)^{340,348,349}. Même s'ils ne portent pas de gènes de toxines, les phages peuvent dans certains cas moduler indirectement la production de toxines déjà présentes dans le génome de la bactérie. C'est le cas chez *C. difficile* pour qui des études ont fait état d'une augmentation de la production de toxines chez les souches

porteuses des phages tempérés φ C2, φ C6, et φ C8³⁵⁰, ainsi que pour le phage φ CD38-2 qui augmente la production de TcdA et TcdB jusqu'à un facteur 2 et également le niveau de transcrits des cinq gènes du PaLoc, lorsqu'il est introduit dans une souche épidémique de *C. difficile*³⁵¹. Les mécanismes moléculaires par lesquels les phages altèrent l'expression des gènes restent encore mal caractérisés, bien qu'une étude ait montré que le régulateur transcriptionnel RepR codé par le phage tempéré φ CD119 peut réprimer directement la transcription des gènes des toxines de *C. difficile*³⁵². Dans les souches contenant le prophage φ CD119, le produit du gène *repR* est capable de se fixer à son propre promoteur mais également à la région en aval du promoteur du gène *tcdR* empêchant ainsi son expression et la production du régulateur positif de l'expression des toxines, TcdR³⁵². Il a également été montré que la dynamique de formation et de dispersion des biofilms peut être fortement influencée par la présence de phages chez certains pathogènes comme *Pseudomonas aeruginosa*, *Streptococcus pneumoniae*, *Bacillus anthracis* et *E. coli*³⁴⁰. De même, la capacité de certaines souches à sporuler a également été associée à la présence de certains phages dans leur génome et à des éléments génétiques mobiles *skin*.

4. Les phages chez *C. difficile*

Les phages sont capables d'infecter presque toutes les espèces bactériennes et des régions prophagiques peuvent être trouvées dans la plupart des génomes bactériens. *C. difficile* ne fait pas exception à la règle. Au cours de son cycle infectieux, *C. difficile* doit subsister parmi les communautés microbiennes de l'intestin riches en phages, et doit donc être muni de systèmes favorisant les échanges génétiques favorables pour survivre dans un environnement. Le génome de *C. difficile* est composé de nombreux EGM et structuré en mosaïque. En effet, 11% du génome de la souche 630 est constitué d'EGM, responsable de l'acquisition de traits adaptatifs favorisant le mode de vie dans le tractus intestinal¹⁵⁴. Ces acquisitions doivent être également accompagnées de mécanismes de défense efficaces contre les envahisseurs abondants tels que les phages et les plasmides. Tous les phages de *C. difficile* isolés et caractérisés à ce jour sont tempérés³⁵³. Au cours d'une étude, une forte prévalence de prophages dans les génomes de *C. difficile* a été révélée par séquençage extensif du génome³⁵⁴. Au moins 24 génomes complets de phages sont actuellement

disponibles dans les bases de données, et la plupart des phages identifiés à ce jour font partie de la famille des *Myoviridae* de l'ordre des *Caudovirales*, et six phages sont membres de la famille des *Siphoviridae* ³⁴⁸. Les différentes souches de *C. difficile* contiennent généralement entre 1 et 3 prophages, pouvant aller plus rarement jusqu'à 6, en plus des îlots génomiques contenant des gènes reliés aux phages ³⁴⁸. Certains prophages de *C. difficile*, y compris ceux dont la taille du génome excède 130 kb, peuvent être maintenus de manière stable sous forme d'épisome ³⁵⁴. De plus, la plupart des phages de *C. difficile* ont un spectre d'hôte relativement étroit qui peut s'expliquer par la disponibilité des récepteurs spécifiques à la surface de la cellule hôte, la présence de prophages endogènes conférant la résistance à la surinfection, la présence de systèmes de défense efficaces contre l'invasion phagique (incluant le système CRISPR-Cas) ³⁵⁵, et le système d'exclusion de surinfection faisant intervenir la protéine CwpV ^{348,356}.

Les phages présents chez *C. difficile* contribuent à sa virulence par différentes manières. Une étude a notamment identifié le *locus* complet de la toxine binaire (CdtLoc) dans la séquence du prophage phiSemix9P1 ³⁵⁷. Le *locus* de pathogénicité PaLoc quant à lui partage plusieurs caractéristiques communes avec des séquences de phage, en particulier en ce qui concerne le gène *tcdE* qui code pour une protéine *holine-like* impliquée dans la sécrétion des toxines ¹⁹⁶. D'autre part, il a été montré que les phages phiCD119 et phiCD38-2 ont la capacité de réguler respectivement négativement ³⁵² et positivement ³⁵¹ certains gènes chez *C. difficile*. Cependant l'impact du phage phiCD38-2 sur la synthèse des toxines semble être souche-dépendant, suggérant que l'influence du prophage sur son hôte dépend de l'environnement génétique. Les résultats d'une étude transcriptomique de la souche R20291 dans laquelle a été introduit le prophage phiCD38-2 ont révélé une expression active des gènes de la souche R20291, dont certains appartenant à son prophage naturel phi027, ainsi qu'un potentiel *cross-talk* entre les deux prophages phi027 et phiCD38-2 ³⁵⁸. Lors d'une analyse transcriptionnelle de la souche TW11 de PCR-ribotype 078, une répression générale de l'expression des gènes de la bactérie hôte incluant les systèmes de défenses anti-phage comme CRISPR-Cas, les systèmes de RM et TA de type II ont été observés après l'infection avec le phage JD032 ³⁵⁹. Ces résultats semblent donc suggérer que les gènes des prophages

présents dans les différentes souches de *C. difficile* pourraient interférer avec l'expression des autres gènes du génome ^{353,354}. De plus, il a été montré que le phage phiCDHM1 présente dans son génome une cassette de gènes Agr-like du QS qui pourrait exercer une influence sur les phénotypes associés à la virulence de *C. difficile* ³⁶⁰. Chez *C. difficile*, des éléments du système de défense CRISPR-Cas sont localisés dans les régions prophagiques du génome, comme dans la souche 630 où les séquences des CRISPR 3/4 et CRISPR 15/16 sont fortement exprimées respectivement dans les prophages phiCD630-1 et phiCD630-2 (**Figure 16**), ou encore deux séquences CRISPR transcriptionnellement actives qui sont retrouvées dans le phage phi027 des souches hypervirulentes de PCR-ribotype 027 et enfin une séquence CRISPR associée au gène *cas3* localisée dans le phage phiCD211-like ^{354,358,360,361}.

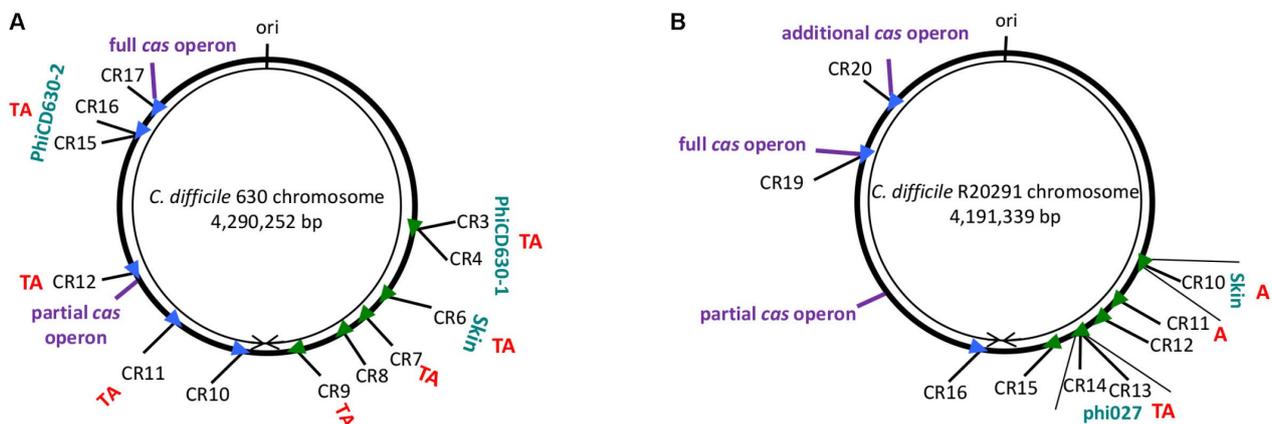


Figure 16 – Vue schématique de la localisation des cassettes CRISPR et de l'organisation des opérons *cas* dans les souches 630 (A) et R20291 (B) de *C. difficile*. ³⁵⁵

Un travail de revue de la littérature a été réalisé sur les interactions entre *C. difficile* et les bactériophages *via* le monde ARN. La revue publiée dans Current Opinion in Microbiology (<https://doi.org/10.1016/j.mib.2021.11.012>) est jointe à la fin du manuscrit en

Publication 1.

I. Les ARN non codants bactériens : présentation générale

1. La découverte des ARN non codants (ARNnc)

Dans les années 1950, les ARN ribosomiaux (ARNr) et les ARN de transfert (ARNt) ont été les premiers ARN non codant (ARNnc) impliqués dans l'expression des gènes à avoir été identifiés. Cependant, le concept d'ARN régulateur de l'expression des gènes peut être attribué au modèle opéron qui a été décrit par Jacob et Monod dans les années 1960. C'est dans les années 1980 qu'est évoqué le rôle d'ARN non codants (ARNnc), les *small nuclear RNA* (ou snRNA), dans l'excision des séquences introniques. Puis s'en est suivi la découverte d'autres ARNnc comme les *small nucleolar RNA* (ou snoRNA), les ARNnc impliqués dans la répression des gènes chez les plantes, la découverte des ARN interférents (siRNA), la compréhension des mécanismes des ARN antisens bactériens et la découverte au début des années 2000 des microARN (miRNA) qui a mis en évidence l'importance que peuvent avoir les ARN dans la régulation post-transcriptionnelle des gènes ³⁶²⁻³⁶⁴. Les petits ARN régulateurs ne sont pas limités exclusivement aux eucaryotes et ont même été identifiés chez les procaryotes quelques années avant la découverte des miRNA et des siRNA chez les eucaryotes ³⁶⁵. Au début des années 1980, de nombreux travaux ont révélé l'existence de riborégulateurs modulant divers processus cellulaires comme l'inhibition de la réplication du plasmide ColE1 d'*E. coli* par l'ARNnc RNA I ^{366,367}, la répression de la transposition du transposon Tn 10 par un ARN transcrit à partir de son propre promoteur pOUT ³⁶⁸ ou encore l'inhibition de la traduction de l'ARNm codant pour la porine OmpF d'*E. coli* par l'ARNnc MicF ³⁶⁹.

Bien que quelques candidats ARN aient été identifiés de manière fortuite ou non, les ARNnc ne sont au centre de l'attention que depuis peu de temps dans la mesure où leur abondance et leur rôle dans la régulation de l'expression des gènes ont été pendant longtemps sous-estimés. En effet, contrairement aux phases ouvertes de lectures (ORFs ; *Open Reading Frame*), les ARNnc n'ont pas été immédiatement révélés par le séquençage des génomes et, par leur petite taille (généralement d'une centaine de nucléotides), sont plus susceptibles d'être manqués. Néanmoins, les avancées technologiques comme

l'amélioration des programmes de prédiction en biologie computationnelle, le séquençage haut-débit ou les puces à ADN couvrant la totalité du génome ont permis l'identification de centaines d'ARN régulateurs chez de nombreuses bactéries³⁷⁰. Chez *E. coli* par exemple, plus de 60 ARNnc ont été découverts (soit environ 1% à 2% du nombre de gènes codants connus) au début des années 2000^{371,372}, et de nombreux miRNA ont été identifiés chez la souris, le nématode, la drosophile et l'humain^{373,374}.

2. Caractéristiques principales des ARN non codants bactériens

Les riborégulateurs bactériens constituent un groupe hétérogène de molécules agissant par différents modes d'action afin de moduler une large palette de processus physiologiques et qui ne présentent généralement pas de phase codante. Il existe cependant quelques exceptions d'ARNnc qui présentent une petite séquence codant pour un peptide fonctionnel : le RNAIII contrôle certains facteurs de virulence chez *S. aureus* comme les protéines de surface et la production des toxines *via* l'appariement par complémentarité de base avec ses ARN cibles tout en codant pour une toxine. L'ARN *psm-mec* code pour un petit peptide pro-inflammatoire chez *S. aureus* et l'ARN SgrS code pour une petite protéine appelée SgrT qui inhibe l'activité du transporteur du glucose PtsG afin de permettre à *E. coli* d'utiliser une autre source de carbone que le glucose 6 phosphate (G6P)^{375,376}. Les tailles des ARNnc sont assez hétérogènes et peuvent varier d'environ 50 à plus de 500 nucléotides. Chez *E. coli* par exemple, la plupart des ARNnc ont une taille d'environ 100 nucléotides ce qui permet de limiter la taille de leur site d'appariement avec leurs ARNm cibles. Bien que l'on puisse penser que la taille du site d'appariement soit proportionnelle à la taille de l'ARNnc, la plupart des riborégulateurs s'apparient avec leur cible *via* des séquences relativement courtes (moins de 20 nucléotides) et conservées appelées région de graine (ou *seed region*)³⁷⁷.

Bien que les ARNnc ont pendant longtemps été associés exclusivement aux régions intergéniques (IGR ; *Intergenic Region*), ces dernières ne sont pas les seules régions hébergeant les ARNnc, et certains d'entre eux peuvent être retrouvés notamment dans les régions non traduites 5' et 3' (5'UTR et 3'UTR) des ARNm³⁷⁷. D'autres types de molécules ARN peuvent héberger de petits ARN régulateurs comme les pré-ARNt et les transcrits des

ARNt. Cette nouvelle famille peut être subdivisée en fonction de la taille des fragments ARNt (15 à 60 nucléotides) et l'origine des molécules d'ARN (ARNt mature ou précurseur)³⁷⁸. Il est donc maintenant bien établi que les ARNnc ne proviennent pas que de plasmides ou d'IGR mais également d'ARNm précurseurs. De plus, si la plupart d'entre eux sont codés sur le génome bactérien, certains sont localisés sur les EGM, les îlots de pathogénicité, les prophages ou les plasmides^{370,379}.

3. Rôles physiologiques des ARN non codants bactériens

Les microorganismes doivent en permanence moduler l'expression de leurs gènes afin de survivre et de s'adapter à un environnement en constante fluctuation. Pour cela, les bactéries utilisent des mécanismes de régulation fine faisant intervenir des facteurs de transcription et des petits ARN régulateurs afin de modifier rapidement leur transcriptome et leur métabolisme³⁸⁰. Il existe très peu d'ARNnc exprimés constitutivement et, pour une grande majorité, leur transcription est induite dans des conditions environnementales spécifiques comme la phase stationnaire de croissance ou un stress environnemental (comme des températures élevées ou basses). Aujourd'hui, le rôle des ARNnc dans la régulation de nombreux processus physiologiques, comme la réponse adaptative et la régulation des facteurs de virulence chez les bactéries pathogènes, ne laisse plus aucun doute et de nombreux exemples de *stimuli* environnementaux engendrant une modulation par ARNnc de ces processus sont caractérisés^{381,382}.

a. Maintien de l'homéostasie du fer

Chez *E. coli*, le petit ARN RyhB est sous le contrôle du régulateur de l'homéostasie du fer Fur. Fur est un répresseur Fe²⁺-dépendant qui, lorsque le fer est abondant dans l'environnement, réprime directement les gènes impliqués dans l'acquisition du Fer, et active indirectement la synthèse d'enzyme permettant le stockage ou l'utilisation du fer en réprimant RyhB. Quand le fer vient à manquer, RyhB n'est plus réprimé et inhibe l'expression de plus de 15 protéines non essentielles fixant du fer^{383,384}.

b. Remodelage de la membrane

Chez les entérobactéries, de nombreux ARNnc (comme MicA, MicC, MicF, MicL, RyhB, OmrA, et OmrB) sont impliqués dans le remodelage de la membrane externe en ciblant les

ARNm des porines et protéines de surface. Par exemple, les ARN MicF et MicC inhibent respectivement la traduction des ARNm codant pour les porines OmpF et OmpC ce qui permet de contrôler la taille des pores de la membrane lors de changement d'osmolarité de l'environnement ^{385,386}. D'autres ARNnc interviennent dans la régulation des mécanismes permettant le maintien de l'homéostasie membranaire en inhibant la production de certaines protéines membranaires durant un stress de l'enveloppe. Par exemple, l'accumulation de protéines de la membrane externe mal conformées dans le périplasme active la transcription σ^E dépendante des ARN MicA, MicL et RybB qui vont exercer un contrôle négatif de l'expression de plus d'une vingtaine de protéines de la membrane externe réduisant ainsi considérablement la synthèse *de novo* de ces protéines et diminuant le stress membranaire ^{387,388}.

c. Motilité et biofilms

La transition entre le mode de vie motile et le mode de vie sessile est sous le contrôle de réseaux de régulation transcriptionnelle complexes et de régulateurs clés. Un second niveau de contrôle dépendant d'ARNnc et permettant de réguler ce passage d'un mode de vie à un autre a été mis en évidence chez certaines entérobactéries. Des travaux ont notamment montré que plusieurs ARNnc (McaS, MicA, ArcZ, OmrA, OmrB, OxyS, SdsR et GadY) ciblent l'opéron codant pour le régulateur de l'expression flagellaire FlhDC afin de moduler son expression par différents mécanismes ³⁸⁹.

d. Métabolisme du sucre

La régulation du transport et du métabolisme des glucides implique également de nombreux ARNnc (**Figure 17**) ³⁹⁰. L'un d'entre eux est l'ARNnc Spot 42 identifié chez *E. coli* et codé par le gène *spf*. Son expression est contrôlée par un mécanisme de répression catabolique qui, lorsque le taux de glucose est faible, inhibe son expression *via* le complexe CRP-AMPc et inversement lorsque le taux de glucose augmente ^{381,390}. Après sa production et accumulation, Spot 42 va interagir avec l'extrémité 5' de l'ARNm de *galK* codé par l'opéron *galETKM* impliqué dans la conversion du galactose en G1P (glucose-1-phosphate), afin d'inhiber la traduction de GalK ³⁹¹. Spot 42 régule négativement d'autres cibles impliquées dans l'absorption et l'utilisation de diverses sources de carbones qui sont également sous le

contrôle du complexe CRP-AMPc³⁹². L'accumulation de G6P dans la cellule crée un stress métabolique ce qui a pour conséquence l'activation de la protéine SgrR qui active la transcription de l'ARNnc SgrS. SgrS interagit avec l'ARNm ptsG qui code pour un transporteur du glucose et va promouvoir sa dégradation *via* la RNase E au niveau de la membrane interne, stoppant ainsi l'importation du glucose dans la cellule et diminuant le stress métabolique^{393,394}. Il est intéressant de noter que SgrS possède également une séquence codant pour la petite protéine SgrT qui permet de bloquer l'importation du glucose par un autre mécanisme³⁹⁵. Un autre exemple de régulation du métabolisme des sucres par des ARN est celui des deux ARNnc hautement similaires GlmY et GlmZ. Chez *E. coli*, le gène *glmS* (codant pour la glucosamine-6-phosphate [GlcN6P] synthase) est sous le rétrocontrôle négatif de son produit final, la GlcN6P. GlmZ active directement la traduction de l'ARNm de *glmS* par un mécanisme antisens. Lorsque le taux de GlcN6P est élevé, GlmZ est clivé en une forme courte inactive par la RNase E recrutée par la protéine adaptatrice RapZ, inhibant ainsi la production de la protéine GlmS³⁹⁶. A l'inverse, lorsque le taux de GlcN6P est bas, GlmY agit en amont de GlmZ en empêchant son inactivation par clivage, promouvant ainsi la traduction de GlmS³⁹⁷.

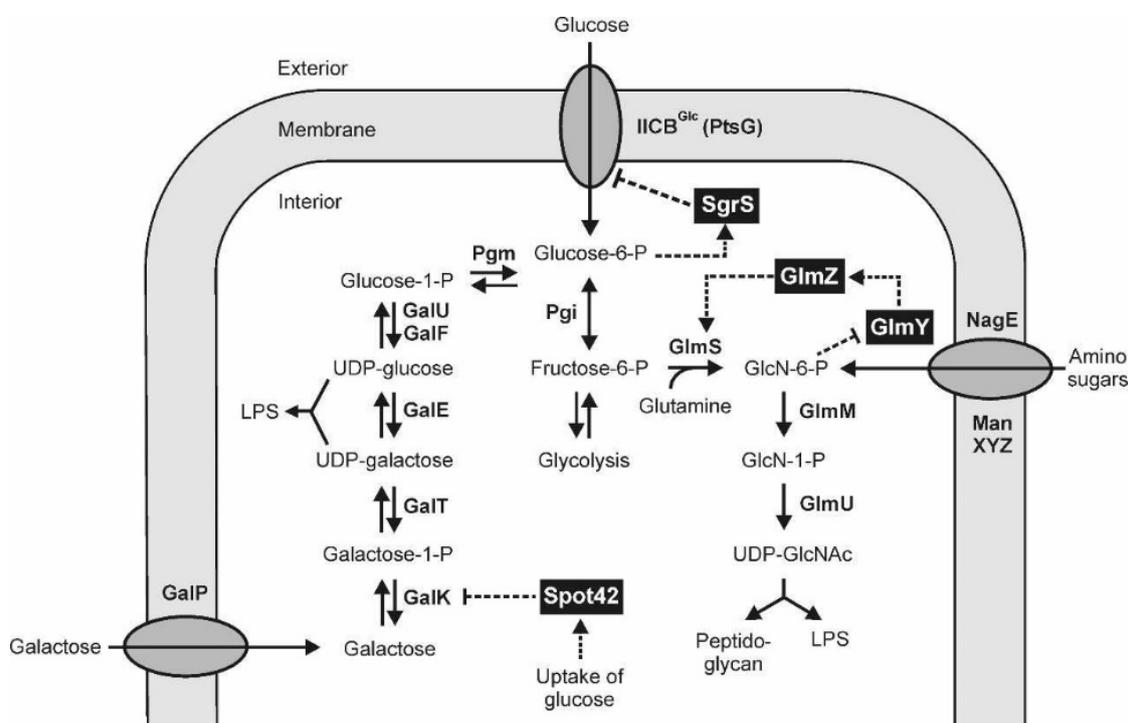


Figure 17 – Voies métaboliques du galactose, du glucose et des osamines régulées par ARNnc.³⁹⁰

e. Régulation des facteurs de transcription

L'expression de la plupart des ARNnc est contrôlée par des facteurs de transcription dont la plupart contiennent dans leur régulon des ARNnc conservés, et inversement de nombreux ARNm de facteurs de transcription sont ciblés par des ARNnc^{398,399}. Par exemple, chez les entérobactéries, trois ARNnc (DsrA, RprA, ArcZ) régulent positivement le facteur RpoS par appariement antisens et stabilisation de son ARNm. Un quatrième ARNnc, OxyS, permet quant à lui de réguler négativement la traduction de RpoS. Ces ARNnc sont synthétisés en réponse à différents stress et sous le contrôle de différents régulateurs, permettant à la bactérie d'intégrer et de répondre à de nombreux signaux de stress pour le contrôle de la traduction de RpoS⁴⁰⁰. Parmi les exemples de facteurs de transcription clés qui sont ciblés par des ARNnc on peut citer le facteur CsgD d'*E. coli* qui régule l'expression des gènes de la protéine Curli et qui est lui-même régulé par les ARNnc OmrA/B⁴⁰¹, le facteur Lrp impliqué dans la régulation de nombreux gènes de la synthèse d'acides aminés qui est la cible de l'ARNnc GcvB⁴⁰², mais aussi les deux régulateurs majeurs du QS LuxR et AphA chez *Vibrio* qui sont respectivement réprimés et activés par l'ARNnc Qrr⁴⁰³.

f. Interactions entre prophages, EGM et bactérie hôte

En plus de leur position centrale dans les mécanismes de défense contre les phages avec les systèmes CRISPR-Cas, les ARNnc sont impliqués dans les interactions entre les phages ou autres EGM et le génome de la bactérie hôte au niveau post-transcriptionnel^{404,405}. Des études récentes ont mis en lumière ces interactions durant lesquelles les ARN des phages modifient l'expression des gènes bactériens et les ARN de la bactérie remodelent à leur tour l'expression des gènes de phages^{404,406}. Les ARN régulateurs peuvent également contribuer à la décision lyse-lysogénie lors de l'infection par les phages tempérés. Chez les bactéries entéropathogènes, les interactions entre EGM et bactéries contribuent également au contrôle de la virulence⁴⁰⁷. Chez *Salmonella enterica*, ces réseaux de régulation incluent à la fois des ARNnc codés par le génome bactérien (SgrS et RprA) et des ARNnc acquis par transfert horizontal (PinT et IsrM des îlots de pathogénicité SPI-1 et SPI-2, IsrK du prophage Gipsy, Stnc6030 du prophage BTP1)^{318,407-410}. Des études récentes ont identifié un petit ARN, StxS, généré par la terminaison prématurée de la transcription du gène codant la Shiga

toxine d'*E. coli* entérohémorragique (EHEC) et agissant en *trans* sur l'expression des gènes de la bactérie ⁴¹¹. Un autre exemple de cross-régulation entre les gènes de prophages et de la bactérie hôte a été illustré avec la découverte d'anti-ARNnc codés par des prophages de la bactérie pathogène *E. coli* qui agissent comme des ARN éponges interférant avec l'activité des ARNnc ciblés ⁴¹². Parmi les 55 ARNnc dépendant de la protéine Hfq identifiés dans les régions phagiques du génome des EHEC, se trouvent notamment les ARNnc AsxR et AgvB qui inhibent respectivement l'action des ARNnc FnrS et GcvB codés par le génome bactérien impactant ainsi la virulence de la bactérie hôte ⁴¹². Des études de transcriptomique ont permis d'identifier d'autres ARNnc codés par les phages et prophages d'autres bactéries pathogènes comme sRNA1 et sRNA2 du phage PAK_P3 de *P. aeruginosa* ^{413,414}.

g. Les systèmes toxines-antitoxines

Les systèmes toxine-antitoxine (TA) sont des modules codant pour des systèmes à deux composants qui sont une toxine stable et une antitoxine instable permettant de réprimer la fonction ou l'expression de la toxine ⁴¹⁵. La surexpression de la toxine a pour conséquence soit la mort de la cellule, soit une inhibition de la croissance. Les systèmes TA ont initialement été découverts sur des plasmides, ce qui confère à ces derniers une stabilité de maintien *via* la mort post-ségrégation des cellules dépourvues de plasmide (PSK; *Post Segregational Killing*) ⁴¹⁶. Ces systèmes sont retrouvés parfois en grand nombre dans les chromosomes des bactéries et des archées et, bien que de nombreuses fonctions aient été proposées et discutées dans la littérature, c'est une question qui reste encore débattue. Parmi les fonctions discutées se trouve notamment le maintien des prophages, la stabilisation de régions chromosomiques (*via* le PSK), la prévention d'infection phagique (*via* des mécanismes d'Abi), la réponse au stress, la virulence et la formation de cellules dormantes et tolérantes aux antibiotiques appelées *persisters* ^{415,417,418}. Les modules TA sont classés en six différents types dépendant de la nature et de l'action de l'antitoxine ⁴¹⁸. Alors que la toxine est systématiquement une protéine, l'antitoxine peut être soit une protéine (types II, IV à VI) ou un ARN simple brin (types I et III) ⁴¹⁵. L'action de l'antitoxine ARN peut soit réprimer la traduction (type I) ou l'activité protéique (type III) de la toxine associée. Le système TA de type II est probablement le mieux étudié et utilise une protéine comme antitoxine qui

neutralise la toxine associée en interagissant directement avec elle. Les antitoxines protéiques appartenant au système de type IV annulent indirectement l'effet de la toxine en agissant comme molécules compétitrices pour la fixation de la toxine à ses cibles cellulaires (**Figure 18**). Pour les TA de type V, la toxine est inactivée au niveau transcriptionnel grâce au clivage de l'ARNm de la toxine par une antitoxine RNase. Enfin, le système TA de type VI implique l'action d'une antitoxine protéique qui induit la dégradation de la toxine par le recrutement d'une protéase ⁴¹⁵. Une étude récente a décrit un système TA permettant le maintien d'un *locus* CRISPR-Cas type I-B ⁴¹⁹. Cette étude a révélé une région intergénique de 311 pb située dans la cassette de gènes *cas* codant pour le complexe effecteur Cascade qui contient deux éléments génétiques, l'un codant pour une toxine (*creT*) et l'autre pour un ARN antitoxine (*creA*). L'ARN *creA* s'apparie à l'ARNm de *creT* grâce au complexe Cascade inhibant ainsi son expression.

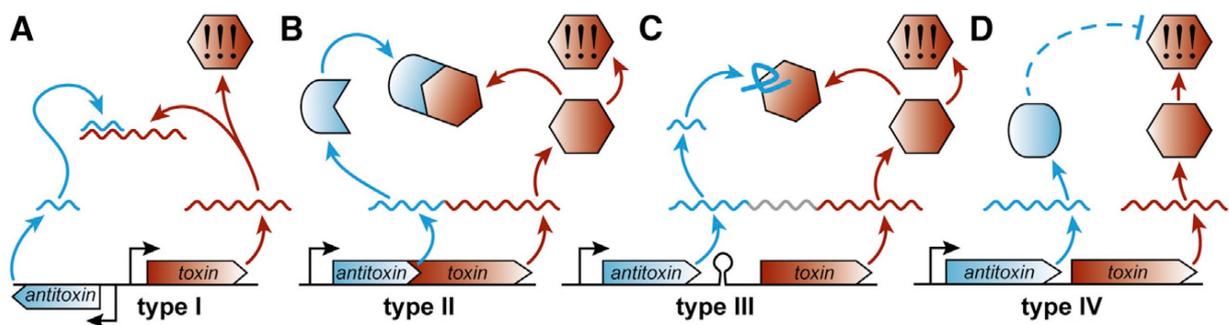


Figure 18 – Aperçu des mécanismes des quatre principaux systèmes TA. ⁴¹⁵

II. Les différents modes d'action des ARN non codants bactériens

Les riborégulateurs bactériens peuvent agir sur la transcription, la traduction, la stabilité des ARNm, l'intégrité de l'ADN et le *silencing* ^{420,421}. Afin de pouvoir exercer leurs actions sur leurs cibles, ils utilisent divers mécanismes d'action, dont les changements de conformation des ARN, la fixation à des protéines, l'appariement par complémentarité de base avec d'autres ARN et en interagissant avec l'ADN ^{362,370}. Le groupe le plus grand et le plus étudié est celui des ARNnc agissant par complémentarité de base avec leur cible, modifiant généralement la traduction et la stabilité des ARNm ^{365,422}. La majorité de ces ARN régule les différentes réponses aux variations des conditions environnementales. Certains d'entre eux, agissant en *cis*, sont situés à proximité de leur ARNm, et d'autres, agissant en *trans*, sont

localisés dans une région génomique distante de l'ARNm qu'ils régulent (**Figure 19**)^{377,379}. Les ARN codés en *cis*, appelés aussi ARN antisens, présentent une séquence entièrement complémentaire de leurs cibles, alors que la séquence des ARNnc agissant en *trans* n'est que partiellement complémentaire avec celle de leurs cibles et leur action requiert généralement l'assistance de la protéine ARN chaperonne Hfq^{362,377}. Enfin, le groupe de riborégulateurs découvert plus récemment connu sous le nom d'ARN CRISPR impliqué dans la lutte contre l'infection phagique et le transfert de certains plasmides.

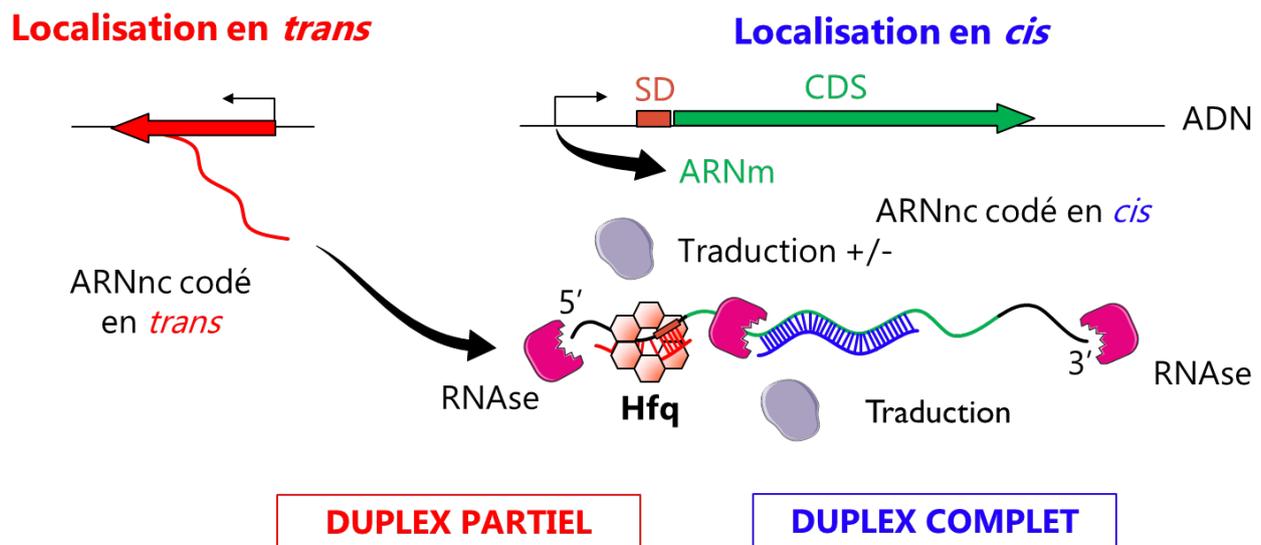


Figure 19 – Mécanismes de régulation par ARNnc agissant par complémentarité de base, localisés en *trans* ou en *cis* de leur cible ARNm.³⁸²

1. Mode d'action par complémentarité de base

a. ARNnc codés en *cis*

Les ARN régulateurs agissant en *cis* exercent leur action grâce au chevauchement de leur séquence avec celle de leur gène cible qui est localisé dans une région génétique très proche. Ils peuvent être divisés en deux sous-catégories (**Figure 20**) : (1) les ARN antisens (ARNas) dits véritables qui présentent une complémentarité parfaite avec tout ou partie d'un ou plusieurs gène(s) localisé(s) sur le brin opposé de leur séquence, et (2) les longues UTRs (*Untranslated Regions*) chevauchantes entre deux ARNm contigus et dont les séquences s'opposent sur les deux brins d'ADN. Les chevauchements des 5'UTR longs sont générés lorsque le promoteur d'un gène se trouve sur le brin opposé codant pour un autre gène, et les chevauchement 3'UTR peuvent être la conséquence de l'absence de terminateur de transcription à proximité du codon STOP ou de la translecture d'un terminateur donnant lieu

à une terminaison de la transcription dans une région chevauchante d'un autre gène localisé sur le brin opposé.

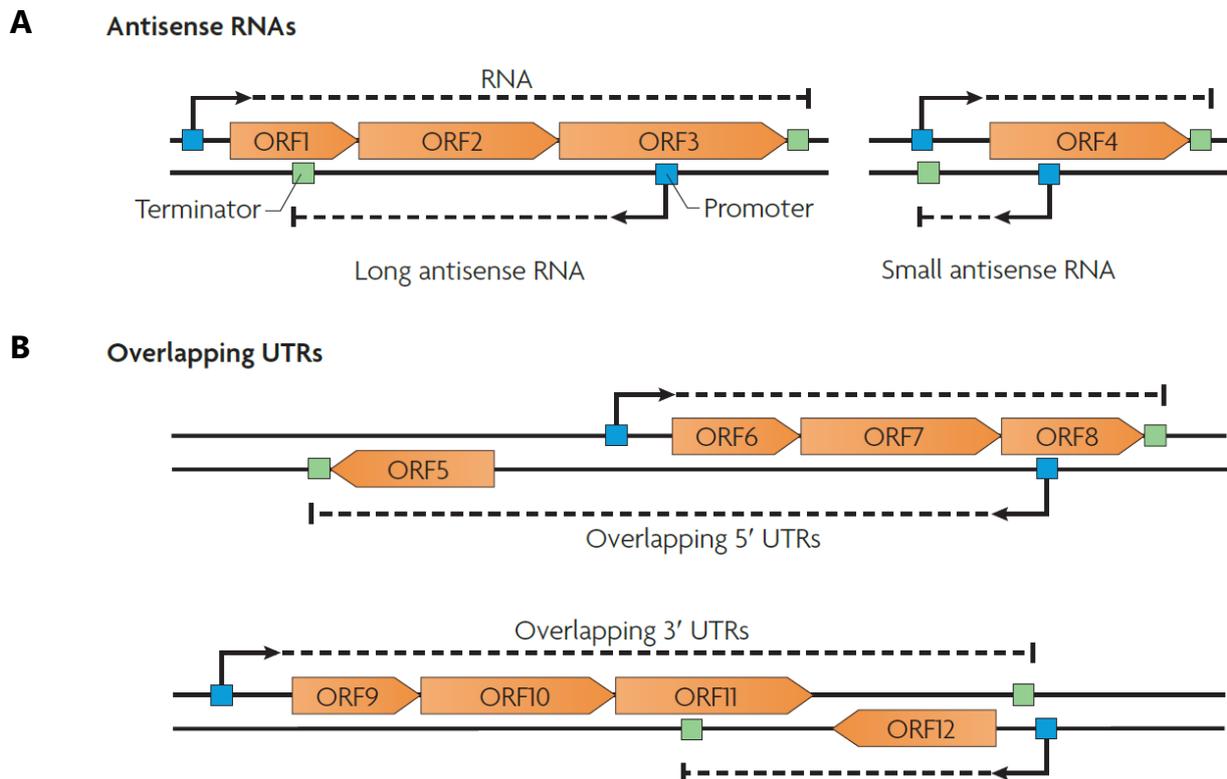


Figure 20 – Représentation schématique des deux catégories d'ARN antisens (ARNs). (A) Les ARNs dits véritables, peuvent être soit de longs ARN couvrant plus d'une ORF (ORF1, ORF2 et ORF3) ou de petits ARN ne couvrant que la séquence Shine-Dalgarno d'une ORF (ORF4) affectant ainsi la stabilité de l'ARNm et/ou sa traduction. (B) Les longues 5'UTR sont générées lorsque le site d'initiation de la transcription de gènes (ORF5) est chevauchant avec la séquence d'autres gènes du brin opposé (ORF6, ORF7, ORF8). A l'inverse, les 3'UTR chevauchantes sont générées lorsque le site de terminaison de la transcription de gènes (ORF9, ORF10 et ORF11) est situé sur le brin opposé d'un autre gène (ORF12).³⁸²

Les ARNs situés sur les plasmides, les phages et les transposons furent parmi les premiers ARNc à être étudiés. Leur action est souvent liée au maintien du nombre adéquat de copies des plasmides et autres éléments génétiques auxquels ils sont associés, et pour ce faire, ils peuvent agir selon différents mécanismes^{423,424}. Par exemple, chez *E. coli* l'ARN RNA I contrôle le nombre de copies du plasmide ColE1 en empêchant la formation de primers de réplication à partir du RNA II^{366,367}. De nombreux ARNs sont codés sur le brin opposé de gène de transposases. L'ARNs RNA-OUT du transposon Tn 10, l'un des premiers ARNs à avoir été découvert, permet de réprimer la transposition en diminuant le niveau de transposase. D'autres ARNs associés à des transposases ont également été identifiés chez *S. enterica*, *Caulobacter crescentus* et *L. monocytogenes*⁴²⁵. Bien que jusqu'à très récemment

la plupart des ARNas n'ait été associée qu'aux seuls bactériophages, plasmides et transposons, il y a de plus en plus d'exemples d'ARNas chromosomiques caractérisés. Les ARNas impliqués dans le contrôle de la synthèse de protéines toxiques, initialement découverts avec l'ARN Sok du plasmide R1 chez *Enterococcus faecalis*, sont de mieux en mieux décrits ⁴²⁶⁻⁴²⁸. Les systèmes TA de type I sont un cas particulier d'ARNas régulant la traduction d'un ARNm codant pour une toxine. On peut citer en exemple le système TA de type I SprF1/SprG1 de *S. aureus* dont l'ARNas SprF1 se fixe à l'extrémité 3' de l'ARNm de *sprG1* (codant pour deux peptides sécrétés ayant une activité hémolytique et antibactérienne) et stimulant sa dégradation ⁴²⁹. D'autres ARNas ont été associés à la régulation de la synthèse de facteur de transcription, comme l'ARN GadY d'*E. coli* situé sur le brin opposé du gène *gadX* codant pour un régulateur transcriptionnel de la réponse acide et dont l'extrémité 3'UTR est chevauchante avec la séquence de GadY, conférant ainsi une stabilité accrue à l'ARNm de *gadX* durant la phase stationnaire ⁴³⁰. Les ARNas peuvent également être impliqués dans des divers aspects de la vie bactérienne comme la régulation de l'expression de certaines enzymes du métabolisme et de facteurs de virulence ^{431,432}.

Les mécanismes d'actions des ARNas passent par l'altération de la transcription, de la stabilité ou de la traduction de l'ARN sens ciblé (**Figure 21**) ⁴²⁵. Les ARNas peuvent agir sur la transcription des gènes codés sur le brin opposé selon deux différentes façons : l'interférence transcriptionnelle ou l'atténuation de transcription (**Figure 21A et B**). L'interférence transcriptionnelle a lieu quand la transcription à partir d'un promoteur est empêchée à cause d'un second promoteur présent en *cis* ⁴³³. On peut citer l'exemple du *riboswitch* S-box chez *C. acetobutylicum* qui, lors d'un changement de niveau de soufre, synthétise un ARNas qui inhibe l'expression des gènes de l'opéron *ubiG* situé en amont non pas par complémentarité de base mais par un mécanisme de compétition avec l'ARN polymérase entre les transcrits sens et antisens ⁴³⁴. L'atténuation de transcription se produit lorsque la transcription du gène sur le brin opposé est prématurément terminée. Dans quelques rare cas, généralement sur les plasmides, l'appariement de l'ARNas à l'ARNm entraîne la formation d'une structure terminatrice sur l'ARNm cible ⁴³⁵.

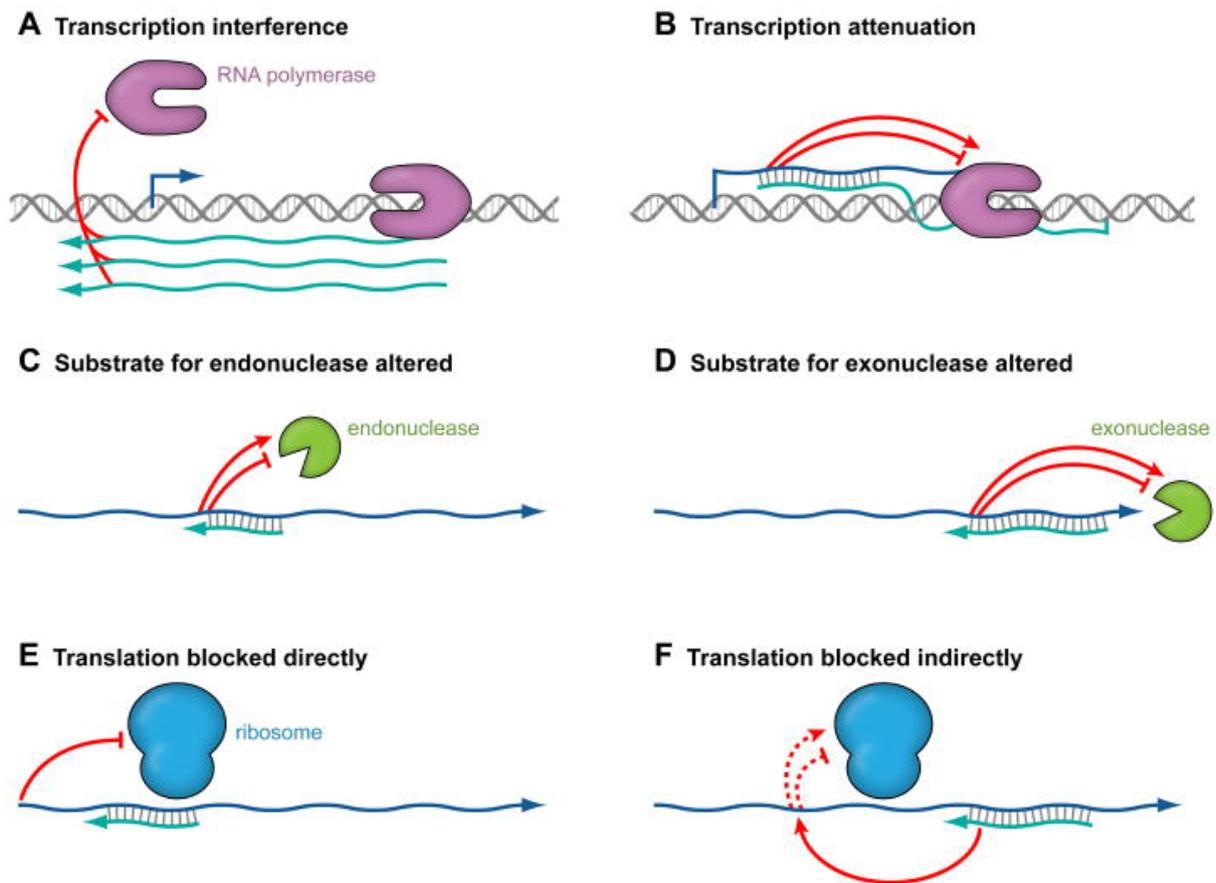


Figure 21 – Mécanismes d'action des ARNs. ⁴²⁵

Les ARNs peuvent également impacter la stabilité de l'ARN cible en induisant ou en inhibant le clivage par des ribonucléases (RNAses) (**Figure 21C et D**). Les mécanismes de maturation et de dégradation des ARN qui ont été étudiés en premier chez *E. coli* puis plus tard chez *B. subtilis* font offices de paradigmes pour les entérobactéries et les Firmicutes respectivement ⁴³⁶. Les études de ces mécanismes ont ensuite été étendues à d'autres bactéries dont des pathogènes majeurs. Une quarantaine de RNAses bactériennes ont été identifiées jusqu'à présent, presque toutes ont été décrites en premier chez *B. subtilis* ou *E. coli* et seulement 9 de ces 40 RNAses sont présentes à la fois chez *E. coli* et *B. subtilis* ⁴³⁶. La dégradation des ARNm peut être influencée par divers facteurs dont la présence de structures secondaires spécifique ou son taux de traduction. Les deux endoribonucléases majeures qui sont liées au clivage des ARN chez *E. coli* et *B. subtilis* sont la RNase E et la RNase Y, respectivement ^{437,438}. Elles sont associées à la membrane bactérienne et peu spécifiques, clivant les ARN au niveau de régions riches en AU. Contrairement à la RNase E qui est essentielle chez *E. coli*, la RNase Y n'est pas essentielle chez *B. subtilis* bien que sa déplétion entraîne de nombreux phénotypes ⁴³⁶. Ces deux RNAses forment des complexes

multiprotéiques affectant globalement la stabilité des ARNm appelés dégradosomes. Bien que les RNAses E et Y soient les deux principaux acteurs dans l'initiation de la dégradation des ARNm chez *E. coli* et *B. subtilis*, d'autres endoribonucléases peuvent également intervenir. Par exemple, *E. coli* possède un paralogue non essentiel du domaine catalytique de la RNase E appelée RNase G qui présente un mécanisme de clivage similaire par ciblage des régions riches en AU des ARN monophosphorylés en 5'. Plusieurs 3'-exoribonucléases (PNPase, RNaseR, RNase PH) et endoribonucléases (RNase III, RNase P, RNase Z) sont conservées entre *E. coli* et *B. subtilis*⁴³⁸. Parmi celles-ci, la RNase III, généralement associée à la maturation des ARNr, a la capacité de cliver l'ARN double brin ce qui lui permet de cibler les duplex formés par l'ARNas et sa cible. Par exemple, la RNase III clive les deux ARNnc RyeA et RyeB codé chacun sur des brins opposés chez *E. coli*⁴³⁹. Très peu d'études ont analysé le rôle des exonucléases dans la dégradation du complexe ARNas/ARNm, mais l'on peut supposer que l'appariement des deux ARN pourrait bloquer l'action des exonucléases ou créer une structure sur l'ARN cible le rendant plus ou moins susceptible d'être clivé. Les ARNas complémentaires de la région 5'UTR peuvent agir sur la traduction en s'appariant directement avec la séquence SD (Shine-Dalgarno) de leurs cibles ARNm empêchant ainsi la fixation des ribosomes (**Figure 21E**). Par exemple, l'ARNas SymR d'*E. coli* chevauche la région 5'UTR de l'ARNm de *symE* inhibant ainsi sa traduction⁴⁴⁰. Un exemple récent et original de régulation post-transcriptionnelle en *cis* a été décrit pour l'ARNm *icaR* (codant pour un répresseur transcriptionnel d'enzymes responsables de la synthèse du PIA [Polysaccharide Intercellular Adhesin]) chez *S. aureus* possédant une longue extrémité 3'UTR (390 pb) qui s'apparie avec l'extrémité 5'UTR masquant la séquence SD et empêchant ainsi sa traduction⁴⁴¹. Les mécanismes permettant d'induire la traduction d'*icaR* ne sont pas connus mais il est possible que des signaux environnementaux soient responsables de la stabilisation de l'interaction entre les deux extrémités. Les ARNas peuvent également agir indirectement sur la traduction en induisant des modifications de structures du RBS empêchant ainsi la fixation du ribosome (**Figure 21F**).

La modulation de la transcription, de la stabilité ou de la traduction d'un gène codé en *cis* n'est pas la seule fonction des ARNas. Certains ARNas sont également connus pour

coder une petite protéine sur leur séquence. Il existe également certains ARNnc qui peuvent agir à la fois en *cis* et en *trans*. Il a été montré que l'ARNnc GadY d'*E. coli* se fixe à la protéine Hfq qui est requise pour l'action des ARNnc agissant en *trans*, suggérant que GadY ait la possibilité de s'apparier avec d'autres cibles ARNm⁴³⁰. De la même manière, l'ARN RliE de *L. monocytogenes*⁴⁴² et les ARN ASdes et Aspks de *M. tuberculosis*⁴⁴³ ont des cibles prédites autres que le transcrite codé par le gène opposé à l'ARNs. Deux mécanismes généraux d'appariement entre un ARNas et sa cible sont proposés. Dans le premier mécanisme, dit à une étape, le contact en un point entre l'ARNs et son ARNm conduit à la formation d'un duplex complet entre ces deux molécules. Dans le second mécanisme, dit à étapes multiples, l'appariement transitoire entre les deux ARN nécessite la stabilisation du complexe par une protéine ou l'ajout de nouvelles paires de base. Une fois stabilisé, le complexe peut alors former un duplex complet entre l'ARNs et son ARNm cible. Cependant, ce processus est lent et ne semble pas requis pour la régulation dans les quelques cas décrits. Dans de nombreux cas d'appariement par complémentarité de bases, des *tiges-boucles* contenant un motif *U-turn* YUNR (motif contenant une pyrimidine [Y] suivie d'un uracile [U] puis de n'importe quel nucléotide [N] et d'une purine) sont requis pour les interactions ARN-ARN⁴⁴⁴.

b. ARNnc codés en trans

Les ARNnc agissant en *trans* constituent un autre groupe d'ARN régulateurs et, contrairement aux ARNnc agissant en *cis*, ne présentent qu'une complémentarité de séquence partielle avec leur cibles ARNm. Ils sont généralement localisés dans les régions intergéniques, à distance de la séquence de leur ARNm cible avec lequel ils vont s'apparier afin d'agir sur sa stabilité et/ou sa traduction. Les ARNnc agissant en *trans* sont également retrouvés dans les régions 5' et 3' UTR des ARNm. La production des ANRnc à partir des régions 3' UTR peut se faire selon deux mécanismes : (1) à partir d'un promoteur ou (2) par maturation de l'ARNm *via* des RNAses⁴⁴⁵. La plupart des ARNnc agissant en *trans* ne sont synthétisés que dans des conditions de croissance très spécifiques et sont souvent impliqués dans la réponse rapide aux changements environnementaux. Chez *E. coli* par exemple, ces ARN régulateurs sont induits dans des conditions de carence en Fer (Fur réprimé par RyhB)

⁴⁴⁶, de stress oxydatif (OxyR activé par OxyS) ³⁶⁵, lors d'un stress membranaire (σ^E induit par MicA et RybB) ⁴⁴⁷, d'un taux élevé de glycine (GcvA induit par GcvB) ⁴⁴⁸, d'un changement de concentration de glucose (CRP réprimé par Spot42 et activé par CyaR) ^{390,449}, ou un taux de glucose-phosphate élevé (SgrR activé par SgrS) ³⁹⁰.

La majorité des régulations décrites pour les ARNnc agissant en *trans* sont de type répressives et l'appariement avec leur cible conduit le plus souvent à la répression de la production de la protéine par inhibition de la traduction et/ou dégradation de l'ARNm ³⁷⁰. Pour cela, ils se fixent généralement à l'extrémité 5'UTR de l'ARNm masquant ainsi le site de fixation du ribosome. De manière moins fréquente, certains se fixent loin en amont de l'AUG et agissent sur la stabilité de l'ARNm. Par exemple, l'ARN RydC de *Salmonella* cible l'ARNm de *cfa* (*cyclopropane fatty acid synthase*) loin en amont du RBS et, avec l'aide de Hfq, interfère avec le clivage par la RNase E ⁴⁵⁰. Ou encore l'ARNnc RyhB qui réprime la traduction du l'ARNm de *sodB* en se fixant au niveau du RBS et entraîne sa dégradation avec Hfq par la RNase ³⁸⁴. L'inhibition de la fixation du ribosome par l'interaction de l'ARNnc avec la séquence SD reste cependant le principal mode d'action des ARNnc agissant en *trans*, et la dégradation consécutive de l'ARNm contribuerait à augmenter l'effet répresseur et rendant également la répression irréversible ⁴⁵¹. Par exemple, plusieurs ARNm chez *S. aureus* sont ciblés par des ARNnc qui se fixent au niveau de la séquence SD et créant par la même occasion un site de clivage pour la RNase III ⁴⁵². Certains ARNnc agissant en *trans* ont également la capacité d'activer l'expression de leur cible *via* un mécanisme antisens par lequel l'appariement ARNnc/ARNm permettrait de dénaturer une structure secondaire inhibant la fixation du ribosome ^{365,370}.

Il n'y a pas de corrélation particulière entre la localisation chromosomique d'un ARNnc agissant en *trans* et celle de sa cible. De plus, chaque ARNnc agissant en *trans* a la capacité de s'apparier avec de multiples cibles ^{365,370}. Cette multiplicité de cibles résulte de la complémentarité imparfaite entre un ARNnc agissant en *trans* et sa cible, offrant ainsi plus de possibilités d'interactions. La taille de la région d'interaction entre un ARNnc et sa cible ARNm est d'environ 10 à 25 nucléotides, mais il est systématiquement observé que seule une petite partie de cette région semble critique pour la régulation. Par exemple, bien que

l'ARNnc SgrS ait la possibilité de s'apparier sur 23 paires de bases avec l'ARNm *ptsG* sur une région de 32 nucléotides, seulement quatre mutations simples dans la séquence de SgrS affecte significativement son action sur *ptsG*⁴⁵³. Le fait qu'un ARNnc puisse réguler plusieurs cibles différentes implique qu'un même ARNnc peut être au centre de la régulation d'une réponse physiologique donnée, de la même manière qu'un facteur de transcription, mais au niveau post-transcriptionnel. Les voies de régulation bien caractérisées impliquant ces ARNnc sont la répression des cluster d'enzyme ferriques dans des conditions de carence de fer, la répression des porines de la membrane externe lors de stress membranaire et la répression du QS dans des conditions de faible densité cellulaire⁴⁵⁴⁻⁴⁵⁶.

Dans de nombreux cas, l'intervention de la protéine chaperonne à l'ARN Hfq est requise afin de médier l'action des ARNnc agissant en *trans*. Elle permet de faciliter les interactions ARN-ARN entre l'ARNnc et sa cible ARNm en raison de la complémentarité de séquence limitée entre ces deux ARN⁴⁵⁷⁻⁴⁶⁰. La protéine Hfq est une protéine homologue des protéines Sm et Sm-like impliquées dans l'épissage et la dégradation des ARNm chez les eucaryotes. Elle est structurée en anneau constitué de six protomères identiques (α - β_{1-5}) d'environ 65 résidus chacun et présentant deux motifs de séquences nommés Sm1 et Sm2 (faisant 32 et 14 acides aminés de longs respectivement) (**Figure 22A-C**)⁴⁶¹. Le motif Sm1 englobe les trois premiers brins β (β_{1-3}) de la structure commune α - β_{1-5} et est hautement conservé entre les membres de cette famille⁴⁶⁰. Le motif Sm2 est formé par les deux derniers brins β (β_{4-5}) et diffère entre la protéine Hfq de bactéries d'une part, et entre la protéine Hfq des archées et les protéines eucaryotes Sm et Sm-Like d'autre part. L'architecture en anneau de Hfq présente deux faces pouvant interagir avec des acides nucléiques : la face distale qui présente l'hélice α N-terminale et la face proximale (**Figure 22D**). L'appariement d'une molécule ARN avec Hfq dépend d'une information structurale codée dans la molécule d'ARN. Récemment, les ARNnc interagissant avec Hfq ont été classés en deux groupes, en fonction de la face avec laquelle ils interagissent⁴⁶². Les ARNnc de classe 1 se fixent à la face proximale de la protéine Hfq qui reconnaît spécifiquement une région riche en U. Les ARN de cette classe peuvent également interagir avec le bord de l'anneau de la protéine Hfq permettant l'exposition de la région d'interaction de certains ARNnc avec leur cibles ARNm⁴⁶³. Les ARN

de classe 2 peuvent également interagir avec la face proximale mais vont s'enrouler autour de l'hexamère de Hfq afin d'interagir avec la face distale qui interagit principalement avec les séquences riches en AU⁴⁶⁴. Ainsi, la protéine Hfq interagit principalement avec les extrémités 3' poly(U) des ARNnc et les sites de fixations sont le plus souvent localisés en aval d'un terminateur Rho indépendant⁴⁶⁵. L'absence de cette caractéristique chez les ARNnc dérivant des régions 5'UTR suggèrent que ces molécules sont probablement Hfq indépendantes. Cela pourrait être la raison pour laquelle si peu d'ARNnc agissant en *trans* et dérivant de la région 5'UTR sont identifiés dans les études exploratoires de nouveaux ARNnc³⁷⁷. La protéine Hfq est capable d'interagir avec deux molécules d'ARN simultanément, permettant ainsi la formation du duplex ARNnc-ARNm.

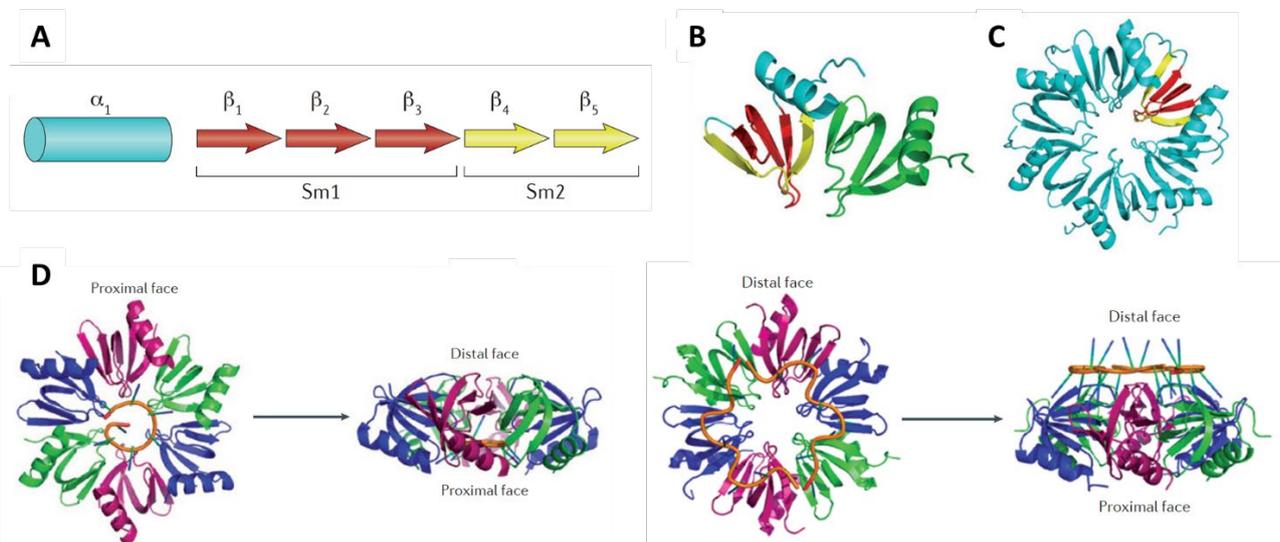


Figure 22 – Structure de la protéine Hfq et modes d'interaction avec l'ARN.⁴⁶⁰

Elle peut agir sur la stabilité ou la traduction des ARN avec lesquels elle interagit, et ce selon plusieurs mécanismes (**Figure 23**)^{457,459}. D'une part, Hfq peut réprimer la traduction d'une protéine en assistant un ARNnc à se fixer sur la région 5'UTR d'un ARNm cible qui lui est complémentaire, rendant ainsi inaccessible le site de fixation du ribosome (**Figure 23A**). À l'inverse, Hfq peut induire la traduction en guidant un ARNnc complémentaire de la région 5' de son ARNm cible afin de dénaturer une structure secondaire inhibant la fixation du ribosome (**Figure 23B**). En amont de l'appariement avec l'ARNm, Hfq joue un rôle stabilisateur de l'ARNnc agissant en *trans* en le protégeant du clivage par les ribonucléases (**Figure 23C**) et, une fois que l'ARNnc est apparié avec sa cible, elle induit dans la plupart des cas la dégradation de l'ARNm par la RNase E et peut également recruter d'autres

éléments de la dégradation des ARN et composants du dégradosome (**Figure 23D**)⁴⁵⁷. Cependant, Hfq n'a pas seulement un rôle d'entremetteur mais peut également intervenir comme régulateur important de l'expression des gènes. En effet, Hfq peut interagir avec l'ARNm et agir sur sa traduction. Elle peut par exemple se fixer à l'extrémité 3' d'un ARNm en la rendant accessible pour une polyadénylation favorisant ainsi sa dégradation par exonucléases (**Figure 23E**), ou encore se fixer au niveau de la région 5'UTR inhibant ainsi l'initiation de la formation de la machinerie traductionnelle (par exemple pour l'ARNm de *cirA* chez *E. coli*⁴⁶⁶). En général, de longues séquences d'appariement, comme c'est le cas pour les ARNas, et/ou une forte concentration d'ARNnc peuvent dispenser de l'action de la protéine chaperonne Hfq. Hfq est requises chez les bactéries à Gram négatif dans de nombreux processus de *fitness* et de virulence chez de nombreux pathogènes et des mutants ne produisant plus la protéine Hfq sont très souvent sensibles aux mécanismes de défense de l'hôte et leur virulence est fortement atténuée dans des modèles animaux^{467,468}. La virulence de mutants Δhfq de pathogènes à Gram négatif tels que *Brucella abortus*, *Salmonella spp.* et *V. cholerae* est très atténuée dans des modèles souris. Des études sur modèles souris ou rat ont également montré l'importance de la protéine Hfq pour la virulence des UPEC (*Uropathogenic E. coli*), *Francisella tularensis*, *Neisseria meningitidis*, *P. aeruginosa*, et *Y. pestis*⁴⁶⁷. Cependant les mutants Δhfq chez *S. aureus* ne semblent présenter aucun phénotype, du moins dans le modèle d'infections de *Caenorhabditis elegans*⁴⁶⁷. Les mutants Δhfq chez les bactéries à Gram négatif présentent également une plus grande sensibilité à certains stress environnementaux comme le stress oxydatif ou un pH faible. Encore une fois des mutants Δhfq de *S. aureus* dans trois souches différentes ne présentent aucun phénotype détectable pour plusieurs stress testés⁴⁶⁹. Ainsi, bien que le rôle de la protéine Hfq dans l'adaptation aux stress et la virulence semble indéniable chez un grand nombre de bactéries à Gram négatif, son importance dans ces processus doit encore être établie dans les bactéries à Gram positif comme *L. monocytogenes* et *S. aureus* qui sont capables de produire cette protéine fonctionnelle. On peut par exemple citer le cas de l'ARNnc LhrC (*Listeria Hfq-binding RNA C*) identifié chez *L. monocytogenes* durant une expérience de co-immunoprécipitation avec la protéine Hfq pour lequel il a été montré que la protéine Hfq n'est pas requise pour sa stabilité ou ses interactions avec ses cibles ARNm

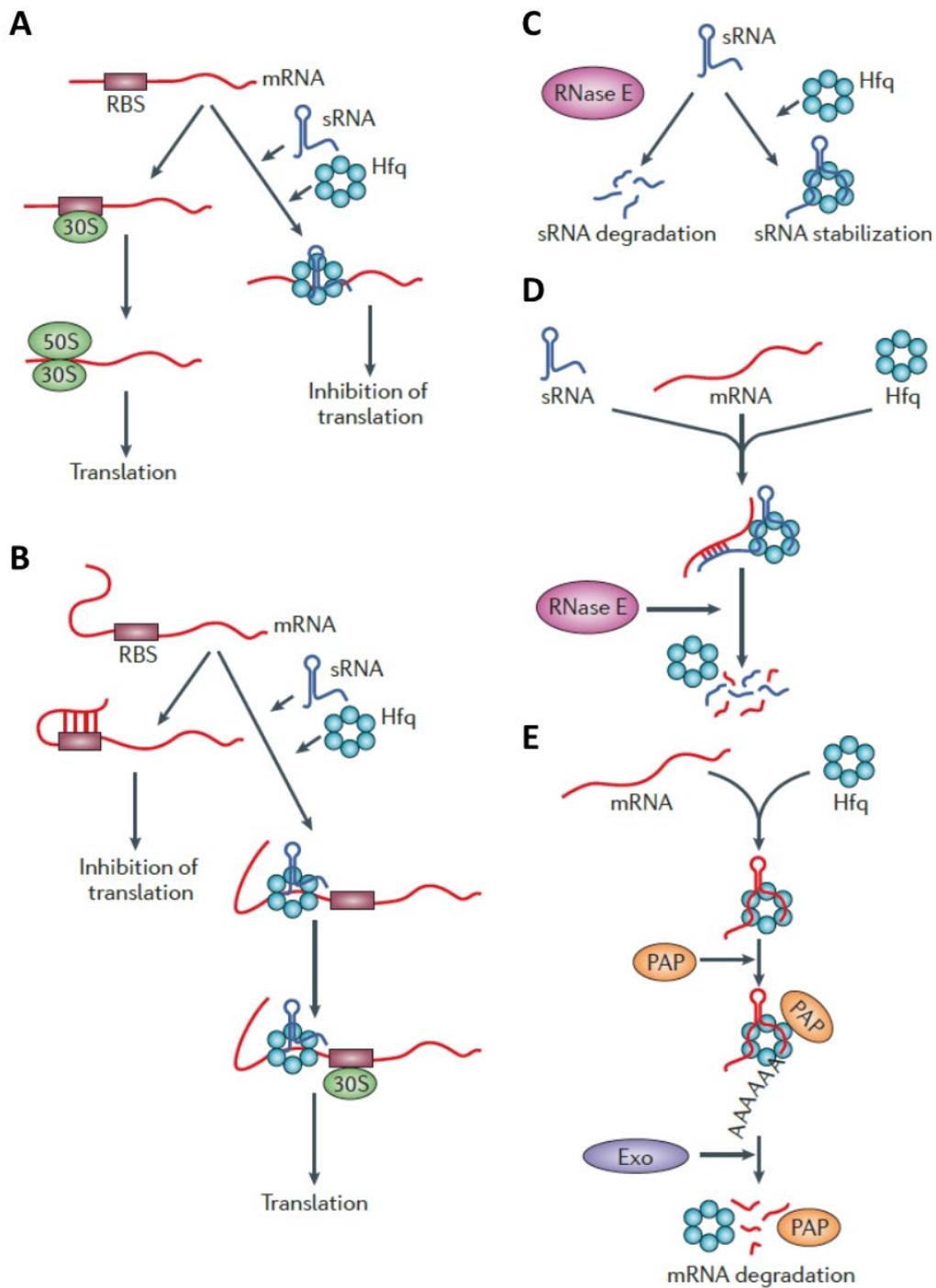


Figure 23 – Modes d'action de la protéine Hfq.⁴⁶⁰

2. Les ARNnc senseurs : thermosenseurs, T-boxes et *riboswitch*

La région de l'ARNm située entre le +1 de transcription et le codon START de traduction est appelée région 5' non traduite (5'UTR). La taille de cette région est variable, pouvant aller de quelques nucléotides à plusieurs centaines, et plusieurs promoteurs

peuvent initier sa transcription donnant lieu à de complexes régulations post-transcriptionnelles ^{471,472}. Chez de nombreux organismes, ces régions sont le siège d'éléments régulateurs de l'expression des gènes en aval qui sont capables de reconnaître différents signaux environnementaux ³⁷⁹.

Parmi ces éléments régulateurs se trouve les ARN thermosensibles généralement situés dans la région 5'UTR d'ARNm codant pour des protéines de choc thermique et de facteurs de virulence ⁴⁷³. D'une manière générale, à des températures basses, la séquence Shine et Dalgarno (SD) de l'ARNm est masquée par une structure en *épingle à cheveux* qui va se déstructurer au fur et à mesure que la température augmente, permettant ainsi une régulation fine et thermo-dépendante de la traduction. L'exemple le mieux étudié est sans doute celui du thermosenseur régulant l'expression du facteur de transcription *prfA* qui active l'expression de la plupart des gènes de virulence chez *L. monocytogenes* à des températures proches de celle du corps humain (37°C) ^{474,475}. Des thermosenseurs de fonction similaire ont été identifiés chez d'autres bactéries pathogènes comme *Yersinia pestis* et chez les protéobactéries.

Les *riboswitch* et les *T-boxes* sont deux autres éléments de régulation senseurs situés dans la région 5'UTR des ARNm qui présentent des domaines hautement structurés pouvant se lier à des métabolites comme des cofacteurs, des vitamines, des acides aminés, des nucléotides, le second messager c-di-GMP, des ions Mg²⁺ ou des ARNt non chargés ^{476,477}. La fixation de ces métabolites induit un changement de structure de la région 5'UTR ce qui a pour conséquence la modification de l'expression de la région avale de l'ARNm en induisant par exemple un arrêt prématuré de la transcription, l'activation/répression de la traduction ou le clivage de l'ARNm ^{370,379}. Les ARN *T-box* ont la capacité de détecter l'état d'aminocyclation d'un ARNt et de contrôler la transcription des gènes en aval codant pour des protéines impliquées dans la synthèse et le transport des acides aminés, ou l'aminocyclation des ARNt. Les *riboswitch*, découvert plus récemment, sont des éléments hautement structurés pouvant détecter la présence de certains métabolites et agissant ainsi comme senseurs et régulateurs de nombreuses voies métaboliques chez les bactéries ⁴⁷⁸⁻⁴⁸¹. Plus d'une vingtaine de classes de *riboswitch* ont déjà été identifiées, et la plupart de leurs

mécanismes d'action et de leurs stratégies de reconnaissance moléculaire ont été décrits à ce jour ⁴⁸². Chacune de ces classes fixe un métabolite spécifique *via* un domaine particulier appelé aptamère, ce qui induit un changement structurel d'un second domaine du *riboswitch* responsable de l'altération de la transcription ou de la traduction (**Figure 24**) ³⁷⁰. Pour les *riboswitch* agissant au niveau transcriptionnel, la fixation du ligand à l'aptamère induit généralement la formation d'une structure de terminateur (en *tige-boucle*) dans le domaine régulateur, empêchant ainsi la transcription du gène en aval. L'effet inverse est également possible avec une dénaturation d'une structure secondaire après la fixation d'un ligand. Quant à ceux agissant sur la traduction, la fixation du métabolite sur le domaine aptamère interfère avec l'interaction entre les séquences SD et anti-SD qui, selon les cas, inhibe ou permet la fixation du ribosome au RBS (*Ribosome Binding Site*). Selon les types et les contextes génétiques, les *riboswitch* peuvent agir comme inducteur ou répresseur de l'expression d'un gène en réponse à la fixation d'un ligand. En général, la plupart des *riboswitch* réprime la transcription ou la traduction en présence du ligand ; parmi les rares *riboswitch* activant l'expression des gènes, les *riboswitch* de type II dépendant du c-di-GMP ont été caractérisés et certains ont été identifiés chez *C. difficile* ⁴⁸³. En raison de la nature modulaire des *riboswitch*, un même aptamère peut être responsable de différentes voies de signalisation ou opérer selon différents mécanismes d'action en fonction du contexte dans lequel il se trouve ⁴⁷⁹. Par exemple, le *riboswitch* fixant la cobalamine (ou vitamine B12) utilise un mécanisme de terminaison de la transcription pour le gène *btuB* chez les bactéries à Gram positif et agit sur l'initiation de la traduction de l'opéron *cob* chez les bactéries à Gram négatif. En général, les *riboswitch* chez les bactéries à Gram positif agissent sur la terminaison précoce de la transcription tandis que chez les bactéries à Gram négatif ils régulent le plus souvent l'initiation de la traduction ⁴⁷⁹. Si les *riboswitch* sont en général des éléments localisés dans les régions 5'UTR des ARNm, certains peuvent être également situés dans la région 3'UTR et agissent alors comme terminateur de transcription du gène en amont lorsqu'ils sont liés à leur ligand. Un exemple de ce type de régulateur est le *riboswitch* de la lysine qui régule l'expression du gène aval *Imo0798* codant un transporteur de la lysine chez *L. monocytogenes* ⁴⁸⁴.

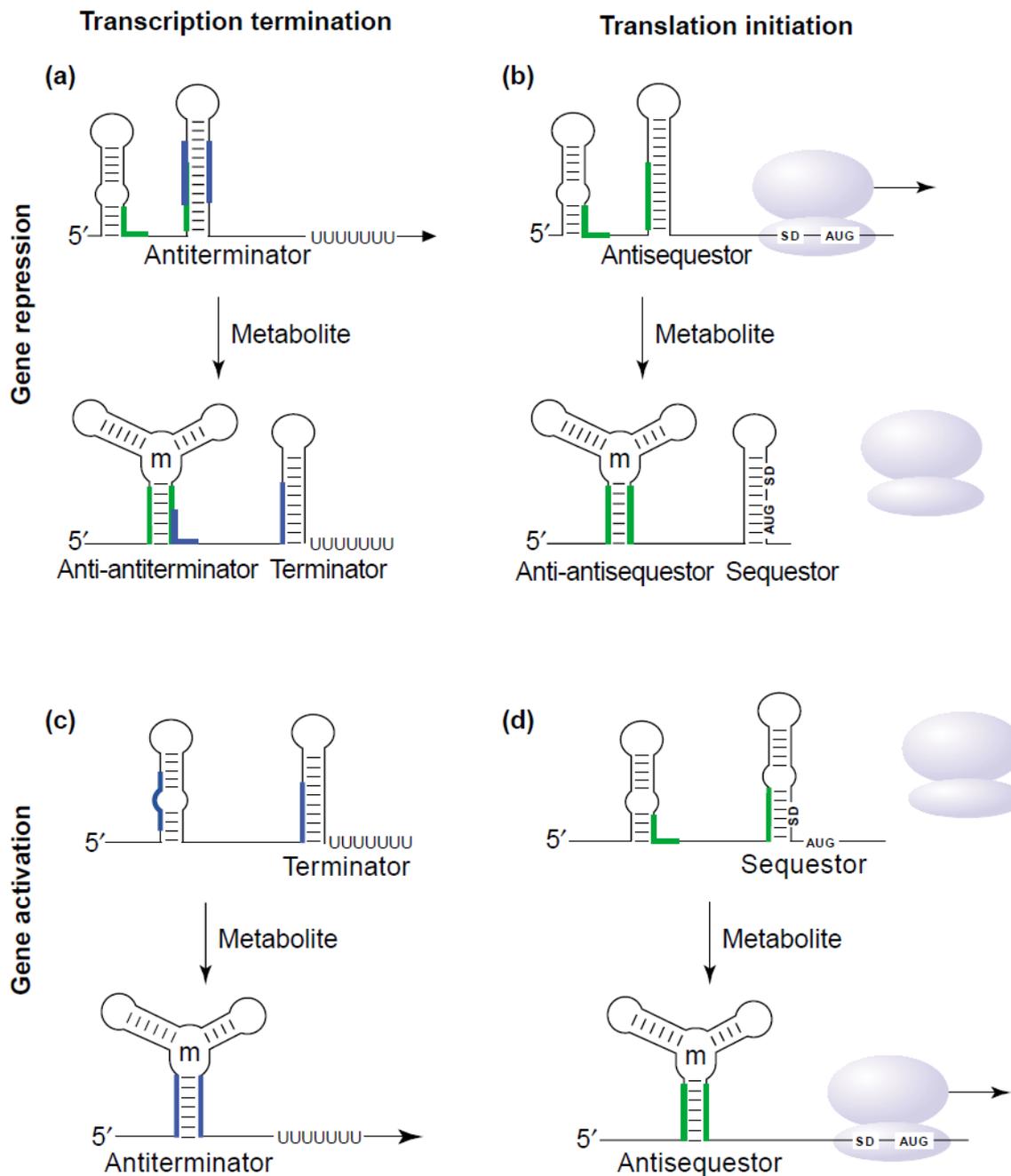


Figure 24 – Mécanismes de contrôle de l'expression des gènes par *riboswitch*. Les *riboswitch* activent ou répriment l'expression des gènes en agissant sur la structure secondaire de l'ARNm qu'ils contrôlent (représenté par la flèche noire débutant en 5'). En réponse à un changement de concentration du métabolite qu'ils fixent, ils peuvent agir sur la terminaison de la transcription (A et C) ou sur l'initiation de la traduction (B et D).⁴⁷⁹

3. Les ARNnc interférant avec l'activité d'une protéine

Parmi les ARNnc se liant à des protéines, trois ont une activité intrinsèque (RNase P) ou contribuent aux fonctions essentielles de la particule ribonucléoprotéique (4.5S et ARNtm). D'autre part, trois autres ARNnc (CsrB, 6S, and GImY) se fixent à des protéines agissant selon un mode d'action antagoniste des protéines auxquelles ils se lient en mimant la structure d'autres acides nucléiques.

Les ARNnc CsrB et CsrC d'*E. coli* modulent l'activité de la protéine CsrA, dont le rôle est de réguler l'utilisation du carbone et la mobilité bactérienne lors de l'entrée en phase stationnaire et autres conditions de carence nutritionnelle ⁴⁸⁵. Les dimères de la protéine CsrA se fixent aux motifs GGA de la région 5'UTR de l'ARNm cible affectant ainsi leur stabilité et/ou leur traduction. Les ARNnc CsrB et CsrC contiennent chacun de multiples sites GGA (22 et 13 respectivement) qui vont pouvoir être reconnus par la protéine CsrA. Ainsi, quand le niveau des ARN CsrB et CsrC augmente, les deux ARNnc vont séquestrer la protéine CsrA l'empêchant d'agir sur ses ARNm cibles. La transcription des gènes *csrB* et *csrC* est induite *via* le régulateur à deux composants BarA-UvrB lorsque les cellules sont en carence nutritionnelle (le signal de cette induction n'est cependant pas encore caractérisé). Le niveau des deux ARN CsrB et CsrC est cependant régulé et stabilisé par la protéine CsrD qui fixe le c-di-GMP et recrute la RNase E pour la dégradation des petits ARN ⁴⁸⁶. Des homologues de CsrB et CsrC ont été identifiés dans d'autres bactéries comme impactant le métabolisme secondaire, le QS et la pathogénèse ^{487,488}.

L'ARN 6S d'*E. coli* mime un promoteur ouvert afin de fixer et séquestrer l'holoenzyme σ^{70} de l'ARN polymérase ⁴⁸⁹. Quand ce petit ARN est abondant, particulièrement lors de la phase stationnaire, il est capable de se lier au facteur σ^{70} (forme de la phase de croissance) mais pas à σ^S (forme de la phase stationnaire) ⁴⁹⁰. Ainsi, l'interaction entre l'ARN 6S et l'holoenzyme σ^{70} inhibe la transcription de certains promoteurs régulés par σ^{70} et augmente la transcription de certains promoteurs régulés par σ^S . L'ARN 6S est produit à partir d'un long transcrit mûri et s'accumule lors de la phase stationnaire mais les détails de sa régulation ne sont pas encore compris ⁴⁸⁹. Il existe de nombreux homologues de cet ARN 6S dont deux chez *B. subtilis* dont le rôle est encore inconnu ⁴⁹⁰.

Il a été également proposé pour l'ARNnc GlmY un mode de régulation par séquestration d'une protéine ciblant l'ARN GlmZ pour le dégrader ³⁹⁰. GlmY et GlmZ induisent par des mécanismes distincts l'accumulation de la glucosamine-6-phosphate synthase GlmS. L'appariement de la forme entière de l'ARNnc GlmZ avec l'ARNm de *glmS* permet d'activer sa traduction tandis que GlmY, bien que fortement homologue en séquence et structure à GlmZ, ne présente pas la région complémentaire de l'ARNm de *glmS* et ne

peut donc pas réguler directement son expression. En revanche, GlmY inhibe les événements de clivage de GlmZ qui le rendent incapable d'activer la traduction de *glmS*. Bien que cela doive encore être démontré, GlmY semble stabiliser la forme longue de GlmZ en entrant en compétition avec lui lors de la fixation à la protéine YhbJ qui cible GlmZ pour le cliver. GlmY est également clivé et son niveau d'expression est régulé par polyadénylation^{396,491}.

4. Le système CRISPR-Cas

Les ARNnc CRISPR (*Clustered Regularly Interspaced Short Palindromic Repeats*) et les gènes qui leurs sont associés (*cas*) constituent un mécanisme de défense unique conférant une réponse rapide et une adaptation efficace contre les infections de bactériophages et la conjugaison et la transformation de plasmides^{492,493}. Certaines études suggèrent que les systèmes CRISPR-Cas peuvent également être impliqués dans d'autres processus que des processus de défenses, comme la régulation du comportement de groupe, la réponse adaptative et la pathogénicité⁴⁹⁴⁻⁴⁹⁶. Les systèmes de défense CRISPR-Cas présentent des caractéristiques qui leurs sont propres, mais partagent également certaines similitudes avec les petits ARN interférents eucaryotes (siRNA). Ces régions sont présentes dans environ 50% des génomes bactériens et 90% des génomes des archées⁴⁹². Elles sont composées d'une séquence *leader* qui est suivie d'une série de blocs de séquences répétées (entre 2 et 249 fois), identiques (allant de 24 à 47 pb) et partiellement palindromiques séparées de manière régulière par des séquences variables courtes appelées *spacers*, complémentaire de séquences virales ou plasmidiques (**Figure 25A**)^{492,493}. Les régions CRISPR sont adjacentes de régions contenant des gènes associés au ARNs CRISPR, ou gènes *cas* (*CRISPR-associated*), codant pour des endonucléases qui s'associent avec les crRNA afin de former des modules CRISPR-Cas impliqués dans les différentes étapes du système de défense CRISPR³²⁷. Les différentes protéines Cas peuvent être divisées en trois modules fonctionnels principaux (**Figure 26**) : un module adaptatif impliqué dans l'acquisition des séquence *spacers*, un module d'expression assurant la maturation des pré-crRNA en crRNA et leur fixation à leur cible et un module d'interférence responsable du clivage de la cible du crRNA. Un module impliqué dans la transduction du signal et un module auxiliaire ont également été décrits mais restent encore mal caractérisés^{328,497}.

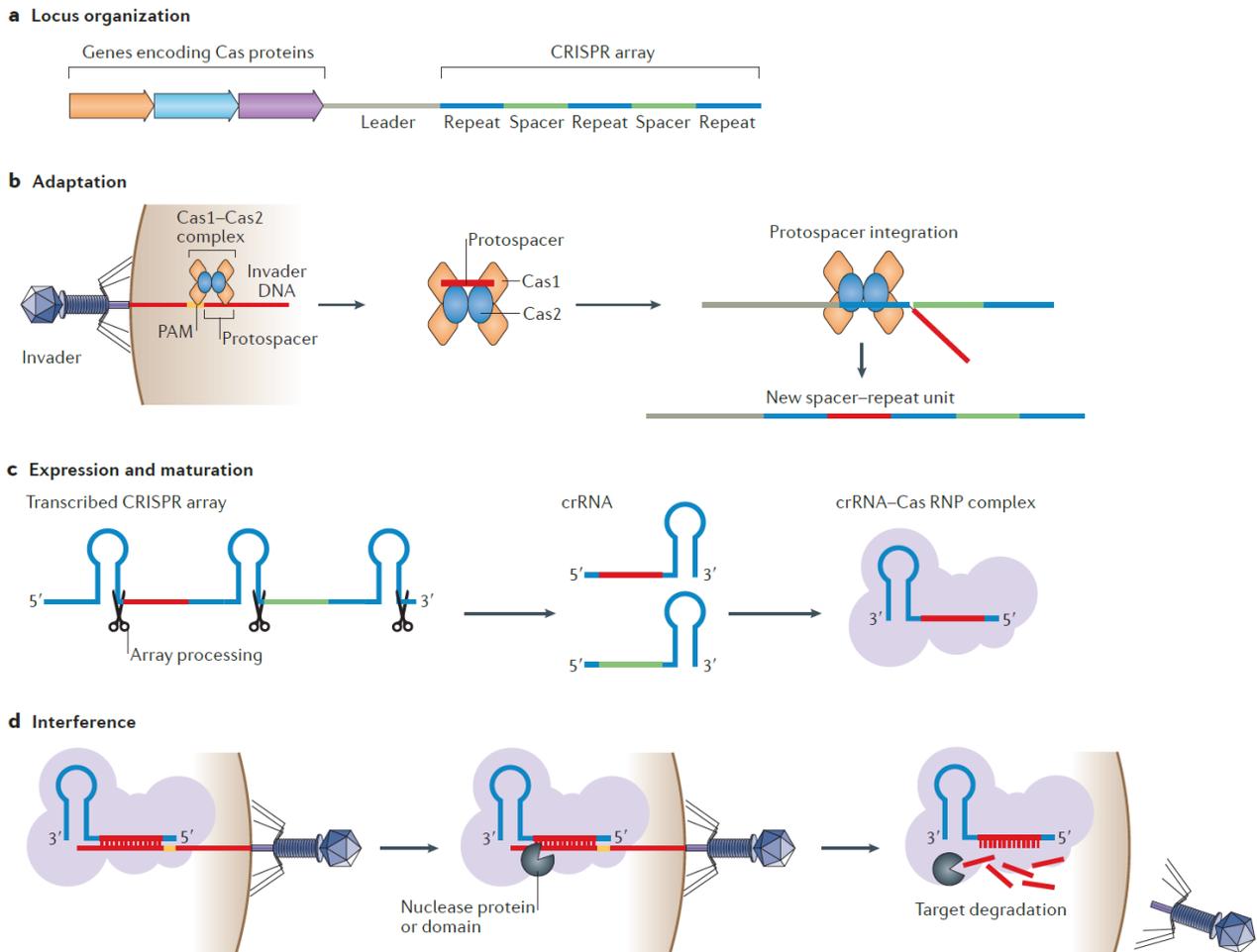


Figure 25 – (A) Organisation du *locus CRISPR-cas* et (B-D) description des trois étapes de l’immunité bactérienne médiée par le système CRISPR-Cas. ³²⁷

Le mode d’action du système de défense CRISPR-Cas peut être décomposé en trois temps : adaptation, expression/maturation et interférence (**Figure 25B-D**) ^{327,498–501}. Durant la phase d’adaptation, les protéines Cas1 et Cas2 qui sont présentes dans tous les systèmes CRISPR-Cas et varient très peu entre les différents organismes permettent l’acquisition de séquences *spacers* dans les régions CRISPR à partir de séquences d’ADN ou d’ARN exogènes (**Figure 25B**). Il a été également proposé un rôle au sein des modules d’adaptation pour la protéine Cas4 présente dans certains système et qui contient un domaine exonucléase RecB permettant de générer des fragment d’ADN sb ³²⁷. L’acquisition des séquences *spacers* dans de nombreux systèmes CRISPR-Cas requiert la reconnaissance d’un court motif adjacent de la cible appelé PAM (*Protospacer Adjacent Motif*) ⁵⁰². Lors de la phase d’expression et de maturation, la région CRISPR est transcrite dans son intégralité en un ARN précurseur (pre-crRNA) qui sera par la suite digéré en plus petites unités appelées ARN CRISPR (crRNA) qui contiennent chacune un *spacer* unique flanqué par une partie de la séquence répétée. Le

pre-crRNA sont maturés en crRNA soit par une sous-unité endonucléase du module effecteur ⁵⁰³, soit par un mécanisme alternatif impliquant la RNase III et un tracrRNA ⁵⁰⁴. Les crRNA sont associés avec une ou plusieurs protéines Cas formant ainsi le complexe actif Cas-crRNA (**Figure 25C**). Enfin, dans l'étape d'interférence le complexe Cas-crRNA va permettre de cibler un acide nucléique étranger possédant la séquence complémentaire du *spacer* contenu sur le crRNA. Lorsque la séquence est reconnue, l'acide nucléique étranger va être clivé puis dégradé (**Figure 25D**).

L'évolution rapide de la plupart des gènes *cas* et la remarquable variabilité de l'architecture génomique des *loci* CRISPR-*cas* font de l'annotation et la classification des protéines Cas un réel défi, qui requiert des mises à jours régulières ^{329,497}. La classification des différents systèmes CRISPR-Cas en fonction de la structure et de la fonction des protéines Cas associées à leurs modules fonctionnels permet de les catégoriser en deux classes (classe 1 et classe 2), elles-mêmes subdivisées en six types (types I à VI) et plus de 20 sous-types (**Figure 26**) ^{328,329}. Les systèmes de la classe 1 englobent les types I et III, très représentés chez les archées mais moins communs chez les bactéries, et le type IV, plus rare et plus rudimentaire, dont les *loci* CRISPR-*cas* ne présentent pas de nucléase effectrice et dans la plupart des cas ne contiennent pas non plus de module d'adaptation ³²⁹. Dans les systèmes de types I et III, les étapes d'expression, de maturation et d'interférence sont exécutées par le complexe multiprotéique Cascade (*CRISPR-associated complex for antiviral defence*) assisté par l'hélicase Cas3 ⁵⁰⁵, le complexe Csm ⁵⁰⁶ ou le complexe Cmr ⁵⁰⁷ pour les types I, III-A et III-B respectivement. Les systèmes de classe 2 présentent une organisation beaucoup plus simple avec un module effecteur constitué d'une seule et grande protéine à domaines multiples comme Cas9 pour les systèmes abondants de type II, et la protéine Cpf1 (et autres protéines similaires) pour les systèmes plus rares de types V et VI ^{328,508}. De plus, la protéine Cas9 requiert l'association avec un ARNnc structuré appelé tracrRNA (*transactivating CRISPR RNA*) pour son activité ³²⁷. Les systèmes de type I, II et V sont capables de reconnaître et de cliver l'ADN, le type VI de l'ARN et le type III à la fois de l'ADN et de l'ARN. Le type de molécule ciblée par le système de type IV reste inconnu ⁵⁰⁹. Certains modules des différents systèmes CRISPR-Cas sont très utilisés comme outils de génie

généétique, et particulièrement ceux basés sur les protéines Cas9 du type II et Cas12a du type V⁵¹⁰. Cependant, la toxicité intrinsèque des protéines Cas9 et Cas12a comme outils dérivés des systèmes CRISPR-Cas pouvant conduire à la mort de certaines cellules a encouragé l'utilisation et le développement de systèmes CRISPR-Cas endogènes de types I et III.

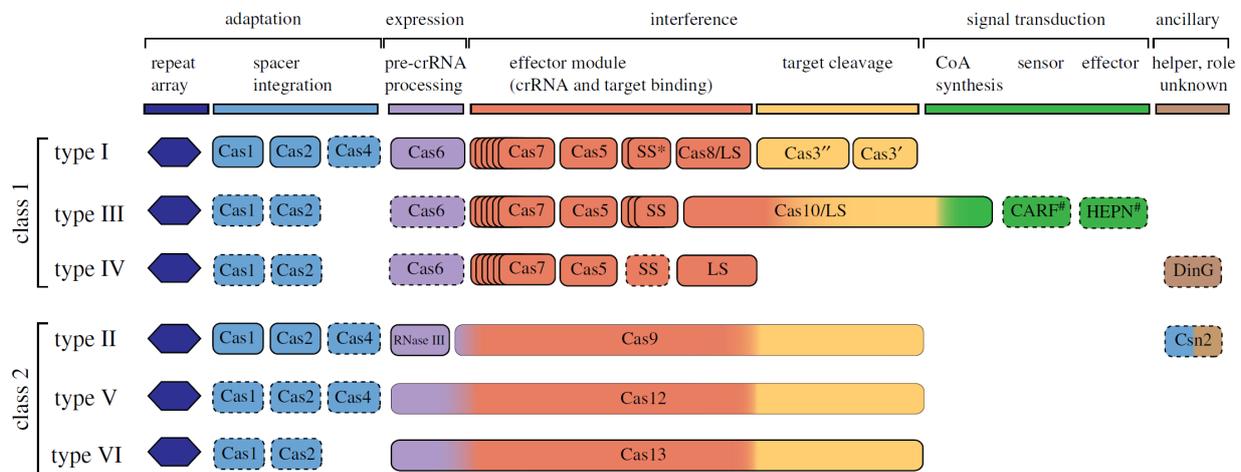


Figure 26 – Organisation modulaire des systèmes CRISPR-Cas. Sont représentés sur cette figure les principaux blocs fonctionnels des différents types de systèmes CRISPR-Cas. L'astérisque (*) indique une petite sous-unité putative (SS ; *Small Subunit*) pouvant être fusionnée à la grande sous-unité (LS ; *Large Subunit*) dans plusieurs sous-types du type I. Le dièse (#) adjacent aux domaines CARF et HEPN indique que d'autres domaines senseurs ou effecteurs inconnus peuvent être impliqués dans la voie de signalisation. Les gènes non essentiels sont indiqués par une ligne en pointillés.⁴⁹⁷

III. Les ARN régulateurs chez les bactéries pathogènes

Les bactéries pathogènes possèdent certaines caractéristiques qui leur permettent de détecter des signaux environnementaux comme des changements dans la disponibilité des nutriments, des modifications du pH, de l'osmolarité ou de la température, et d'adapter leur métabolisme afin de pouvoir passer d'un mode de vie hors de l'hôte à un mode de vie dans leur hôte^{511,512}. De plus, les bactéries pathogènes doivent coordonner l'expression d'un arsenal de facteurs de virulence dont des systèmes de production et de sécrétion de puissantes toxines, et d'autres gènes impliqués dans la pathogenèse afin de pouvoir survivre et proliférer durant l'infection de leur hôte. A cela s'ajoute un coût énergétique de production et de sécrétion élevé rendant ainsi indispensable la régulation de l'expression de ces facteurs par la bactérie. Un autre aspect fondamental de la virulence est la capacité des bactéries pathogènes à former des biofilms qui leur permettent *via* la production d'une matrice extracellulaire de polymères de se protéger de la réponse immunitaire et des

antibiotiques, mais également de maximiser la mobilisation de nutriment par le biofilm bactérien ^{513,514}. La communication entre les bactéries au sein des communautés *via* le QS, c'est-à-dire la production de signaux moléculaires diffusibles dans le milieu extracellulaire, constitue un aspect fondamental de la régulation de la virulence des bactéries pathogènes qui sous-tend de nombreux réseaux de régulation ⁵¹⁵. Il est maintenant bien établi que les ARNnc jouent un rôle pivot dans la coordination des changements d'expression des gènes, que cela soit au niveau transcriptionnel, de la traduction ou de l'activité même de la protéine. De nombreux ARN régulateurs jouant un rôle clé dans la réponse adaptative et dans de nombreux processus pathogéniques ont été identifiés chez plusieurs bactéries pathogènes comme *S. aureus*, *Listeria monocytogenes*, les souches uropathogènes d'*E. coli* (UPEC), *Helicobacter pylori*, *Pseudomonas aeruginosa*, *Salmonella Typhimurium*, *Vibrio cholerae* et *Streptococcus pneumoniae* (**Figure 27**) ⁵¹⁶⁻⁵²⁰. Des exemples de pathogènes utilisant divers mécanismes de régulation par ARNnc durant leur cycle sont décrits ci-dessous.

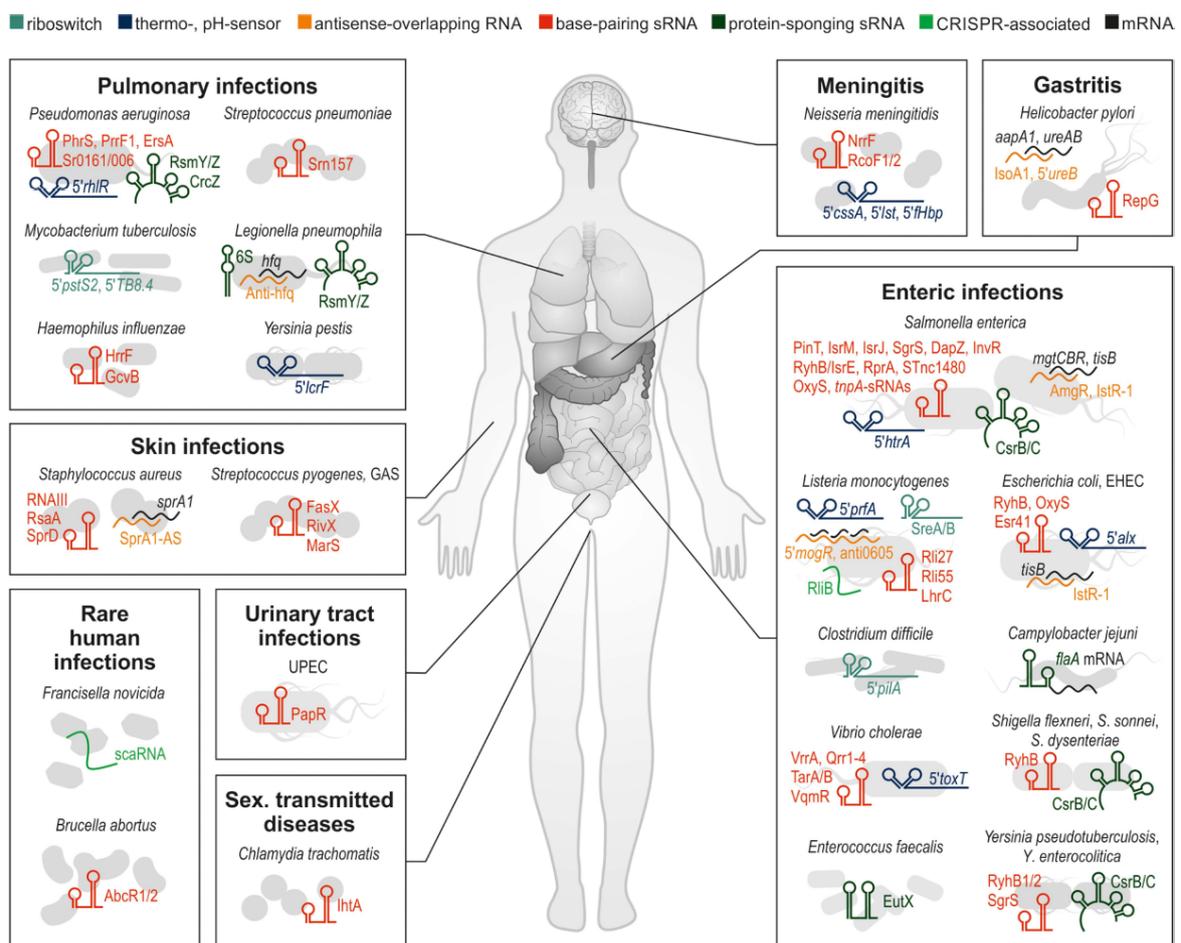


Figure 27 – ARN régulateurs associés à la régulation de la virulence chez des pathogènes humains majeurs. Les différentes classes ARN sont indiquées *via* le code couleur en haut de la figure. ⁵²⁰

1. Bactéries à Gram négatif

a. *Escherichia coli* uropathogène (UPEC)

Les souches uropathogènes d'*E. coli* (UPEC) représentent la cause principale des infections du tractus urinaire (ITU) qui peuvent affecter la vessie (cystites) ou les reins (pyélonéphrites), ces dernières étant associées à une forte morbidité voir même mortalité⁵²¹. Ces souches disposent de divers facteurs de virulence dont des toxines, des adhésines, des sidérophores et la capacité de former des biofilms extracellulaires ou des communautés bactériennes intracellulaires⁵²¹. Les ARN régulateurs sont des acteurs importants de la régulation de la virulence des UPEC. Il a été montré notamment que la protéine Hfq occupe une place importante dans les processus de colonisation du tractus urinaire, dans la formation de biofilms et de microcolonies intracellulaires, dans le maintien de l'homéostasie du lipopolysaccharide (LPS), dans la régulation de la tolérance à différents stress et dans la motilité⁵²².

Une expérience de co-immunoprécipitation d'ARN liés à la protéine Hfq a permis l'identification de nombreux petits ARN associés à Hfq durant l'infection⁵²³. En fonction des conditions de développement de la souche UPEC (laboratoire ou infectieuse), les profils d'ARNnc co-immunoprécipités avec Hfq varient fortement et en particulier ceux concernant les ARNnc associés au stress de l'enveloppe cellulaire qui sont très enrichis dans les conditions infectieuses. La même technique a permis d'identifier l'ARNnc agissant en *trans* PapR, qui régule négativement l'ARNm de *papI* codant pour un régulateur positif de l'expression des pili de type P, un facteur de virulence critique contribuant à l'adhérence de la bactérie au tissu rénal⁵²³. L'ARNnc PapR s'apparie avec l'ARNm de *papI* via une séquence complémentaire d'environ 80 nucléotides localisée en aval du site d'initiation de la traduction, empêchant ainsi la traduction du facteur PapI.

Chez les souches d'*E. coli*, l'ARNnc de 90 nucléotides RyhB régule l'assimilation et l'utilisation du fer³⁸⁴. Chez les souches UPEC, ce petit ARN induit la synthèse de sidérophores chélateur du fer (entérobactine, aérobactine et salmochéline) fondamentaux pour la pathogénèse en s'appariant avec les ARNm des précurseurs de ces molécules et en les stabilisant afin de permettre leur traduction⁵²⁴. Sa délétion entraîne des déficits de

colonisation de la vessie dans des modèles animaux. RyhB régule également certains processus infectieux chez d'autres pathogènes comme *Shigella*⁵²⁵ et *Vibrio*⁵²⁶.

b. Helicobacter pylori

H. pylori est un agent pathogène de la muqueuse gastrique, dont la prévalence dans le monde est de 50%, qui régule sa virulence par des ARNnc. Les symptômes des infections à *H. pylori* peuvent aller d'inflammations modérées du tissu gastrique à des ulcérations chroniques pouvant évoluer vers des tumeurs malignes dans les cas les plus sévères⁵²⁷. *H. pylori* a la capacité de moduler l'expression de miRNA du tissu gastrique, altérant ainsi la réponse immunitaire de l'hôte à son avantage⁵²⁸. Ce pathogène possède également d'autres facteurs de virulence, dont des petits ARN lui permettant de subsister et proliférer dans son hôte⁵²⁹. Des analyses *in silico* ont mis en évidence que de multiples ARNnc peuvent exercer une influence importante sur son caractère infectieux.

TlpB est un récepteur requis pour la colonisation des zones endommagées de la muqueuse gastrique par *H. pylori* et le chimiotactisme pH dépendant^{530,531}. L'ARNm de *tlpB* possède une séquence répétée de 6 à 16 guanines dans sa région 5'UTR. Des variations dans la longueur de cette séquence sont observées entre les différents isolats de *H. pylori* de différents patients et parfois même entre les isolats d'un même patient. Cette séquence répétée de G est ciblée par l'ARNnc RepG (*Regulator of polymeric G repeats*) et sa taille détermine l'interaction de l'ARNm de *tlpB* avec RepG⁵²⁹. L'appariement de RepG avec l'ARNm de *tlpB* se fait au niveau de la région 5'UTR, en amont du RBS inhibant ainsi la traduction par un réarrangement probable de la structure de cette région ou par occlusion du site de fixation du ribosome. De plus, RepG diminue la stabilité de l'ARNm de *tlpB* et la formation du duplex entraîne sa dégradation. Outre *tlpB*, RepG possède un plus large *targetome* avec lequel il interagit probablement *via* sa boucle terminatrice constituée de C/U.

La production abondante d'uréase est un mécanisme clé qui permet à *H. pylori* de survivre dans un environnement gastrique acide⁵³². L'uréase métabolise l'urée disponible en ammoniac et en bicarbonate, qui servent tous deux de tampon permettant de maintenir un pH équilibré dans la cellule bactérienne. Exprimées à partir du même opéron, UreA et UreB

sont les deux sous-unités de l'apoenzyme précurseur de l'uréase. La transcription de *ureAB* est régulée positivement par le système à deux composants (TCS ; *Two Component System*) ArsRS qui est activé dans des conditions acides, assurant ainsi une synthèse abondante d'uréase à faible pH et une diminution de sa synthèse à des pH neutres ou alcalins. La répression de *ureAB* dans des conditions de pH neutres ou alcalines est médiée par l'ARNas 5'*ureB* qui inhibe la transcription du gène *ureB* en se fixant à son extrémité 5'UTR^{533,534}. En effet, alors que le régulateur de réponse ArsR phosphorylé favorise la transcription sens de *ureAB* dans des conditions acides, la forme non phosphorylée (dans des conditions neutres/alcalines) augmente l'expression de l'ARNas 5'*ureB*. Lorsque l'ARNas 5'*ureB* est exprimé (à pH neutre à alcalin), l'ARNm bicistronique sens de *ureAB* est clivé en une forme plus courte de 1400 nucléotides au lieu de sa forme complète de 2700 nucléotide qui peut être traduite en deux sous-unités de l'apoenzyme. La forme courte de l'ARNm sens *ureAB* est le résultat d'une terminaison précoce de sa transcription lorsqu'il est apparié à l'ARNas 5'*ureB*.

c. *Pseudomonas aeruginosa*

P. aeruginosa est un pathogène opportuniste notamment connu pour donner des infections pulmonaires sévères chez les patients atteints de mucoviscidose, où il forme des biofilms persistants⁵³⁵. Une fois acquises, il est presque impossible d'éradiquer ces infections qui diminuent significativement la qualité et l'espérance de vie des patients atteints^{536,537}. La formation de biofilms est favorisée par le mucus épais et collant caractéristique de la mucoviscidose qui fait office de niche favorable au développement du pathogène⁵³⁸. De plus, l'infection à *P. aeruginosa* entraîne forte réponse immunitaire avec un recrutement massif de leucocytes qui ne peuvent pas éliminer l'infection et causent de nombreux dégâts tissulaires. Ce processus appelé *phagocytose frustrée* (ou incomplète)⁵³⁹ contribue à la pathogénèse en fournissant de l'ADN extracellulaire issu de la lyse des cellules qui servira à l'élaboration du biofilm^{540,541}. *P. aeruginosa* a la capacité de détecter et de répondre à des signaux environnementaux en modulant l'expression de ses gènes, en grande partie par des mécanismes de régulation post-transcriptionnel liés à des ARNnc. En effet, plus de 570 ARNnc ont été identifiés chez *P. aeruginosa*^{542,543}.

Près d'un dixième du transcriptome de *P. aeruginosa* fait partie du régulon Rsm (*Regulator of Secondary Metabolites*). Ce régulon agit comme un système de régulation post-transcriptionnel contrôlant plusieurs déterminants de la virulence, dont la transition entre les modes de vie planctonique (infections aiguës) et de biofilm (infections chroniques)⁵⁴³. Les composants clés du système Rsm sont deux protéines de liaison à l'ARN, RsmA et RsmF (appelée également RsmN), qui sont des orthologues de la protéine CsrA (régulateur du stockage du carbone) d'*E. coli*. RsmA et RsmF partagent 31 % d'identité et une arginine conservée qui est essentielle pour leur capacité de liaison à l'ARN. Les deux protéines peuvent se lier directement aux ARNm cibles et affecter positivement ou négativement leur stabilité et/ou leur traduction. La présence d'un (RsmA) et de deux (RsmF) motifs GGA conservés est importante pour la reconnaissance de la cible⁵⁴⁴. RsmA et RsmF activent toutes deux un phénotype d'infection aiguë et répriment la formation de biofilm. Les protéines RsmA et RsmF sont étroitement régulées par les quatre ARNnc RsmV, RsmW, RsmY et RsmZ qui vont pouvoir se lier à elles *via* leur séquence consensus GGA et les séquestrer les empêchant ainsi d'agir sur leurs ARNm cibles⁵⁴⁵. Ainsi, ces ARNnc sont des activateurs du phénotype d'infection chronique *biofilm*, tout en étant des répresseurs du phénotype *infection aiguë*. Bien que ces quatre ARNnc agissent de manière similaire, il existe des différences dans leurs affinités de liaison avec leur cible, ainsi qu'au niveau de leur schéma d'expression⁵⁴⁵. La régulation du régulon complexe Rsm est assurée, du moins en grande partie, par le système à deux composant GacAS qui régule positivement l'expression des gènes *rsmY* et *rsmZ* en fonction de certains signaux environnementaux (comme le taux de calcium intracellulaire ou la concentration en magnésium)⁵⁴⁶⁻⁵⁴⁸. Un nombre important d'autres facteurs vont intervenir dans la régulation du régulon Rsm comme la polynucléotide phosphorylase (PNPase) qui stabilise RsmY et RsmZ, ou encore la protéine TspR qui inhibe la transcription de *rsmYZ*^{549,550}. En revanche, contrairement à RsmYZ, les ARNnc RsmW et RsmV ne sont pas régulés par le système GacAS⁵⁴⁵.

L'ARNnc conservé de 100 nucléotides ReaL (*Regulator of alkyl quinolone*) est impliqué dans la régulation du QS chez *P. aeruginosa*⁵⁵¹. Il est sous le contrôle négatif du système du QS Las et active la traduction du transcrit *pqsC*, faisant ainsi le lien entre deux systèmes

importants du QS (*las* et *psq*) lui conférant ainsi un effet pléiotrope sur les traits de virulence majeurs. L'expression de *ReaL* est sous le contrôle de *RpoS* et est par conséquent fortement induite durant la phase stationnaire de croissance. *ReaL* inhibe également le *swarming* mais augmente la formation de biofilms et la sécrétion de pyocyanine et de pyoverdine (deux facteurs de virulence de *P. aeruginosa*)^{552,553}. L'expression de l'ARNnc *PhrS*, impliqué dans la production de pyocyanine et la synthèse du régulateur transcriptionnel *PqsR*, est également induite durant la phase stationnaire de croissance, mais aussi dans des conditions d'hypoxie grâce à la protéine ANR se fixant à l'ADN⁵⁵⁴⁻⁵⁵⁶. Le *locus prrF* chez *P. aeruginosa* contient les séquences des deux ARNnc *prfF1* et *prfF2* séparées par 95 bases et identiques à 95%^{557,558}. Ce sont des homologues fonctionnels de l'ARNnc *RyhB* de *E. coli* qui sont par conséquent impliqués dans le maintien de l'homéostasie du fer et la régulation de la virulence, mais également dans la motilité, le métabolisme des acides aminés et l'homéostasie du zinc^{557,559,560}. Leur expression est augmentée dans des conditions de carence en fer, durant lesquelles ils répriment la synthèse des protéines non essentielles se liant au fer comme *SodB*⁵⁶⁰. Les ARNnc *PrrF1* (116 nt) et *PrrF2* (114 nt) régulent le QS chez *P. aeruginosa* en s'appariant avec l'ARNm du gène *antR* et en bloquant sa traduction⁵⁶¹. *AntR* est un facteur de transcription qui active la transcription des *loci antABC* et *catBCA*, responsables de la dégradation de l'antranilate, un composé précurseur de la molécule de signalisation quinolone⁵⁴².

L'ARNnc de la réponse au stress de l'enveloppe *ErsA* est activé par le facteur de réponse au stress σ^{22} ⁵⁶². Des signaux de l'environnement pulmonaire des patients atteints de mucoviscidose, tels qu'une carence en fer et un faible taux d'oxygène, favorisent sa transcription. *ErsA* s'apparie directement au niveau du RBS de l'ARNm du gène *algC* codant pour une phosphomannomutase produisant des résidus de sucres pour la synthèse des exopolysaccharides de la matrice du biofilm⁵⁶³. *ErsA* régule également négativement la traduction de la protéine de la membrane externe *OprD* en s'appariant à l'extrémité 5'UTR de son ARNm, ce qui a pour conséquence une réduction de l'entrée des carbapénèmes dans la cellule, à l'origine d'une résistance potentielle à ces antibiotiques⁵⁶⁴. Un autre ARNnc appelé *Sr0161* réprime également l'expression d'*OprD* par le même mécanisme⁵⁶⁵.

2. Bactéries à Gram positif

a. *Staphylococcus aureus*

Bien que *S. aureus* soit souvent retrouvé parmi les espèces du microbiome cutané humain sain, c'est l'un des agents pathogènes les plus couramment impliqués dans des infections de plusieurs organes dont les tissus mous et la peau, les os, le cœur, les voies respiratoires et le système circulatoire ⁵⁶⁶. Il est également réputé pour sa capacité à former des biofilms très résistants sur les prothèses et les implants ⁵⁶⁷. *S. aureus* est l'un des principaux agents responsables des infections associées aux soins, et la majorité des souches isolées en clinique sont résistantes aux antibiotiques ⁵⁶⁸. Il appartient au groupe bactérien ESKAPE (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa* et *Enterobacter sp.*) qui regroupe les espèces les plus résistantes aux antibiotiques ⁵⁶⁹. Il possède tout un arsenal de facteurs de virulence comprenant des toxines, des adhésines, des molécules d'immunoévasion et la capacité de former des biofilms ⁵⁷⁰. Ces éléments de pathogénèse sont étroitement régulés, et les ARNnc jouent un rôle important dans la modulation de leur expression ^{571,572}.

Le RNAIII est sans nul doute l'ARNnc le mieux caractérisé chez *S. aureus* ^{376,573}. Son expression est sous le contrôle du *locus agr* sensible au QS qui influence l'expression de nombreux gènes de virulence chez *S. aureus*. Ce *locus* comprend deux ORF transcrits à partir de deux promoteurs différents (P2 et P3) et dans deux directions opposées. P2 permet la transcription de l'ARNm tétra-cistronique RNAII qui code pour : le peptide auto-inducteur (PAI) AgrD produit dans sa forme inactive, le PAI associé à la membrane AgrB qui permet la maturation de AgrD dans sa forme active et son exportation extracellulaire, le senseur histidine kinase AgrC et son régulateur de réponse AgrA du TCS AgrAC. Dans des conditions de densité cellulaire élevée, le PAI AgrD est détecté par le senseur AgrC qui va transmettre le signal par la phosphorylation du régulateur de réponse AgrA qui, à son tour, régule positivement la transcription du RNAIII (produit à partir de P3) qui joue un rôle pléiotrope chez *S. aureus*. La région centrale du RNAIII code pour la δ -hémolysine qui confère à *S. aureus* son activité hémolytique. Le RNAIII peut également agir en tant que riborégulateur, régulant de nombreuses cibles ARNm. La demi-vie du RNAIII est de 45 minutes et il est

structurellement caractérisé par 14 *tiges-boucles* et deux longues structures hélicoïdales. Des structures spécifiques en *tête d'épingle* sont impliquées dans l'appariement avec les cibles ARNm avec plusieurs *tiges-boucles* agissant de concert pour l'achèvement de la régulation. Afin de réprimer la traduction, le RNAIII peut se fixer soit au RBS, soit à la fois au RBS et au 5'UTR par de nombreuses *tiges-boucles*, ou bien à la région codante ⁵⁷⁴. L'action du RNAIII est très souvent négative, et les seules cibles connues pour être régulées positivement par le RNAIII sont les ARNm *hla*, codant pour l' α -hémolysine, et *eap*, codant pour une protéine d'adhérence extracellulaire ⁵⁷⁵. Bien que le RNAIII possède un très grand nombre de cibles auxquelles il se fixe directement, il peut également moduler indirectement plusieurs régulateurs en inhibant par exemple la traduction du répresseur transcriptionnel *rot* (*repressor of toxins*) ⁵⁷⁶. Le RNAIII réprime principalement la traduction des gènes codant des protéines de surface ou des protéines associées au recyclage du PG qui sont généralement requises lors des stades précoces de l'infection. A l'inverse, il active la synthèse d'exotoxines sécrétées, nécessaires à la dissémination bactérienne à des moments plus avancés de l'infection lorsque la densité cellulaire est plus importante. Il a été décrit que le RNAIII facilite le passage d'un mode de colonisation *silent bacteria* vers un mode de destruction des tissus de l'hôte ⁵⁷³. Le RNAIII occupe ainsi une place centrale dans la régulation de la virulence chez *S. aureus* et, hormis quelques exceptions, tous les effets de régulation en aval du QS sont médiés par le RNAIII ⁵⁷⁷. Le RNAIII est donc un régulateur polyvalent intervenant de manière directe ou indirecte dans des voies de signalisation en cascade.

Au moins 11 autres ARNnc ont été identifiés chez *S. aureus* et nommés RsaA à K (pour *RNA from Staphylococcus aureus A to K*) ⁵⁷⁸. Parmi ceux-là, l'ARNnc RsaA, contrairement au RNAIII, favorise un mode infectieux chronique et de persistance ⁵⁷⁹. La transcription du gène *rsaA* est fortement exprimé en phase stationnaire mais peu exprimé en phase exponentielle. Son expression est sous le contrôle positif du facteur σ^B et les deux endoribonucléases RNAse III et RNAse Y régulent sa dégradation ⁵¹⁷. L'ARNnc RsaA permet d'inhiber la traduction de l'ARNm du gène *mgrA* codant pour le régulateur transcriptionnel MgrA en s'appariant avec lui et en induisant ainsi sa dégradation. RsaA peut se fixer à l'ARNm *mgrA* au niveau de deux

régions différentes : un motif riche en C cible la séquence SD et deux motifs en *épingle à cheveux* dans la région 5' de RsaA interagissent avec la séquence codante de *mgrA*. Une étude récente a validé l'interaction entre l'ARNnc RsaA et l'ARNm *mgrA* par une technique de co-purification de complexes ARN-ARN et l'identification des cibles par séquençage (MAPS ; *MS2-affinity purification*), et élargit le *targetome* de RsaA ⁵⁸⁰. Le régulon MgrA possède environ 350 gènes dont le sien qu'il autorégule ⁵⁸¹. Le rôle principal de MgrA est l'activation de la synthèse de la capsule polysaccharidique et l'inhibition de la formation de biofilm en réprimant l'expression et le relargage d'ADN extracellulaire ⁵⁸².

L'ARNnc SprD est transcrit à partir d'un îlot de pathogénicité et sa cible principale est l'ARNm *sbi*, codant pour la protéine d'immunoévasion Sbi (*Second binder of immunoglobulins*) ⁵⁸³. Sbi inhibe l'opsonisation, qui est généralement suivie d'une phagocytose et d'une action du système du complément, en se liant spécifiquement aux IgG et aux facteurs du complément C3b et H ⁵⁸⁴. Cependant, bien que l'action de Sbi ait pour but l'échappement au système immunitaire, cette protéine possède un fort potentiel immunogène qui a pour conséquence la production de cytokines majeures et le recrutement de leucocyte au site d'infection. L'ARNnc a donc pour rôle de contrôler son expression en inhibant sa traduction afin d'éviter sa production constitutive. SprD agit de manière indépendante de la protéine Hfq et peut se fixer au niveau d'une partie de la région 5'UTR, de la séquence SD et du codon START de la traduction de l'ARNm *sbi*. La délétion de SprD diminue considérablement la morbidité et la mortalité en modèle animal, confirmant ainsi son rôle majeur dans la pathogénèse de *S. aureus*.

b. Listeria monocytogenes

L. monocytogenes est un pathogène intracellulaire et un modèle particulièrement intéressant de régulation de la pathogénèse par les ARNnc. Bien que les infections à *L. monocytogenes* (listérioses) soient relativement rares, la mortalité associée à ces infections demeure assez élevée (20% à 30%), particulièrement chez les enfants, les personnes âgées, les immunodéprimés et les femmes enceintes ⁵⁸⁵. La contamination se fait principalement par de la nourriture contaminée et, pour une personne vulnérable, la dose infectieuse peut être inférieure à une centaine de bactéries ⁵⁸⁵. En tant que pathogène gastro-intestinal, *L.*

monocytogenes franchit la barrière de l'épithélium intestinal afin de disséminer dans tout le corps (particulièrement la rate et le foie) *via* les systèmes circulatoires (sanguin et lymphatique). Plus de 150 ARNnc ont été identifiés chez *L. monocytogenes*⁵⁸⁶.

L'ARNnc Rli27 contrôle positivement la traduction de la protéine de surface Lm0514 en se fixant à l'extrémité 5'UTR de son ARNm et en exposant le RBS⁵⁸⁷. Le niveau d'expression de cette protéine est fortement augmenté (200 fois) durant la phase de croissance intracellulaire durant laquelle elle contribue à la survie de la bactérie, mais en revanche son niveau d'ARNm n'augmente que d'un facteur 6 ce qui va bien dans le sens d'une régulation au niveau traductionnel⁵⁸⁸.

L'ARNnc LhrC (*Listeria Hfq-binding RNA C*) est un petit ARN découvert lors d'une expérience de co-immunoprécipitation avec la protéine Hfq chez *L. monocytogenes*⁴⁷⁰. Des études ultérieures ont cependant mis en évidence que, comme pour la plupart des cas de régulation par ARNnc chez les espèces à Gram positif, la protéine Hfq n'est pas requise pour la stabilité de LhrC ou ses interactions avec ses cibles ARNm. Il existe cinq copies du gène *lhrC* dont la taille varie de 111 à 114 nucléotides⁵⁸⁹. Les expressions des ARNnc LhrC1-5 sont augmentées durant des conditions de stress de l'enveloppe cellulaire pouvant être générés par les céphalosporines ou les acides biliaires⁵⁸⁹. Il a été montré que l'hème, fréquemment en contact avec *L. monocytogenes* lors de la dissémination systémique, permet d'activer l'expression de LhrC⁵⁹⁰. L'une de cibles de LhrC est l'ARNm *lapB* qui code pour une adhésine attachée à la paroi cellulaire⁵⁸⁹. LhrC possède trois régions riches en C formant des motifs UCCC qui sont très importants dans les interactions avec la région riche en AG de la séquence SD des ARNm cibles dont il réprime la traduction. Les cinq copies de LhrC et les trois sites de fixation CU-riches redondant reflètent l'importance que peut avoir LhrC dans la transduction d'un signal faible en réponse forte. De plus, les cinq copies de LhrC semblent agir de manière synergique plutôt que de manière redondante. LhrC peut également se fixer (*via* ses motifs UCCC) et réprimer les ARNm de gènes impliqués dans l'influx de l'hème dans la cellule et dans son métabolisme jouant ainsi un rôle important dans la survie de la bactérie dans des niches à forte lyse érythrocytaire⁵⁹⁰. En utilisant toujours le même mécanisme de fixation, LhrC réprime également la traduction d'une autre protéine membranaire, OppA,

impliquée dans la fixation d'oligopeptides ⁵⁹¹. Des travaux récents ont également mis en évidence l'inhibition de la production d'un antigène des lymphocyte T (TcsA) en agissant sur la stabilité de son ARNm ⁵⁹².

Durant l'infection du tractus gastro-intestinal des mammifères, *L. monocytogenes* utilise l'éthanolamine comme source de carbone et d'azote en activant l'expression de son *locus eut* (*ethanolamine utilization*) dont la transcription est régulée par des ARNnc ^{518,593}. Le premier niveau de régulation est médié par le TCS EutWV (le senseur histidine kinase EutW et le régulateur de réponse EutV) qui est activé par la détection d'éthanolamine dans le milieu extracellulaire. Le domaine de fixation à l'ARN ANTAT du régulateur EutV phosphorylé permet de lever la répression de la transcription des gènes *eut* due à des structures *tige-boucle* de leur ARNm natif ⁵⁸⁶. En amont du *locus eut* se trouve le gène de l'ARNnc Rli55 agissant comme second niveau de régulation des gènes *eut*. La région 5' de Rli55 présente un *riboswitch* liant la vitamine B12. En absence de vitamine B12, Rli55 est transcrit en un ARN de 450 nucléotides de long qui a pour fonction de séquestrer EutV, empêchant ainsi la transcription des gènes *eut*. A l'inverse, en présence de vitamine B12, le *riboswitch* est complexé à son ligand entraînant la transcription d'une forme plus courte de Rli55 (200 nucléotides) qui ne peut pas séquestrer EutV, lui permettant donc d'exercer son rôle d'antitermineur de la transcription des gènes *eut* ⁵⁹⁴. L'éthanolamine et la vitamine B12 sont donc essentiels pour l'activation des gènes permettant le métabolisme de l'éthanolamine chez *L. monocytogenes*. Un autre exemple de *riboswitch* liant la vitamine B12 chez *L. monocytogenes* est AspocR qui régule l'utilisation du propanediol comme source de nutriment durant l'infection ⁵⁹⁵. L'expression de l'opéron *pdu* (*propanediol utilization*) est médiée par le facteur de transcription PocR ⁵⁹⁶. Sur le brin opposé du gène *pocR* se trouve un ARNas contrôlé par le *riboswitch* AspocR. En l'absence de vitamine B12, le *riboswitch* agit comme antitermineur de l'ARNas de *pocR* qui, une fois transcrit, inhibe la transcription de *pocR* et par conséquent des gènes *pdu*. A l'inverse, en présence de vitamine B12, la conformation du *riboswitch* atténue la transcription de AspocR entraînant la production d'un ARNas tronqué qui est incapable de s'apparier à l'ARNm de *pocR*, permettant ainsi l'utilisation de propanediol.

La régulation traductionnelle du gène *prfA* codant un activateur transcriptionnel de gènes de virulence est un autre exemple de mécanisme de régulation par ARNnc chez *L. monocytogenes*⁵⁹⁷. Le thermosenseur localisé dans la région 5'UTR de *prfA* forme une structure en *épingle à cheveux* lorsque la température est de 30°C, bloquant ainsi le RBS et empêchant la traduction de *prfA*. Lorsque la température atteint 37°C, le thermosenseur est déstructuré permettant ainsi la traduction de *prfA*. L'activation de l'expression de *prfA* permet la production d'un grand nombre de toxines, d'enzyme lytiques et de protéine remodelant l'actine, assurant donc la production de facteurs de virulence uniquement à température physiologique (37°C) lorsque la bactérie est en contact avec son hôte⁵⁸⁶. Un second mécanisme de régulation est associé à la disponibilité en nutriments et est médié par les *riboswitch* SreA et SreB⁵⁹⁸. En présence du ligand de ces deux *riboswitch*, la S- adénosylméthionine (SAM), leur transcription est prématurément arrêtée ce qui permet la production d'un petit transcrit non codant qui se fixe au niveau du RBS de l'ARNm de *prfA*, empêchant ainsi sa traduction.

Le concept d'excludon définit un *locus* de gènes ayant des fonctions liées et souvent antagonistes. Les transcrits de ces gènes peuvent avoir à la fois une fonction antisens de l'ARNm du gène situé sur le brin opposé et coder pour un ARNm fonctionnel⁵⁹⁹. Ainsi un excludon peut réguler négativement un gène sur le brin opposé et positivement les gènes voisins localisés sur le même brin d'ADN. Un exemple classique d'excludon chez *L. monocytogenes* est la régulation de la biosynthèse du flagelle codé par les gènes *Imo0675-0689* de l'opéron *fli*^{586,599}. Les gènes *Imo0676* et *Imo0677* codent pour les protéines FliP et FliQ qui font partie intégrante de la machinerie de sécrétion de l'assemblage flagellaire. Sur le brin opposé est codé le long ARN Anti0677 qui présente une séquence complémentaire des gènes *Imo0675*, *Imo0676* et *Imo0677* et agit comme ARNs inhibant leur expression. La transcription de Anti0677 à partir d'un second promoteur permet également de transcrire le gène voisin répresseur de motilité *mogR* faisant également de Anti0677 un ARNm fonctionnel. Cet excludon régule donc la synthèse du flagelle selon deux mécanismes : la diminution de sécrétion du flagelle par un mécanisme ARNs et l'augmentation de la production de MogR qui diminue également la production du flagelle.

3. Le cas de *Clostridioides difficile*

Des ARN régulateurs issus de différents groupes fonctionnels (ARNs CRISPR, ARNs antisens, *riboswitch* et riborégulateurs agissant en *trans*) ont récemment été identifiés dans divers processus de régulation de l'expression des gènes de *C. difficile* et agissent de concert avec des facteurs σ et des protéines régulatrices spécifiques^{174,600,601}. En combinant analyses *in silico*, séquençage ARN (RNA-seq) et TSS-mapping (*Transcription Start Site Mapping*), plus de 200 ARNnc ont été identifiés chez ce pathogène^{174,602-604}. La plupart des ARNnc identifiés chez *C. difficile* sont des *riboswitch*, avec 77 *riboswitch* potentiels détectés par TSS-mapping et RNA-seq^{174,604}. *C. difficile* possède notamment un grand nombre de *riboswitch* dépendant du c-di-GMP et ces éléments furent parmi les premiers ARNnc fonctionnels à avoir été identifiés chez *C. difficile*^{173,174,483,605,606}. Le génome de la souche 630 de *C. difficile* contient 16 *riboswitch* dépendant du c-di-GMP dont 12 de type I (possédant un motif GEMM ; *Genes for the Environment, for Membranes and for Motility*) et 4 de type II^{173,483,607}. Chez *C. difficile*, les *riboswitch* de type I exercent globalement une régulation négative en induisant la terminaison de la transcription en présence de c-di-GMP. A l'inverse, les *riboswitch* de type II induisent l'expression des gènes qu'ils contrôlent en présence de c-di-GMP. Onze de ces *riboswitch* dépendant du c-di-GMP ont été identifiés comme fonctionnels et régulant l'expression des gènes en aval^{174,483,606}. Tandis que les *riboswitch* fonctionnels sont très conservés entre les souches de *C. difficile* (à moins que les gènes qu'ils régulent ne soient absents), les *riboswitch* non fonctionnels le sont moins⁶⁰⁶. *C. difficile* 630 code 37 diguanylates cyclases (DGC) et phosphodiesterases (PDE) impliquées dans le *turnover* du c-di-GMP qui, avec la large utilisation des *riboswitch* comme molécules effectrices, suggèrent un rôle clé du c-di-GMP dans le contrôle des processus infectieux de *C. difficile*^{174,483,600,605}. Deux de ces *riboswitch*, *cdi1_4* et *cdi1_5*, sont localisés dans les prophages ϕ CD630-1 et CD630-2 de la souche 630 de *C. difficile* et sont associés avec des ARN antisens appartenant à un système TA de type I⁶⁰⁸. Deux *riboswitch* dépendant du c-di-GMP ont également été identifiés en amont des gènes codant pour la métalloprotéase *Zmp1* et son substrat CD630_28310. Le *riboswitch* situé en amont du gène de *Zmp1* est de type I et celui situé en amont du gène de CD630_28310 est de type II suggérant une régulation positive de l'expression de *Zmp1* dans des condition de faible concentration de c-di-GMP et un

régulation positive de CD630_28310 en présence de concentration de c-di-GMP élevée ¹⁷⁴.

La protéine chaperonne Hfq de *C. difficile* présente des caractéristiques uniques en comparaison à celles des autres bactéries à Gram positif, puisqu'elle semble un élément clé des réseaux de régulation post-transcriptionnels chez *C. difficile*, et une centaine d'ARNnc localisés dans des IGR pouvant nécessiter l'assistance de la protéine Hfq ont été identifiés ^{600,602,604,609,610}. Une analyse RIP-seq (*RNA immunoprecipitation high-throughput sequencing*) a permis d'identifier un grand nombre d'ARNnc agissant en *trans* associés à la protéine Hfq (comme l'ARNnc RCd1) ^{602,604}. En plus des ARNnc agissant en *trans*, un certain nombre d'ARNnc agissant en *cis* dont des ARNnc antitoxine appartenant aux systèmes TA de type I ^{608,611,612}, de nombreux *riboswitch* et des ARN CRISPR ont été identifiés comme étant associés à la protéine Hfq, constituant ainsi une nouvelle catégorie de ligands pour Hfq ⁶⁰². Les complexes ARNnc-ARNm et ARNnc-ARNnc prédits suggèrent un grand potentiel de ciblage pour les ARN associés à Hfq chez *C. difficile*. Des régulations croisées entre les ARN codés par les prophages et les ARN bactériens pourraient être envisagées et mériteraient des validations expérimentales supplémentaires en accord avec le rôle de Hfq dans les mécanismes basés sur les interactions phages-bactéries médiées par les ARN ^{404,412}. De nombreuses interactions ARN-protéines chez *C. difficile* ont récemment été révélées par approche Grad-seq, ouvrant ainsi des perspectives intéressantes pour l'identification future de nouvelles protéines liant des ARN et leurs cibles ⁶¹³.

C. difficile possède un système actif CRISPR de type I-B et de nombreux ARN CRISPR ont été détectés par RNA-seq ^{174,355,361,600}. Des expériences de séquençage du génome, de RNA-seq et des analyses bio-informatiques ont permis d'identifier un très grand nombre de cassettes CRISPR chez *C. difficile*, soit 8,5 par génome en moyenne et jusqu'à 3 opérons *cas* dans la majorité des souches séquencées ^{154,174,355,360,361,614}. Parmi ces cassettes CRISPR identifiées, plusieurs d'entre elles sont localisées à l'intérieur de prophages (**Figure 16**) ^{360,361}. Les souches *C. difficile* CD630E et R20291 présentent une expression active de toutes les régions CRISPR, et les crRNA des régions CRISPR codées par les prophages sont parmi les plus abondants ^{174,361}. Plus d'un tiers de tous les *spacers* analysés ciblent les phages de *Clostridium*, et la moitié des régions chromosomiques de *C. difficile* ciblées correspondaient

à des régions prophagiques³⁶¹. Étant donné le nombre limité de séquences de génomes de phages des Clostridies, ces proportions pourraient être sous estimées³⁵³. Les interactions importantes des souches de *C. difficile* avec les phages ont permis l'acquisition de séquences dérivées de phages conservées dans des régions CRISPR comme mémoire génétique d'infections antérieures^{355,360,361}. Des expériences ont permis de démontrer l'efficacité de la fonction adaptative et d'interférence du système CRISPR-Cas de type I-B chez *C. difficile*, et un motif PAM 5'-CCW (CCA/T) a été déterminé comme essentiel dans la distinction du soi et du non soi^{361,615}. Un lien fonctionnel entre les machineries d'adaptation et d'interférence du système CRISPR-Cas de *C. difficile* a été mis en évidence puisque les *spacers* nouvellement acquis sont dérivés de séquences associées à un motif PAM CCN, ce qui correspond au motif consensus d'interférence YCN (C/TCN). À partir de cette observation, une nouvelle méthode d'édition du génome utilisant le système natif CRISPR-Cas a été développé chez *C. difficile*⁶⁰⁹. Parmi les applications médicales potentielles, la diversité au sein des séquences *spacer* pourrait être mise à profit dans le développement de nouvelles approches de typage pour la caractérisation des souches de *C. difficile*⁶¹⁴. À proximité des cassettes CRISPR, des ARNas de gènes codant des petites protéines de 50 à 53 acides aminés ont été détectés, et des analyses approfondies ont permis de révéler que ces protéines et leurs ARNas associés présentaient toutes les caractéristiques des toxines et antitoxines des systèmes TA de type I^{174,611}. L'association et la corégulation de ces modules TA avec les régions CRISPR par le facteur de la réponse générale au stress σ^B et les facteurs liés à la formation de biofilm ont soulevé la question du lien fonctionnel potentiel entre le mécanisme de dormance cellulaire et le système immunitaire adaptatif. Cinq systèmes TA de type I conservés chez les prophages de *C. difficile* ont également été décrits comme contribuant à la stabilité des régions génomiques prophagiques⁶⁰⁸. Jusqu'à présent, un total de 13 modules TA de type I ont été détectés dans le génome de la souche 630 de *C. difficile*, dont 8 au sein de prophages⁶¹². D'autres variants de petites protéines de 30 à 56 acides aminés de systèmes TA de type I ont été révélés par une étude plus approfondie des génomes des phages de *C. difficile*⁶⁰⁸. Des nombreux points restent encore à élucider concernant ces modules TA de type I.

IV. Les ARN régulateurs chez les eucaryotes

La découverte des ARN régulateurs chez les eucaryotes est arrivée quelques années après celle chez les procaryotes ⁶¹⁶. Depuis, de nombreuses classes d'ARN régulateurs eucaryotes ont été décrites, dont les micro-ARN (miRNA), les ARN interagissant avec les protéines Piwi (piRNA) et les longs ARNnc (lncRNA). Bien que leur étendue et leur diversité aient été pendant longtemps sous-estimées, les ARN régulateurs sont maintenant reconnus comme étant des régulateurs cellulaires puissants pouvant répondre à une très grande variété de signaux environnementaux et intracellulaires, et il a même été suggéré que le nombre de gènes exprimant des ARN régulateurs dans le génome humain serait supérieur à celui des gènes codant pour des protéines ⁶¹⁶. De plus, les ARN régulateurs occupent une place centrale dans les régulations épigénétiques de tous les organismes vivants et, bien que différents sur plusieurs aspects, des similitudes peuvent être observées dans les mécanismes et la biogenèse des ARN régulateurs issus des eucaryotes et des procaryotes ⁶¹⁷.

1. Aperçu des différentes classes et de leurs modes d'action

La diversité des ARN régulateurs chez les eucaryotes se reflète au travers des différents *loci* codant ces ARN, leurs processus de biogenèse et les différents effecteurs auxquels ils se lient, permettant ainsi de distinguer les différentes classes auxquelles ils appartiennent. Il existe plusieurs types de *loci* permettant de générer des ARN régulateurs fonctionnels parmi lesquels on peut globalement distinguer les petits ARNnc et les long ARNnc. Les petits ARN peuvent induire la répression de gènes *via* l'appariement par complémentarité de bases avec leur molécule cible. Dans la mesure où ces phénomènes ont été considérés comme indépendants les uns des autres au moment de leur découverte, plusieurs noms leur ont été attribués comme ARN interférence (ARNi), co-suppression, *gene silencing* ou *quelling*. Toutes ces voies de répression de l'expression de gènes sont maintenant collectivement appelées *RNA silencing*, bien que le terme ARN interférent soit aussi utilisé dans les phénomènes de clivage d'ARNm ciblé par un ARN double brin complémentaire. Les petits ARN régulateurs peuvent être classés en trois catégories principales selon leur aspect, leur origine, leur structure, les protéines effectrices avec lesquelles ils sont associés et leurs rôles biologiques : les micros-ARN (miRNA), les petits

ARN interférents (siRNA) et les ARN interagissant avec la protéine Piwi (piRNA). La plupart des petits ARN sont obtenus par maturation d'ARN précurseurs par des protéines de la famille Dicer qui sont des protéines d'environ 200 kDa présentant un domaine ATPase/hélicase, un domaine PAZ (Piwi/Argonaute/Zwille), deux domaines catalytiques RNase III et un domaine de fixation à l'ARN double brin. D'autres ARN sont obtenus après amplification d'un transcrit primaire par une ARN polymérase ARN dépendante (RrRP) en ARN double brin.

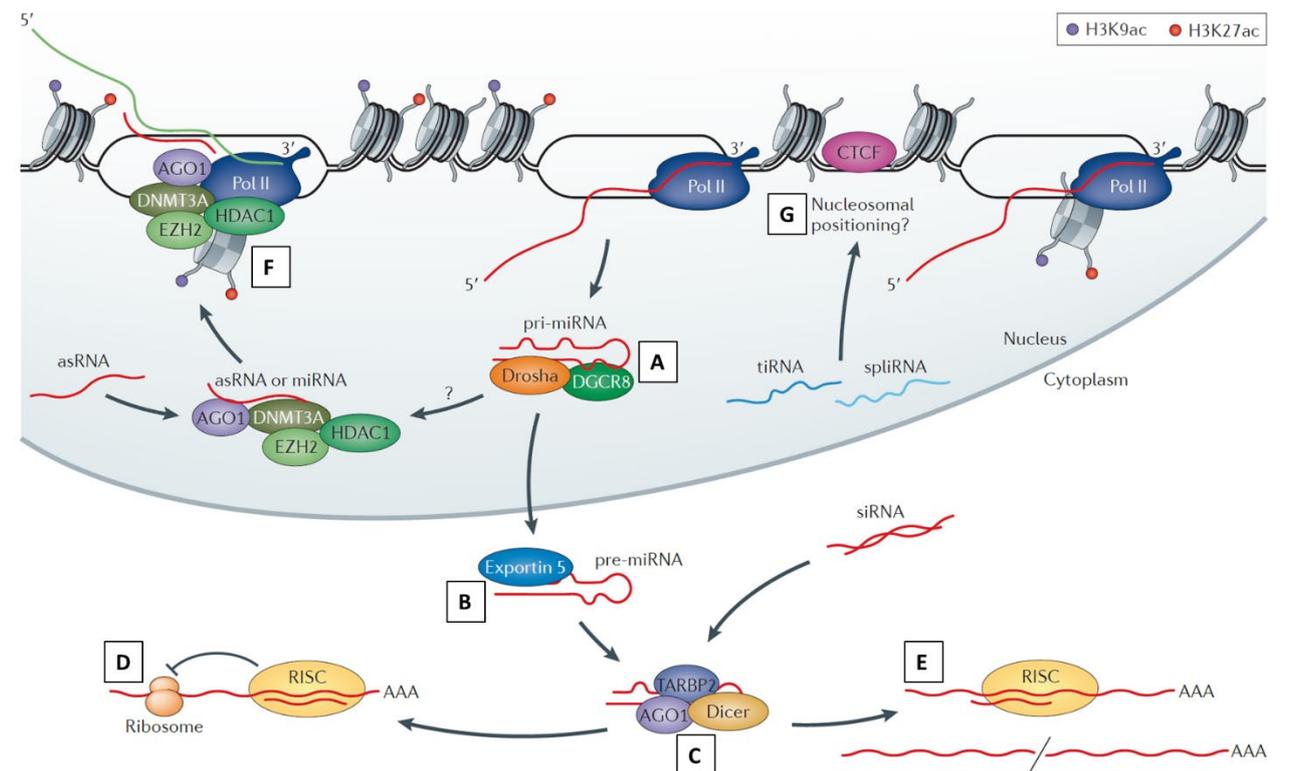


Figure 28 – Mécanismes de régulation par ARNnc chez les eucaryotes. (A) Les précurseurs des miRNA (pri-miRNA) présentent des structures en tige-boucle qui interagissent avec les protéines Drosha et DGCR8. (B) Ils sont ensuite maturés en pré-miRNA et exportés hors du noyau par l'exportine 5. (C) Puis ils sont coupés en plus petits ARN double brin (21-23 nucléotides), dont l'un des brins est chargé sur la protéine Argonaute (AGO) du complexe RISC (*RNA-induced silencing complex*). Les petits ARN interférents (siRNA) sont également maturés par RISC. Les miRNA et siRNA peuvent agir sur leur cible (D) en réprimant leur traduction et/ou (E) en induisant leur dégradation. (F) Quelques petits ARN sont fonctionnels dans le noyau comme les ARNs exogènes et les miRNA qui peuvent induire la répression de l'expression de leur cible en modifiant les marques épigénétiques du *locus* ciblé. (G) Les tiRNA (*Transcription initiation RNAs*) et les spliRNA (*splice site RNAs*), dont l'origine et la fonction sont inconnues, pourraient jouer un rôle dans le positionnement des nucléosomes et/ou dans d'autres niveaux d'organisation de la chromatine. Ils sont représentés dans cette figure comme agissant sur la localisation chromatiniennne du répresseur transcriptionnel CTCF (*CCCTC-binding factor*).⁶¹⁶

a. Les micro-ARN (miRNA)

Les miRNA sont des régulateurs post-transcriptionnels d'une vingtaine de nucléotides retrouvés dans les génomes des eucaryotes pluricellulaires et des plantes unicellulaires, et

dont les profils d'expression et de régulation varient entre les différents tissus et types cellulaires ⁶¹⁸. Ils sont le plus souvent transcrit par l'ARN polymérase II en ARN mono ou polycistroniques, et sont majoritairement ($\approx 70\%$) localisés dans les régions introniques et exoniques des unités de transcription (UT) ⁶¹⁹. Ils sont générés à partir de transcrits précurseurs (pri-miRNA) contenant des structures en *épingle à cheveux* qui interagissent dans le noyau avec la RNase III Drosha et la protéine DGCR8 afin de subir une première étape de maturation donnant les pre-miRNA qui sont ensuite exportés du noyau par l'exportine 5 (**Figure 28 A et B**). Les pre-miRNA sont ensuite pris en charge et clivés dans le cytoplasme par la RNase de type III Dicer afin de former des miRNA matures double brin de plus petite taille (21-23 nucléotides) dont l'un des brins est chargé sur la protéine Argonaute (AGO) du complexe RISC (*RNA-induced silencing complex*) qui est responsable de l'activité répressive ⁶²⁰. La région 5' des miRNA (2nt à 8nt) s'apparie à la région 3'UTR de leur ARNm cible grâce à une séquence partiellement complémentaire pouvant conduire à la répression traductionnelle ou la dégradation de cet ARNm (**Figure 28C**).

b. Les ARN interférents (siRNA)

Un autre groupe fonctionnel de petits ARN régulateurs est également connu sous le nom d'ARN interférents ou siRNA (**Figure 28C**). ⁶²¹. Ces petits ARN ont pendant longtemps été considérés comme exclusivement exogène jusqu'à la découverte d'autres sources de siRNA endogènes (endo-siRNA) incluant des transcrits convergents et autres paires sens-antisens, des duplex de transcrits antisens dérivés de pseudogènes et de leurs ARNm apparentés, et des ARN en *épingle à cheveux* ⁶²². L'inducteur canonique de ce groupe de répresseurs est un long ARN double brin linéaire qui est pris en charge et maturés en siRNA par la protéine Dicer ^{621,623}. Les siRNA ont été initialement observés lors de l'extinction de gènes chez les plantes induite par des transgènes et des virus ⁶²⁴. Des centromères, transposons et autres séquences répétées ont également été identifiées comme sources de siRNA ^{625,626}. Des études ont également révélé chez les plantes une sous-catégorie de siRNA agissant en *trans* (tasiRNA) qui sont issus de *loci* spécifiques du génome (*TAS*) et qui régulent un ensemble précis de gènes. Les transcrits *TAS* sont dépendant de la RNA polymérase II et fonctionnent comme des précurseurs hyperspécialisés de siRNA dont la synthèse dépend

du clivage de miRNA (en fragment de 21 nucléotides)⁶²⁷. Il est donc maintenant clair que les siRNA ne sont pas limités aux seuls acides nucléiques exogènes mais peuvent également provenir de *loci* génomiques endogènes également.

c. Les ARN interagissant avec la protéine Piwi (piRNA)

Enfin, la dernière classe des petits ARN est constituée par les ARN interagissant avec la protéine Piwi (ou piRNA) qui sont retrouvés dans les lignées germinales de mammifères et d'insectes. Chez *Drosophila melanogaster*, les piRNAs proviennent de transposons et de régions répétées, tandis que chez les mammifères ils proviennent de cluster de gènes éteints par le développement qui contiennent fréquemment des transposons et de rétroéléments^{628,629}. Chez les insectes et les mammifères, les piARN sont associés à des membres de la sous-famille Piwi des protéines Argonaute qui sont requis pour la biogenèse. Un modèle proposé de biogenèse des piRNA implique un ciblage initial des transcrits de transposons et de rétroéléments par une protéine Piwi-like qui est associée à un petit ARN. Le clivage du transcrit génère l'extrémité 5' du nouveau siRNA et l'extrémité 3' sera obtenue à partir d'un clivage réalisé par une autre protéine Piwi-like (comme AGO3)⁶³⁰.

d. Les long ARNnc (lncRNA)

Les longs ARNnc (lncRNA) sont définis comme des ARN de plus de 200 nucléotides ne présentant pas d'ORF (de plus de 100 codons)⁶³¹. Etant donné la grande diversité de lncRNA décrite actuellement, leur mode de fonctionnement ne peut pas être déduit de leur séquence ou de leur structure contrairement aux miRNA. Le large répertoire fonctionnel des lncRNA inclue le remodelage de la chromatine, la biologie et l'organisation structurelle subcellulaire des télomères mais également la régulation de l'expression de gènes voisins codant des protéines⁶³¹. Les lncRNA peuvent intervenir sur la structure de la chromatine en recrutant des complexes de remodelage de la chromatine qui modifie les marques épigénétiques de *loci* spécifiques du génome. Les lncRNA peuvent réguler la transcription des gènes par différents mécanismes. Ils peuvent agir sur le recrutement de protéines modulant l'activité de certaines histones acétyltransférase afin de réprimer la transcription, agir en tant que co-activateur d'un facteur de transcription ou former des triplex ADNdb-ARN au niveau de promoteur afin d'empêcher la fixation des facteurs de transcription⁶³²⁻

⁶³⁴. Les lncRNA peuvent également se fixer à des ARNm présentant une séquence complémentaire et agir sur diverses étapes de régulation post-transcriptionnelle comme l'épissage, l'édition des ARNm cibles, le transport, la traduction et la dégradation ⁶³¹.

2. Comparaison des rôles des ARNnc eucaryotes et procaryotes

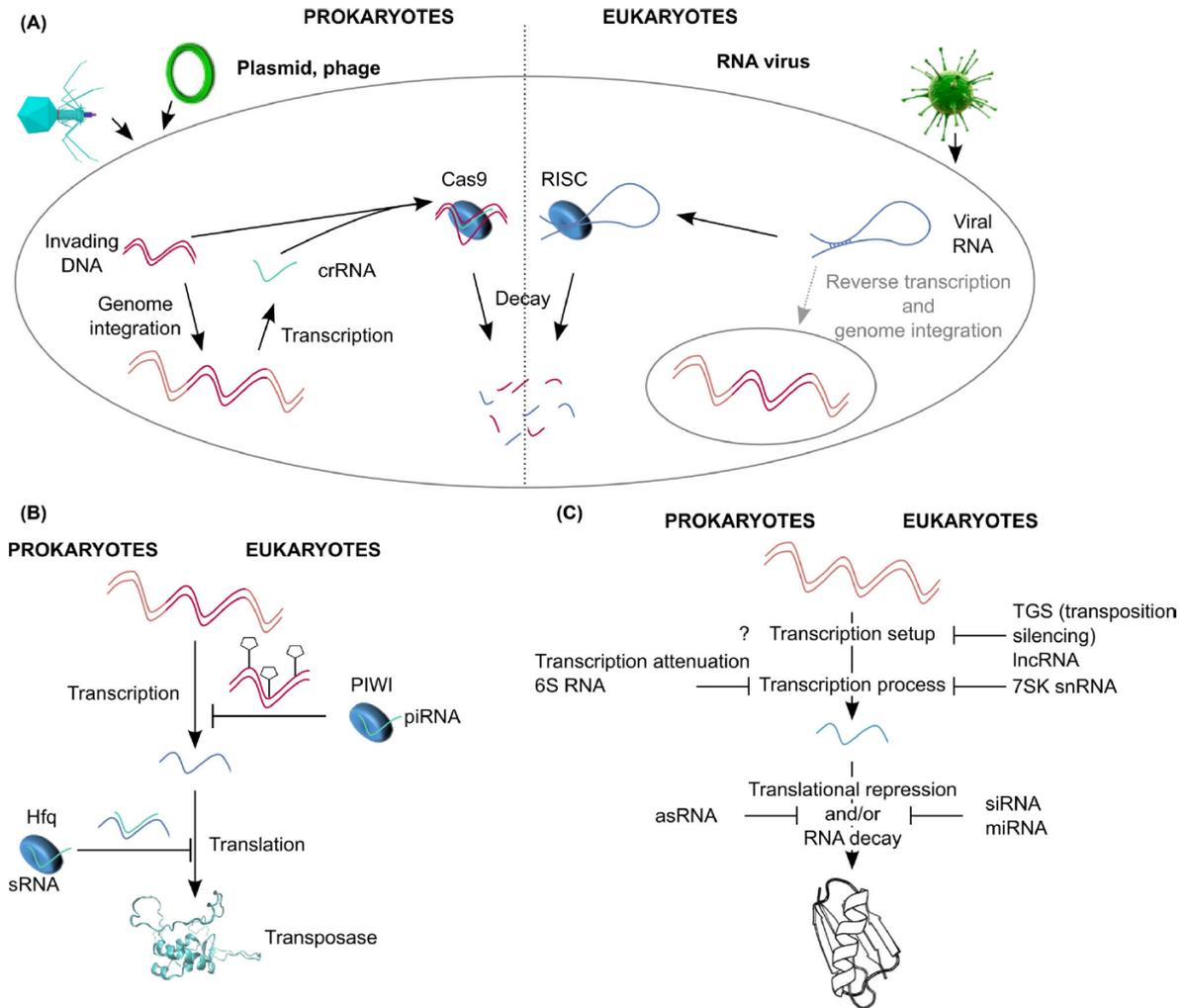


Figure 29 – Comparaison des mécanismes de régulation dépendant des ARNnc et leurs rôles chez les procaryotes et les eucaryotes. ⁶¹⁷

a. Protection contre les virus : l'immunité médiée par les ARN

Les cellules eucaryotes et procaryotes font face au cours de leur existence à de nombreux évènements infectieux par des virus (ou des phages chez les bactéries) et doivent être capables de se protéger en identifiant et neutralisant ces éléments parasites. Les bactéries ont développé tout un arsenal de défense contre ces envahisseurs comme les systèmes restriction-modification, le système CRISPR-Cas ou encore le système Abi, présentés plus haut. Le système d'ARN interférence (ARNi) des eucaryotes peut être dans ce cas comparable au système de défense CRISPR-Cas des procaryotes (**Figure 29A**). Pour

beaucoup d'eucaryotes, il s'agit de la première et de la principale ligne de défense antivirale. En effet, des souches de plantes, de la drosophile et de nématodes déficientes en ARNi sont extrêmement sensibles à différentes infections virales⁶¹⁷. Le système de défense ARNi contre les infections virales des mammifères a également été caractérisé chez la souris^{635,636}. Les mammifères ont également développé une défense antivirale alternative à l'ARNi : l'infection virale est stoppée quand la protéine kinase R est activée par un ARN double brin, entraînant l'arrêt de la machinerie traductionnelle et l'apoptose⁶³⁷. Cependant les contributions relatives de ces deux systèmes dans la défense contre les infections virales chez les mammifères restent encore débattues. Enfin, bien que des complexes ribonucléoprotéiques médient la dégradation des acides nucléiques étrangers à la fois chez les eucaryotes et chez les procaryotes, le système ARNi eucaryote ne peut fournir qu'une réponse immédiate visant à détruire l'ARN viral mais, contrairement au système CRISPR/Cas, ne présente pas de mémoire.

b. Contrôle des éléments génétiques mobiles

Les transposons sont présents dans tous les organismes et ont des effets majeurs sur leurs hôtes. Beaucoup de transposons ont développé des relations symbiotiques avec leur hôte afin que chacun puisse bénéficier de l'existence de l'autre⁶³⁸. Les transposons contribuent au *fitness* de leur hôte en influençant l'expression et la structure de ses gènes, en apportant des gènes passagers ou des séquences codant pour des régulateurs. Les eucaryotes et les procaryotes ont développé des mécanismes permettant de limiter la mobilité des transposons. Chez les bactéries par exemple, les ARNnc avec l'aide de la protéine Hfq bloquent la traduction des transposases en s'appariant avec leur ARNm⁶³⁹. Comme chez les procaryotes, la transposition est inhibée au niveau transcriptionnel chez les eucaryotes par les piRNAs et les complexes protéiques PIWI et AGO au niveau des lignées germinales. Ainsi, contrairement aux systèmes de défense antivirale siRNA-RISC qui dégrade l'ARN étranger, le système piRNA-PIWI va immobiliser le transposon et ne pas le détruire (**Figure 29B**)⁶¹⁷.

c. Régulation globale de l'expression des gènes

De nombreuses classes d'ARN participent à la régulation de l'expression de gènes

pendant ou après leur transcription. Comme détaillé lors de ce chapitre, les ARN régulateurs bactériens peuvent atténuer l'expression d'un gène en s'appariant avec l'ARNm de ce dernier afin d'agir sur sa transcription directement, sa stabilité ou sa traduction (**Figure 29C**). De nombreux ARN régulateurs bactériens régulent négativement l'initiation de la traduction en s'appariant directement au RBS de leur ARNm cible les rendant également plus sensibles aux attaques de ribonucléases. Moins fréquemment, les ARN régulateurs bactériens peuvent favoriser l'initiation de la traduction en modifiant certaines structures secondaires au niveau du RBS ou en stabilisant les ARNm⁴²². Ces mécanismes d'interférence transcriptionnels sont également retrouvés chez les eucaryotes. En effet, les ARN régulateurs eucaryotes (les siRNA et les miRNA) peuvent aussi agir sur la traduction et la dégradation des ARNm en faisant office d'ARN guide pour le complexe RISC contenant la protéine AGO. Les eucaryotes peuvent également moduler la transcription de leurs gènes par le biais de modifications épigénétiques comme des marques spécifiques de l'ADN (méthylation et hydroxyméthylation) et des histones (acétylation, méthylation, phosphorylation des lysines, sérines ou arginines) sans en modifier la séquence génétique. En interagissant avec les enzymes responsables des modifications de marquages des histones et de l'ADN, les ARN régulateurs participent largement à ces régulations épigénétiques en sélectionnant spécifiquement les régions de la chromatine à modifier permettant soit l'induction ou l'extinction de l'expression d'un gène⁶¹⁶. Les ARN régulateurs responsables de ces modifications épigénétiques peuvent être de diverses tailles, incluant donc les lncRNA (comme le transcrite inactivant le chromosome X [*XIST*] et l'ARNas de *HOX* [*HOTAIR*]) et les petits ARN (principalement les piRNA). Les ARN régulateurs eucaryotes peuvent également contrôler l'expression des gènes directement durant l'étape de transcription, comme l'ARN régulateur 7SK des vertébrés qui contrôle l'élongation de la transcription par l'ARN polymérase II (**Figure 29C**)⁶⁴⁰. Lors du développement des organismes multicellulaires, les cellules sont amenées à exprimer progressivement divers ensembles de gènes permettant la différenciation cellulaire. L'expression de ces répertoires spécifiques de gènes dépend du rôle qu'ils ont à jouer et est régulée de manière coordonnée avec les autres cellules. Ainsi, une mauvaise régulation de l'expression de ces gènes peut conduire à un phénotype défectueux de l'organisme. La construction de mutants de délétion chez le nématode

Caenorhabditis elegans a permis de mettre en évidence que la plupart des miRNA de *C. elegans* ne sont pas individuellement indispensables pour le développement ou la viabilité du nématode ⁶⁴¹. Des résultats contradictoires chez la drosophile ont révélé qu'une large majorité de mutants présentent des phénotypes visibles dont un retard développemental, une viabilité réduite, un nombre de cellules germinales altéré ou des déficiences de morphologies externes ⁶⁴². Chez la drosophile, les miRNA contrôlent donc individuellement les processus de prolifération cellulaire, de différenciation ou d'apoptose ⁶¹⁷. Certains gènes de miRNA ont également été inactivés chez la souris entraînant un grand nombre de conséquences dont la létalité embryonnaire, un défaut dans la spermatogenèse, une mauvaise différenciation des muscles lisses et une neurogenèse défectueuse ⁶¹⁷. Il semble donc que la régulation des gènes par les miRNA soit fondamentale dans le développement de la plupart des métazoaires.

d. Adaptation de l'expression des gènes à l'environnement

En plus de leur implication dans le contrôle global de l'expression des gènes, les ARN régulateurs permettent une réponse adaptative rapide d'une cellule en fonction de signaux qu'elle reçoit de son environnement. Cependant, bien que les ARN régulateurs soient présents chez tous les eucaryotes, peu d'études ont mis en évidence leur rôle dans l'adaptation aux stress environnementaux. De nombreuses études ont également mis en évidence l'implication des lncRNA dans des processus biologiques variés incluant les infections par des pathogènes. Certains lncRNA peuvent agir comme molécules de défense afin de prévenir la prolifération de pathogènes chez l'hôte, tandis que d'autres peuvent être utilisés par le pathogène lui-même pour favoriser sa propagation ⁶⁴³. L'implication des ARN régulateurs dans l'adaptation aux conditions environnementales est bien mieux caractérisée chez les procaryotes. Afin de pouvoir subsister à l'environnement hostile dans lequel elles vivent, les bactéries ont développé des processus d'adaptation afin d'assurer leur survie et leur dispersion. Agissant de concert avec des TCS, les ARN régulateurs sont capables d'intégrer un grand nombre de signaux environnementaux permettant aux bactéries de s'adapter à de nombreux stress ⁶⁴⁴. Les bactéries sont également capables d'adapter l'expression de leurs gènes en fonction de la présence d'autres bactéries de la même espèce

dans leur environnement. Le QS contrôle l'expression de centaines de gènes, permettant aux bactéries d'adopter certains modes de vies qu'elles ne pourraient pas avoir en tant que cellules individuelles (la bioluminescence, la sporulation, la compétence, la production d'antibiotiques, la formation de biofilms et la sécrétion de facteurs de virulence). Les ARN régulateurs jouent notamment un rôle très important dans la régulation du passage d'un mode de vie planctonique à un mode de vie sessile grâce à la détection de signaux environnementaux spécifiques pouvant induire la synthèse de flagelle pour la motilité ou la formation de biofilm. De plus, les bactéries ne sont pas seulement capables de détecter la présence de bactéries d'une même espèce mais également à la présence d'autres cellules. Chez de nombreux pathogènes, les ARN régulateurs et les *riboswitch* contrôlent des réseaux de gènes impliqués dans l'adaptation au stress, la communication cellule-cellule, les processus de phagocytose par les cellules immunitaires mais également le bon *timing* de la production de leur facteurs de virulence, leur permettant de survivre et de disséminer après infection de leur hôte ⁶⁴⁵.

Projet scientifique

Mon projet de thèse consiste en l'étude des aspects originaux du contrôle du cycle infectieux chez *C. difficile* par des ARNs. Il fait suite à la récente identification d'un grand nombre des ARNnc chez *C. difficile* suggérant l'importance et la diversité des mécanismes de régulation par des ARNs et s'inscrit dans un projet plus large d'étude des réseaux de régulation chez ce pathogène^{174,600,601}. Mon projet de recherche se subdivise en trois axes de recherches différents qui abordent ces processus de régulation de la virulence par ARNnc selon différents aspects.

Afin d'étudier la dynamique d'expression globales des ARNs au cours de l'infection, j'ai eu recours à une approche transcriptomique particulière, le *dual RNA-seq*, qui permet de suivre simultanément les changements d'expression des gènes chez *C. difficile* et de son hôte au cours de l'infection^{646,647}. Cette expérience a été réalisée dans un modèle murin conventionnel avec l'induction d'une dysbiose par traitement antibiotique afin de mimer chez l'animal les conditions de l'infection humaine. Les données obtenues ont permis d'identifier plusieurs gènes et ARNnc différentiellement exprimés chez *C. difficile* et chez la souris au cours de l'infection.

Parmi les ARNnc identifiés chez *C. difficile*, une centaine d'entre eux sont localisés dans des IGR et semblent nécessiter l'assistance de la protéine Hfq ce qui constitue une caractéristique unique de cette protéine chez *C. difficile* en comparaison à celles des autres bactéries à Gram positif^{600,602,604}. Je me suis donc intéressé au rôle de la protéine Hfq dans la régulation des processus de virulence chez *C. difficile*. Pour cela j'ai réalisé plusieurs expériences de caractérisation de phénotypes associés à la virulence chez un mutant de délétion pour cette protéine, ainsi qu'une expérience en modèle animal (souris) afin d'étudier l'impact de la délétion de la protéine Hfq sur la virulence de *C. difficile*.

Enfin, le dernier axe de mon projet consiste en l'étude d'un ARNnc spécifique de la souche hypervirulente R20291 de PCR-ribotype 027 qui a été identifié lors d'une étude comparative RNA-seq avec la souche de référence CD630E. La structure et la localisation de la séquence de cet ARNnc (*IGR-1461*) suggèrent une fonction régulatrice en *cis* de type *riboswitch* sur le gène situé en aval. Plusieurs expériences ont été réalisées afin de caractériser ce nouvel élément régulateur spécifique de la souche hypervirulente R20291.

Résultats

Partie 1 : Dynamique d'expression des ARNnc chez l'hôte et le pathogène au cours d'une ICD

I. Principe et mise au point de la technique pour *C. difficile*

1. Principe de la technique de dual RNA-seq

Les études transcriptomiques des interactions hôte-pathogène ont pendant longtemps été limitées à l'analyse des changements d'expression des gènes soit de la bactérie, soit des cellules hôtes infectées. Le développement de la technique de dual-RNA-seq a permis de dépasser ce niveau de simplicité en donnant un aperçu simultané des transcriptomes de la bactérie et de l'hôte⁶⁴⁶. Une expérience classique de dual-RNA-seq se divise en plusieurs étapes : (1) une étape expérimentale où les deux espèces (eucaryote et bactérie) interagissent, (2) une lyse cellulaire, (3) la purification des ARN totaux des deux espèces mélangées, (4) la préparation des banques et (5) leur séquençage haut débit. Le point de départ de cette expérience est la préparation des échantillons. La collecte des échantillons, la lyse cellulaire, la purification des ARN et la synthèse de ADNc constituent les étapes déterminantes pour la précision et la fiabilité des résultats finaux. La procédure d'extraction des ARN doit être suffisamment rapide pour éviter à la fois les changements de profils transcriptionnels et les dégradations des ARN par les RNAses (bien que le maintien à 4°C et un traitement chimique stabilisateur des ARN puissent amoindrir ces deux problèmes)⁶⁴⁸. Les *reads* obtenus après le séquençage doivent être traités en suivant les étapes classiques d'analyse de données, c'est-à-dire un contrôle qualité avec élimination des *reads* de basse qualité, une cartographie sur les génomes de référence de l'hôte et du pathogène, un comptage des *read* pour l'analyse de l'expression différentielle des gènes (EDG), puis enfin l'analyse fonctionnelle des gènes différentiellement exprimés.

Au cours de ce projet de thèse, la technique de *dual RNA-seq* a été appliquée à *C. difficile* afin d'identifier la dynamique d'expression des ARNs du pathogène et de son hôte simultanément au cours de l'infection (**Figure 30**). Il s'agit d'un projet réalisé en collaboration avec l'équipe BaPS de Micalis (Faculté de Pharmacie de Chatenay-Malabry, Claire Janoir) et l'équipe de bio-informatique de l'I2BC (Daniel Gautheret). Bien que plusieurs études *in vivo* de transcriptomique aient déjà été réalisées sur *C. difficile* et sur la souris

séparément, aucune jusqu'à présent ne s'est intéressée à la dynamique d'expression des ARN (non codant compris) simultanément chez le pathogène et l'hôte^{120,649–653}. Cette étude pionnière dans son genre a donc nécessité plusieurs étapes de mise au point technique pour l'extraction puis le séquençage simultané des ARN totaux des deux espèces.

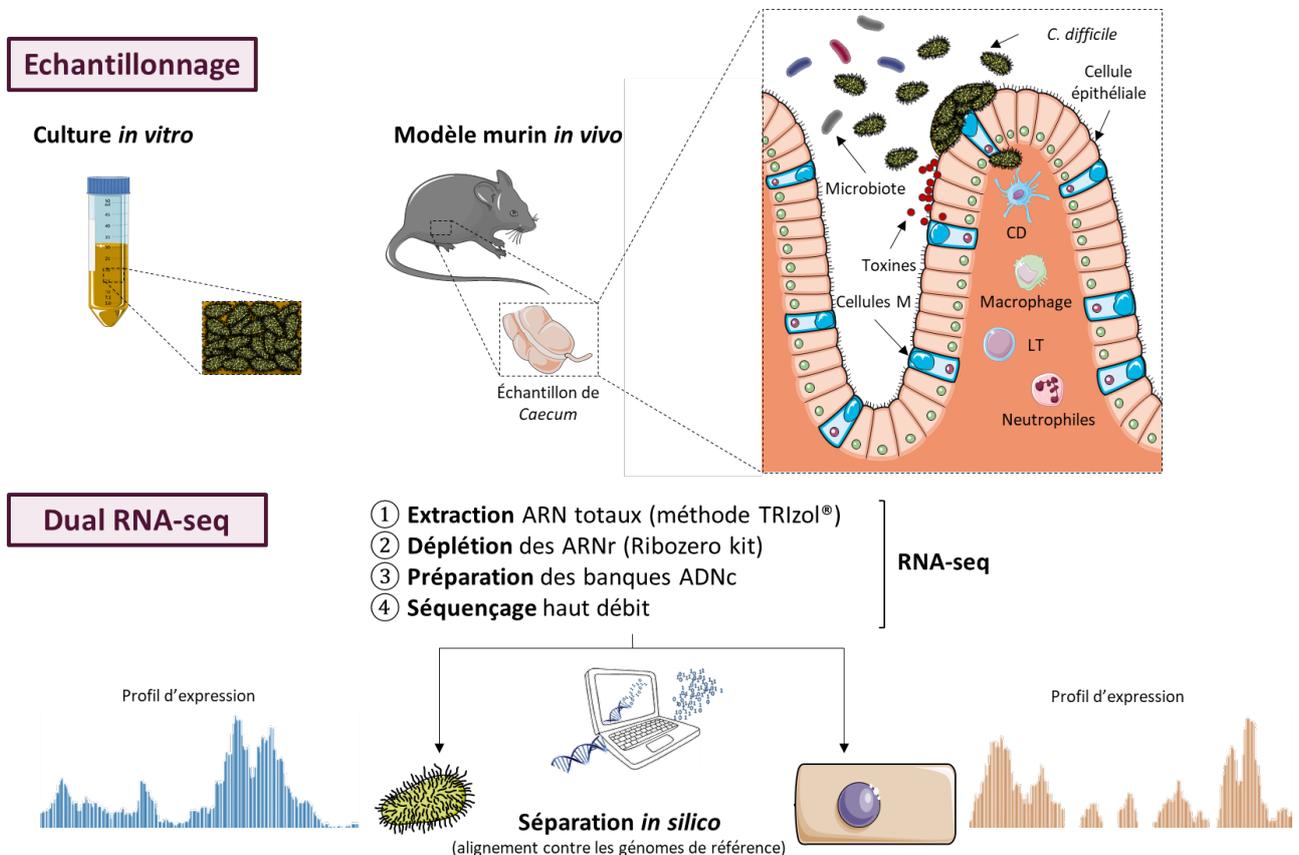


Figure 30 – Aperçu des étapes de l'expérience de dual RNA-seq adapté à *C. difficile*. (LT = Lymphocytes T ; CD = Cellules Dendritiques)

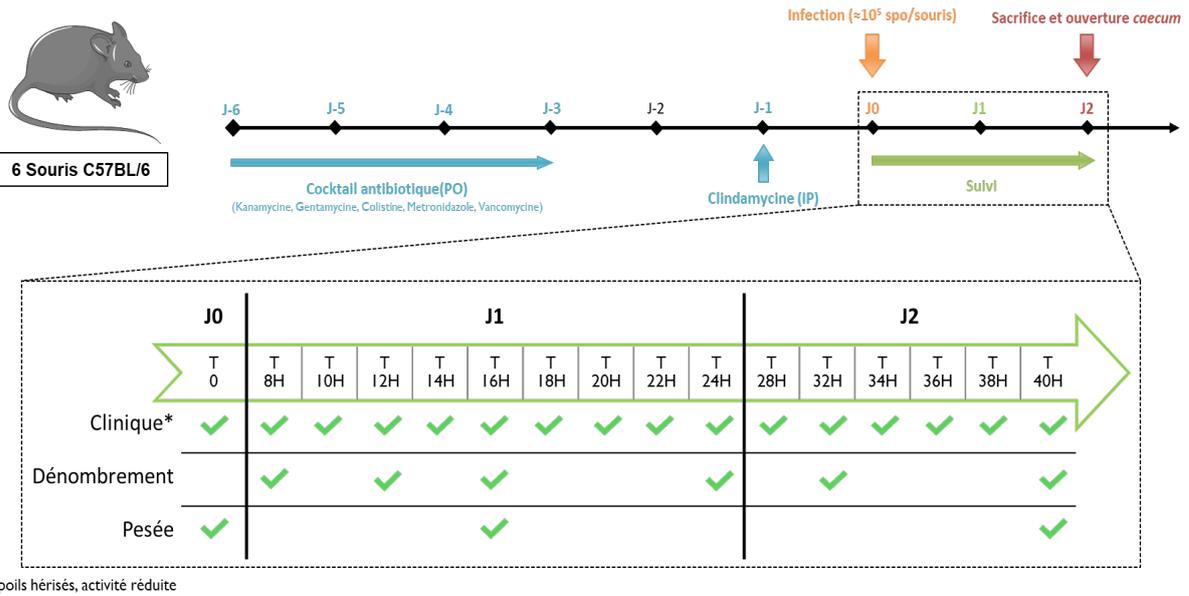
Le modèle d'infection de *C. difficile* utilisé lors de ce projet est celui proposé par Chen et al.⁶⁵⁴ faisant usage de souris conventionnelles C57BL/6 (ou *black 6*) traitées par un cocktail de cinq antibiotiques (kanamycine 0,4 mg/ml ; gentamycine 0,035 mg/ml ; colistine 850 U/ml ; métronidazole 0,215 mg/ml ; vancomycine 0,045 mg/ml), administrés dans l'eau de boisson pendant trois jours, puis par une injection intra-péritonéale de clindamycine (10 mg/kg) deux jours plus tard avant d'être infectées un jour après par une suspension de spores de la souche CD630E de *C. difficile*. L'administration d'antibiotiques en amont de l'infection permet de mimer chez l'animal les conditions de l'infection humaine en induisant une dysbiose du microbiote intestinal des souris permettant la colonisation par *C. difficile*. Toutes les souris, infectées ou contrôles, reçoivent systématiquement le même traitement

antibiotique. Les essais animaux ont été réalisés en accord avec les règles d'utilisation d'animaux de laboratoires et les procédures ont été élaborées dans le respect de la règle des 3R (Réduire, Raffiner, Remplacer). Les protocoles ont été approuvés par le Comité d'éthique en expérimentation animale C2EA-26 (n° d'autorisation APAFIS#4617-2016032118119771 vI). Des souris conventionnelles C57Bl/6, âgées de 6 à 7 semaines, provenant des Laboratoires Charles River (France), ont été utilisées dans les trois essais animaux décrits ci-dessous. Les souris étaient hébergées dans l'animalerie de la faculté de Pharmacie de l'Université Paris-Saclay (n° d'agrément B92-019-01), par groupe de 3 souris par cage, et ont eu accès pendant toute la durée des expériences à de l'eau et de la nourriture stérile *ad libitum*.

2. Expérience de suivi clinique de souris

Dans un premier temps, une étude de suivi clinique a été réalisée sur six souris infectées avec 10^5 spores de *C. difficile* (CD630E) par souris afin d'établir les conditions optimales d'infection permettant à la fois d'observer l'apparition des premiers signes cliniques de l'infection mais également de récupérer suffisamment de contenu caecal pour la purification des ARN bactériens lors de l'expérience de dual RNA-seq. L'évolution des symptômes, de la colonisation et du poids des souris infectées a été suivie durant 40 heures avant le sacrifice des animaux, et le prélèvement des *caeca* (**Figure 31A**). Les premiers symptômes sont apparus à partir de 32h se traduisant par des selles molles chez toutes les souris, le poil légèrement hérissé et la moitié des souris présentaient une activité réduite. Les symptômes se sont intensifiés jusqu'au sacrifice des animaux à 40h post-infection où les *caeca* ont été prélevés et analysés. Les *caeca* présentaient, pour la plupart, peu de contenu et un aspect hémorragique indiquant une inflammation (**Figure 31B**). Ce tableau clinique s'accompagnait d'une perte de poids marquée chez toutes les souris. Le suivi de la colonisation des formes végétatives de *C. difficile* dans les fèces a mis en évidence un plateau de colonisation atteint dès T8h post-infection avec environ 10^8 UFC/g de fèces.

A



B

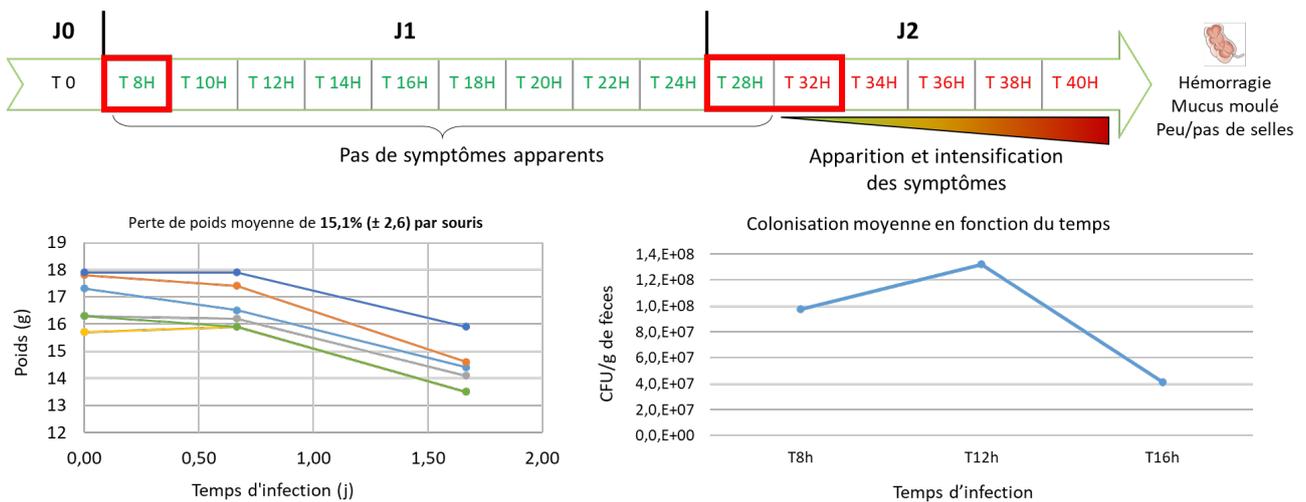


Figure 31 – Aperçu de l’essai clinique et de ses résultats. (A) *Workflow* de la première expérimentation animale réalisée sur six souris conventionnelles femelles C57BL/6 ayant reçu un cocktail d’antibiotiques par voie orale (kanamycine, gentamycine, colistine, métronidazole, vancomycine) et une injection intra-péritonéale de clindamycine en amont de l’infection (10^5 spores de la souche CD630E par souris) afin d’induire une dysbiose favorisant l’implantation de *C. difficile*, comme décrit dans la littérature⁶⁵⁴. Les différents points d’observation clinique, de dénombrement de cellules végétatives à partir des fèces et de pesée des souris sont indiqués dans le tableau sous la frise chronologique. (B) L’état de santé général des souris a été défini par un score prenant en compte la qualité des fèces, l’apparence des poils (hérissés ou non) et leur activité. L’évolution globale de ce score est schématisée dans la frise chronologique par le code couleur suivant : les points pour lesquels aucun symptôme n’a été observé sont représenté en vert ; les points lors desquels des symptômes sont apparus sont représenté en rouge. Les temps 8h, 28h et 32h encadrés en rouge correspondent aux temps sélectionnés pour l’expérience de dual RNA-seq. Sont présentés sous la frise chronologique les graphiques représentant l’évolution du poids pour chaque souris et la colonisation moyenne par *C. difficile*.

Cette étude a donc permis de sélectionner trois points au cours de la cinétique d'infection pour l'extraction des ARNs durant l'expérience de dual RNA-seq : un point précoce à T8h post-infection auquel le plateau de colonisation est atteint, un point à T28h post-infection pour lequel les souris ne sont pas encore symptomatiques et un point à 32h post-infection où les premiers symptômes apparaissent. Ce dernier point a été choisi relativement précocement afin d'être certain de pouvoir récupérer assez de matériel caecal pour l'extraction des ARN totaux bactériens. En effet, la diarrhée qui survient plus tardivement entraîne la vidange du contenu caecal et cela se traduit par une quantité trop faible de bactéries pour obtenir des ARNs en quantité suffisante pour faire du RNA-seq.

3. Mise au point de l'expérience de dual RNA-seq et de purification des ARNs

À partir de ces résultats, une deuxième expérimentation animale a été réalisée pour l'extraction des ARN totaux bactériens et de souris pour le dual-RNA-seq. Au total 18 souris ont été utilisées. Toutes ont reçu le même traitement antibiotique que lors de l'essai de suivi clinique, puis les souris ont été réparties de manière aléatoire dans 6 groupes de 3 souris chacun. Neuf souris, soit 3 groupes, ont été infectées par des spores de *C. difficile* (10^5 spores par souris administrées par gavage oral) et les autres animaux ont reçu une solution d'eau saline à la place des spores. Trois souris infectées et trois souris non infectées ont été sacrifiées à chacun des temps d'infection choisis lors du suivi clinique (T8h, T28h et T32h) afin de prélever leurs *caeca*. Les *caeca* ont été prélevés immédiatement après euthanasie, ouverts sur leur longueur et une spatule de contenu *caecal* a été prélevée pour le dénombrement des formes végétatives luminales de *C. difficile*. Immédiatement après, les *caeca* ouverts et leur contenu sont plongés dans 10 mL d'une solution stabilisatrice des ARN (RNAlater), nettoyés des impuretés (par sédimentation et centrifugation) puis ont été lysés par FastPrep et traités afin de purifier les ARNs totaux (protocole TRIzol). Après l'obtention des échantillons ARN « dual » (eucaryotes et procaryotes mélangés), un séquençage test a été effectué afin de déterminer la part relative des ARN bactériens et eucaryotes obtenus après extraction. Ce premier run a permis d'obtenir environ 3 000 000 *reads* eucaryotes uniques et environ 3 000 *reads* procaryotes uniques s'alignant avec leurs génomes de référence respectifs. La faible proportion d'ARN procaryotes dans le dual RNA-seq est très

souvent décrite dans la littérature et présente l'un des principaux défis de cette technique⁶⁴⁷. Dans de nombreuses études, afin de pallier la très faible proportion de *reads* bactériens obtenus, un enrichissement artificiel en ARN bactériens peut être réalisé à l'aide d'un kit de déplétion d'ARN eucaryotes (MICROBEnrich). Un tel traitement a donc été effectué sur les échantillons ARN obtenus, mais n'a pas permis d'obtenir une augmentation significative de la proportion d'ARN bactérien, comme établi par un nouveau séquençage. Lors de cette expérience, la lyse des cellules totales (cellules épithéliales murines et bactéries) et l'extraction des ARN eucaryotes et procaryotes ont été réalisées simultanément. Une lyse incomplète des cellules bactériennes a donc été envisagée et, afin d'y remédier, nous avons décidé de réaliser séparément la lyse et la purification des ARN eucaryotes et procaryotes pour l'expérience de dual RNA-seq.

Une seconde expérimentation animale a donc été réalisée avec des conditions d'infection identiques à la précédente (**Figure 32A**), mais des conditions d'extraction d'ARN différentes : cette fois-ci, les *caeca* des souris et leur contenu ont été lysés séparément, et les ARNs présents dans ces lysats ont été purifiés également séparément. Les étapes de préparation des *caeca* et de leurs contenus sont détaillées dans la **Figure 32B**. La numération des formes végétatives de *C. difficile* à partir d'un aliquot de contenu caecal des 9 souris infectées a été réalisé par étalement des dilutions sur un milieu sélectif pour *C. difficile* (gélose CLO, BioMérieux). Les boîtes ont été incubées 48h en anaérobiose (5% H₂, 5% CO₂, 90% N₂) à 37°C puis les UFC ont été dénombrées. Les résultats confirment que les souris sont toutes colonisées par le pathogène avec un niveau de colonisation moyen d'environ 10⁸ CFU/g de contenu aux 3 temps (moyenne ± écart-type de 1,63x10⁸ ± 6,15x10⁷, 1,41x10⁸ ± 2,25x10⁷ and 9,67x10⁷ ± 1,3x10⁸ UFC/ g de contenu à T8h, T28h and T32h post-infection, respectivement). À l'exception des souris sacrifiées à T32h, les variations inter-individuelles de niveau de colonisation chez les souris d'un même groupe sont faibles (résultats détaillés, **Tableau S1**). À la fin du processus de purification, deux échantillons ont donc été obtenus pour chaque souris : un échantillon d'ARN (principalement) eucaryotes et un échantillon d'ARN (principalement) procaryotes. La concentration des échantillons ARN procaryotes était environ 5 à 10 fois plus faible que celle des échantillons eucaryotes, avec une moyenne

de 120 ng/μL et 1 200 ng/μL respectivement (résultats détaillés, **Tableau S2**). Les échantillons ont été mélangés pour chaque souris avant le séquençage dans des proportions contrôlées permettant de maximiser la quantité d'ARN procaryotes séquencés. Afin de s'assurer d'une couverture suffisante du génome de la souris (\approx 3,4 Gb) plus grand que celui de *C. difficile* (\approx 4,3 Mb), le ratio de mélange choisi pour chaque échantillon « dual » était de 2 ARN eucaryotes pour 1 ARN procaryote. Pour chaque échantillon, les ARNr ont été déplétés avec un kit Ribozero. Quinze banques d'ADNc ont été ensuite préparées avec un kit *TruSeq Stranded mRNA* (Illumina) : neuf banques pour les souris infectées comprenant un triplicat par temps d'infection (T8h, T28h et T32h), trois banques pour les souris non infectées (pour minimiser les coûts et le nombre d'échantillons à traiter seul le temps contrôle T8h pour les souris non infectées a été séquencé bien que des échantillons contrôles aient été obtenus pour chacun des trois points d'infection), et trois banques pour les échantillons *in vitro* obtenus après une culture de 8h en milieu BHI (*Brain Heart Infusion*). Les échantillons ont été multiplexés et séquencés avec un système NextSeq 500 (Illumina) en *paired end* (PE) 50-35 nt (75 cycles). Entre 100M et 160M de *reads* PE par échantillon ont été obtenus (**Tableau 1**).

Tableau 1 – Nombre de *reads* totaux (PE) obtenus pour chaque échantillon.

Conditions expérimentales	Nom de l'échantillon	Nombre total de <i>reads</i> (PE)
<i>In vitro</i>	IV1	79 628 801
	IV2	81 917 460
	IV3	158 092 060
Infection 8h	S1	132 949 897
	S2	125 441 845
	S3	102 737 714
Infection 28h	S4	110 603 227
	S5	122 935 683
	S6	172 505 220
Infection 32h	S7	143 164 258
	S8	164 233 996
	S9	92 981 633
Non infectées	S10	148 203 445
	S11	117 889 871
	S12	148501662

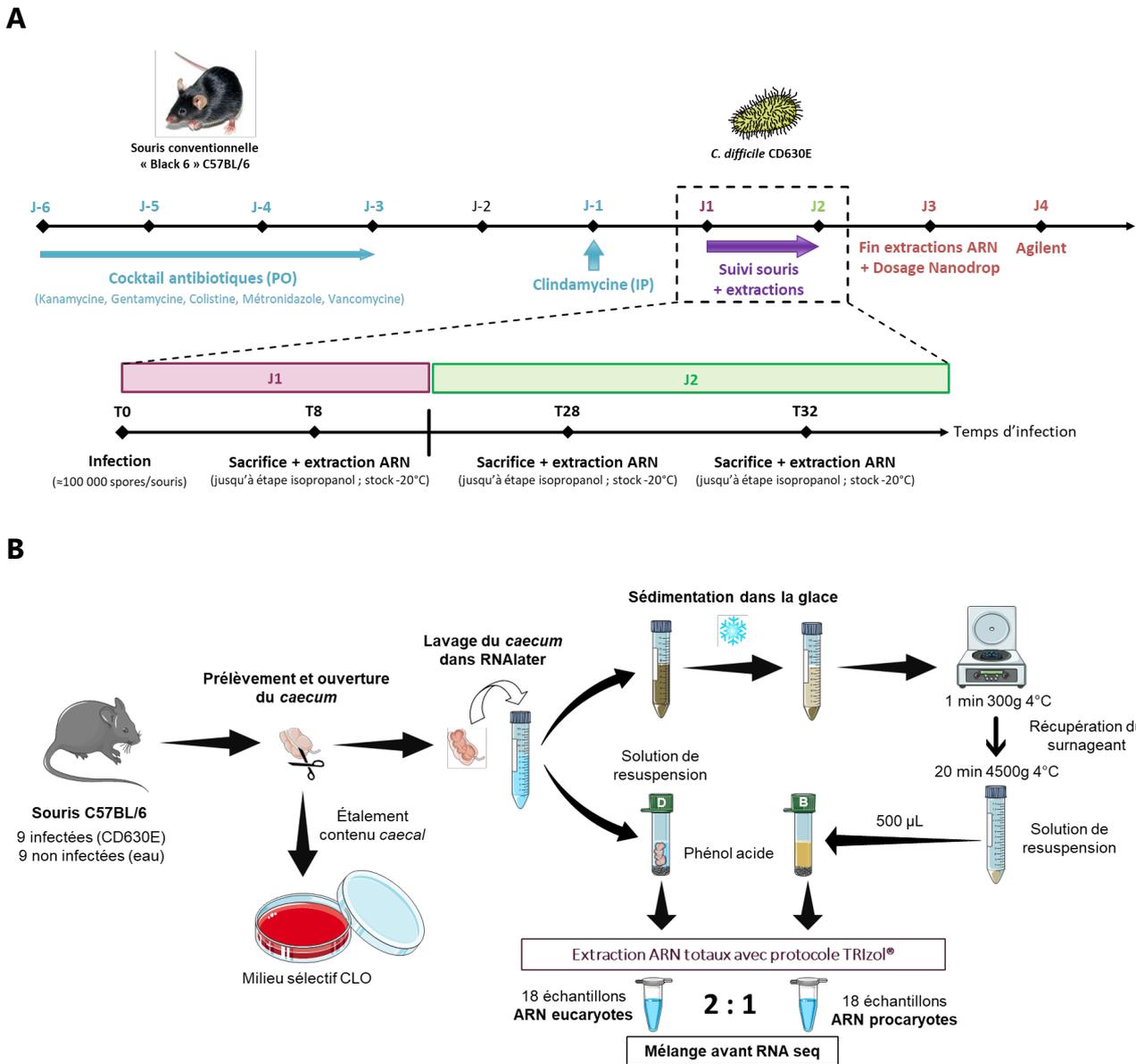


Figure 32 – Workflow de l'expérience de dual RNA-seq et étapes du protocole d'extraction des ARN totaux. (A) Le modèle d'infection utilisé est le même que celui utilisé pour l'expérience de suivi clinique⁶⁵⁴. La totalité des souris, y compris les souris contrôles ont reçu le même traitement antibiotique en amont de l'infection. Neuf souris ont été infectées avec une solution de 10^5 spores de la souche CD630E et neuf souris contrôles ont reçu de l'eau stérile. Les temps 8h, 28h et 32h de sacrifice et d'extraction des ARN ont été choisis à la suite de l'expérience de suivi clinique décrite précédemment. (B) Pour chaque temps, après le sacrifice, les souris (infectées et non infectées) ont été traitées selon le même protocole. Après le prélèvement et l'ouverture du *caecum*, un échantillon de contenu caecal est prélevé pour dénombrer les formes végétatives de *C. difficile*. Le *caecum* (tissu et contenu) est ensuite rapidement placé et lavé dans 10 mL de RNAlater afin de stabiliser les ARN puis, une fois séparé de son contenu, la moitié du tissu est placée dans un tube contenant 500 µL de phénol acide, 500 µL d'une solution aqueuse et une matrice de lyse D (billes de 1,4 mm). Le contenu caecal récupéré dans le RNAlater est ensuite traité par sédimentation et centrifugation afin d'éliminer les matières fécales et de ne récupérer que les bactéries du microbiote. Le culot bactérien est remis en suspension dans 500 µL d'une solution aqueuse et ajouté à 500 µL de phénol acide dans un tube contenant une matrice de lyse B (billes de 0,1 mm). Les ARN totaux ont été purifiés à partir des lysats séparés du *caecum* et des bactéries avec un protocole TRIzol (Sigma) comme décrit précédemment⁶⁵⁵. Les échantillons ARN eucaryotes et procaryotes ont été mélangés avec un ratio 2 : 1 respectivement avant le séquençage.

II. Analyse des données obtenues

1. Contrôle qualité et alignement des *reads*

Les données brutes obtenues après séquençage ont été analysées selon le *workflow* présenté en **Figure 33**.

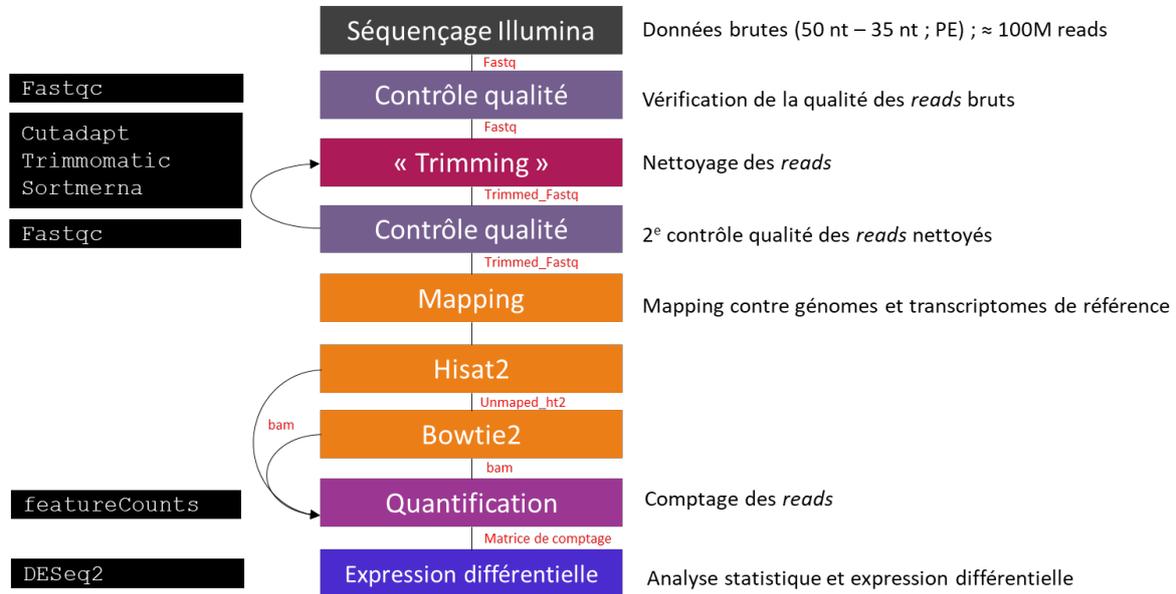
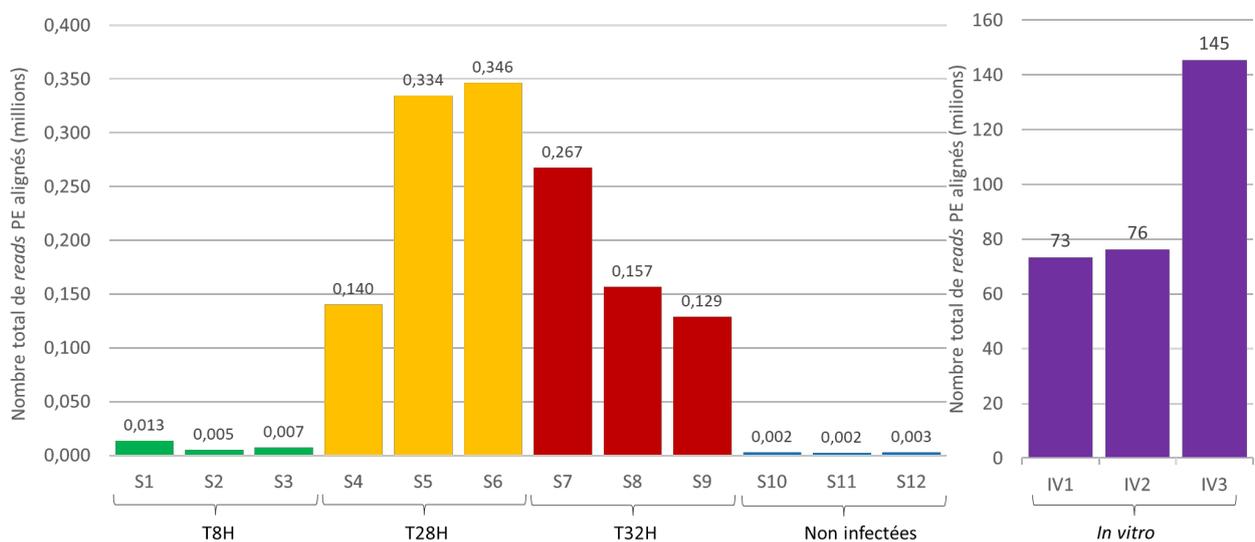


Figure 33 – Workflow d'analyse des données après récupération des *reads* (PE) bruts.

Les *reads* ont d'abord été testés pour leur qualité par FastQC puis les *reads* de trop courte taille (< 10 nt) et de basse qualité ont été supprimés par le programme Trimmomatic⁶⁵⁶ (**Figures S1 et S2**). Les fichiers fastq ont également dû être traités par sortmeRNA⁶⁵⁷, un programme permettant de filtrer les ARNr, en raison d'un nombre non négligeable de *reads* d'ARNr pour la plupart des échantillons. L'étape de cartographie des *reads* nettoyés sur les génomes de référence a été réalisée à l'aide de deux programmes d'alignement (HISAT2⁶⁵⁸ et Bowtie2⁶⁵⁹) afin de maximiser le pourcentage de *reads* s'alignant sur le génome de *C. difficile*. En effet, le facteur limitant dans le dual RNA-seq étant généralement le nombre de *reads* obtenus pour le pathogène, nous avons optés pour la cartographie en deux étapes. Le programme HISAT2, capable de résoudre les jonctions exons-introns chez la souris a été utilisé en premier puis, les *reads* qui n'ont pas été alignés par ce programme ont été récupérés et traités avec le second programme d'alignement Bowtie2. Ce processus de *mapping* en deux étapes a permis d'obtenir entre 0,1% et 0,3% de *reads* s'alignant sur le génome NC_009089 de la souche CD630E *C. difficile* (soit en moyenne 155 000 *reads* par échantillon) (**Figure 34A**) et entre 20% et 70% de *reads* s'alignant sur le génome GRCm39

de la souris (soit en moyenne 54 millions de *reads* par échantillon) (**Figure 34B**). Bien que le plateau de colonisation ait été atteint au temps T8h (**Tableau S2**), le nombre de *reads* s'alignant sur le génome de *C. difficile* est beaucoup plus faible que pour la souris. Cependant, le nombre de *reads* de *C. difficile* provenant des échantillons de culture *in vitro* est beaucoup plus élevé et a permis d'aligner environ 90% des *reads* des échantillons sur le génome de *C. difficile* (soit entre 70 et 145 millions de *reads* par échantillon) (**Figure 34A**). Afin de pouvoir normaliser ces données avec les autres échantillons contenant beaucoup moins de *reads* de *C. difficile*, un échantillonnage aléatoire représentatif de 0,5% des *reads* alignés a été réalisé sur les échantillon *in vitro*.

A



B

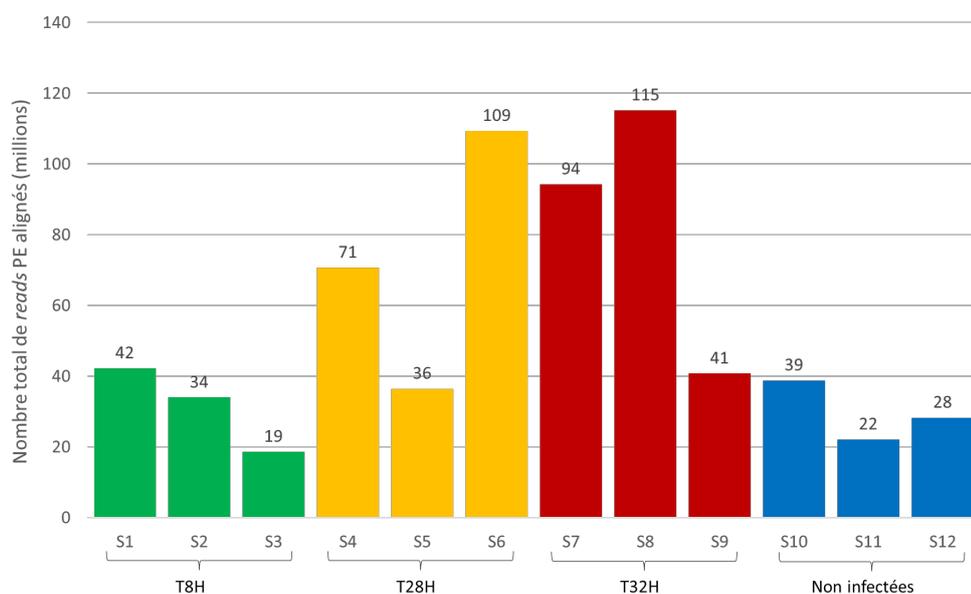


Figure 34 – Nombre de *reads* PE s'alignant sur les génomes de *C. difficile* (A) et de la souris (B).

D'autre part, après cartographie des *reads* sur les génomes de *C. difficile* et de la souris, il restait en moyenne 78 millions de *reads* correspondant aux *reads* des autres microorganismes du microbiote (**Figure S3**). Ces *reads* ont été analysés par le programme kraken2⁶⁶⁰ avec six bases de données de la RefSeq incluant des bactéries, champignons, virus, plasmides, archées et protozoaires afin d'avoir un aperçu des différents profils de microbiotes intestinaux des souris infectées par rapport aux souris non infectées. Dans la mesure où les données traitées ne sont pas de l'ARNr 16S mais des ARN totaux (dépourvus d'ARNr), la quantification des différentes espèces microbiennes se fait en pourcentage de transcrits (**Figure 35**).

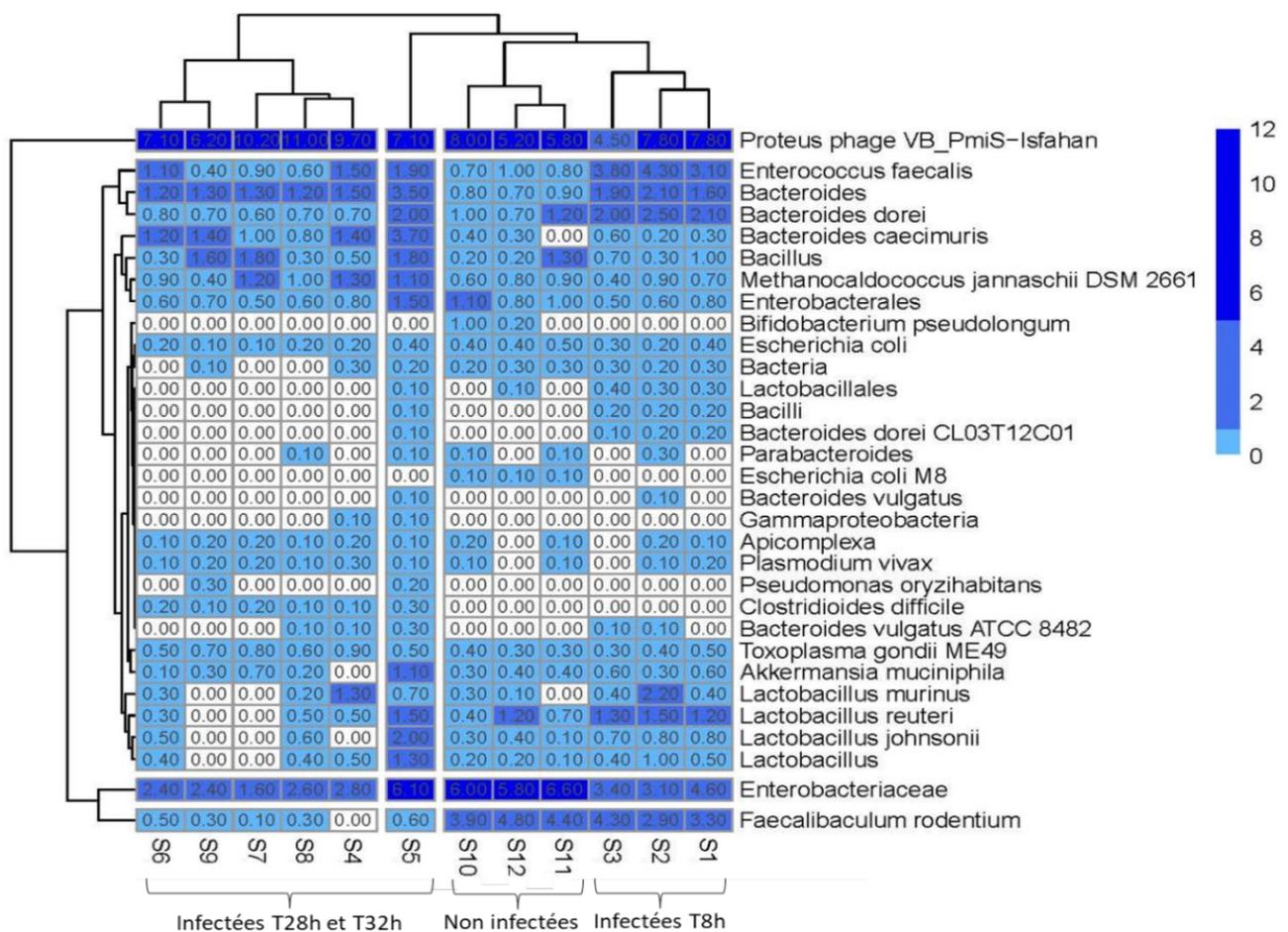


Figure 35 – Taxonomie des espèces microbiennes du microbiote basée sur le pourcentage de transcrits totaux par échantillon. Le pourcentage de transcrit de chaque espèce pour un échantillon donné est indiqué dans les cases. Seules les espèces présentes dans au moins un échantillon avec une abondance de transcrit supérieur à 0,5% sont représentées.

Cela ne permet donc pas de conclure quantitativement sur la composition du microbiote entre les différents souris puisque le pourcentage de transcrits par espèce peut aussi bien refléter une activité transcriptionnelle importante qu'un nombre élevé

d'individus pour une espèce donnée (ou les deux). Ces données ont cependant permis de renforcer la validation du modèle expérimental utilisé puisque l'on peut constater d'une part que les souris non infectées sont regroupées ensemble avec les souris infectées au temps T8h (ce qui peut s'expliquer par le très faible nombre de *reads* de *C. difficile* obtenu à ce temps d'infection) et, d'autre part, que les souris infectées sont regroupées entre elles (**Figure 35**), à l'exception de la souris S5 correspondant au temps d'infection T28h.

2. Quantification et analyse différentielle des données

Une fois les *reads* alignés sur leurs génomes de référence, un processus de quantification a été réalisé avec le programme featureCounts⁶⁶¹. Ce programme permet d'obtenir pour chaque échantillon séquencé une matrice de comptage déterminant le nombre de *reads* s'alignant sur chacun des *loci* annotés des génomes de référence. Ainsi deux matrices de comptages sont obtenues (une pour la souris et une pour *C. difficile*) pour chaque échantillon séquencé. Ces matrices correspondent chacune à une condition biologique particulière, et ont par la suite été comparées entre elles afin de pouvoir analyser l'expression des gènes différentiellement exprimés pour chacun des deux organismes étudiés. Pour cela un *pipeline* en R (SARTools) basé sur le programme DESeq2 a été utilisé⁶⁶². Il permet de comparer les matrices de comptages deux à deux et d'analyser statistiquement leurs variations de comptages⁶⁶². Plusieurs modèles comparatifs ont été analysés afin de pouvoir répondre à différentes questions biologiques. Parmi ces modèles seulement un pour *C. difficile* et trois pour la souris ont été estimés statistiquement significatifs et ont été retenus pour la prochaine publication.

Les profils d'expression des gènes des conditions *in vitro* ont été comparés à ceux des conditions d'infection *in vivo* (MI pour *Mouse Infected*). Le nombre de *reads* trop faible des échantillons extraits à T8h (en moyenne 250 000 *reads* par échantillon) en comparaison à celui des échantillons T28h et T32h n'a pas permis l'utilisation de ces données lors de l'analyse différentielle. Seules les données obtenues à T28h et T32h ont été conservées et analysées. La comparaison des profils d'expression des gènes aux temps T28h et T32h a révélé qu'il n'existe aucune différence significative du profil d'expression des gènes entre ces deux points ce qui a donc permis de les combiner comme une seule et même condition (MI)

pour leur analyse lors des simulations ultérieures afin de gagner en puissance en statistique (**Figure S4**). De plus, l'analyse en composante principale (ACP) semble également suggérer une similitude des échantillons de ces deux groupes qui sont en revanche séparés à 78% des échantillons *in vitro* par l'axe PC1 (**Figure S5**). Ainsi les échantillons issus des cultures *in vitro* ont été comparés aux échantillons prélevés à partir des souris infectées (T28h+T32h). Cette analyse a permis d'identifier 1372 gènes significativement (p value < 0.05) différentiellement exprimés dont 618 induits et 754 réprimés dans des conditions d'infection par rapport à des conditions de culture *in vitro* (avec un seuil de $\text{Log}_2\text{FoldChange}$ > 1 ou < -1 respectivement) (**Figure 36A**). Des catégories fonctionnelles ont été attribuées aux gènes différentiellement exprimés avec le programme COG (*Cluster of Orthologous Groups*) (**Figure 36B**).

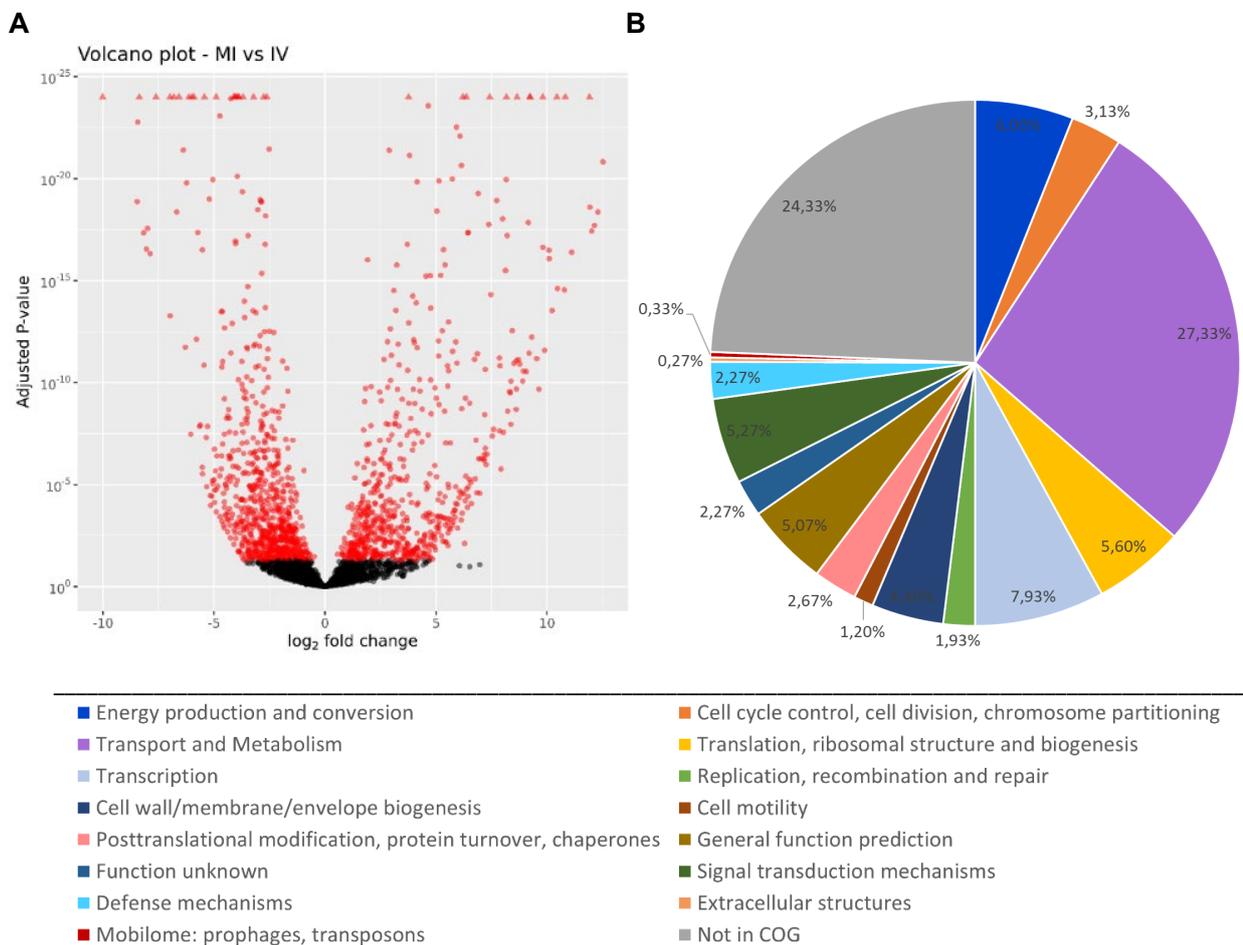


Figure 36 – Gènes différentiellement exprimés entre les groupes de souris infectées à T28h et T32h (MI) et les cultures *in vitro* de *C. difficile* (IV). (A) *Volcano plot* représentant le logarithme de la p -value ajustée en fonction du ratio logarithmique (Log_2 fold change ; FC) des gènes différentiellement exprimés entre les deux conditions. Les points rouges correspondent aux gènes significativement différentiellement exprimés (la partie droite correspond aux gènes induits dans les conditions MI ; la partie gauche correspond aux gènes réprimés dans les conditions MI). Les triangles correspondent aux gènes qui ont un Log_2FC trop bas ou trop haut pour être représenté sur le graphique. (B) Groupes fonctionnels des gènes différentiellement exprimés identifiés par COG. Les différents groupes liés au transport et au métabolisme ont été fusionnés en un seul groupe. La correspondance entre les couleurs et les groupes fonctionnels est précisée dans la partie basse de la figure.

Parmi les gènes induit lors de l'infection, on retrouve notamment certains facteurs de virulence comme les toxines TcdA et TcdB, ce qui est similaire avec les précédents essais de transcriptomique *in vivo* réalisées chez la souris^{652,653,663} ou chez le porc⁶⁶⁴ ainsi que l'adhésine CwpV (CD630_05140) qui favorise l'auto-agrégation de *C. difficile*^{356,665} et qui est surtout un élément de résistance vis-à-vis des phages. On retrouve également l'induction de certains gènes codant des protéines de l'assemblage flagellaire comme FliQ (CD630_02610), FlhA (CD630_02630), FliH/FliB (CD630_02620) et MotA (CD630_02560). Toutefois, le gène codant la sous-unité majeure du flagelline, *fliC*, est quant à lui sous-exprimé *in vivo* par rapport à la culture *in vitro*, de même que le gène *flgK* codant une protéine de jonction entre le crochet et le filament flagellaire. De ce fait, l'expression d'un flagelle fonctionnel et la motilité bactérienne qui en résulte n'est pas certaine à cette étape de l'infection. En revanche, l'expression de la plupart des gènes de l'opéron codant la synthèse du *pilus* de type IV (CD630_31030-35130) est diminuée *in vivo* d'environ un facteur 8 ($\text{Log}_2\text{FC} \approx 3$), notamment le gène de la piline majeure PilA1. Il faut noter qu'*in vitro*, l'expression du flagelle et du pilus est sous la dépendance de *riboswitch* antagonistes dépendant du c-di-GMP, un *riboswitch* de type I¹⁷⁵ et un *riboswitch* de type II¹⁷² respectivement. La plupart des gènes ribosomaux étaient induits avec un Log_2FC maximal de 2 et très peu étaient réprimés, reflétant une activité cellulaire importante dans les conditions *in vivo* avec un renouvellement constant de nutriments.

Les capacités adaptatives métaboliques de *C. difficile* sont un élément fondamental du processus infectieux car elles permettent à la bactérie de s'adapter à son environnement^{663,666}. L'un des principales sources d'énergie de *C. difficile* provient de la fermentation des glucides et des acides aminés. De nombreux gènes dédiés au transport et au métabolisme des glucides sont différentiellement exprimés lors de l'infection, suggérant que la capacité d'utiliser une grande variété de glucides est un atout important pour la colonisation de sa niche par *C. difficile*. Une dizaine de gènes codant pour des PTS (*Phosphoenolpyruvate dependent Phosphotransferase System*) permettant l'acquisition et la phosphorylation du sorbitol (CD630_24170), du fructose (CD630_24860, CD630_31340), du mannitol (CD630_22800) et du galactitol (CD630_00430) sont induits dans des conditions infectieuses.

Inversement, une vingtaine de gènes codant pour le transport d'autres sucres comme le mannose, le lactose, mais aussi le glucose et les glucosides sont réprimés en conditions *in vivo*. Cela peut s'expliquer par l'absence de ces sucres dans le *caecum* des souris (par exemple, le glucose est absorbé totalement dans la partie haute de l'intestin et la dysbiose induite par les antibiotiques doit limiter considérablement la dégradation en monosaccharides des fibres complexes contenues dans l'alimentation des souris) et/ou l'utilisation préférentielle d'autres sources de carbone.

L'étape finale de la glycolyse conduit à la production de pyruvate qui est ensuite métabolisé par plusieurs voies de fermentations pour la production d'énergie ou pour des réactions anaboliques. Certaines de ces voies, qui aboutissent à la production de butyrate, d'éthanol ou de butanol passent par un intermédiaire majeur du métabolisme carboné et énergétique bactérien, l'acétyl-CoA. On notera que la ferrédoxine pyruvate oxydoréductase (CD630_26820) qui catalyse la décarboxylation du pyruvate en acétyl-CoA est induite en condition *in vivo*. L'acétylCoA est également le produit final de dégradation de l'éthanolamine, composé abondant *in vivo* qui peut provenir soit de la dégradation des membranes cellulaires (en conséquence de la production des toxines par *C. difficile*) soit directement de l'alimentation. De manière remarquable, tous les gènes de l'opéron *eut* (CD630_19070-19250) qui codent les protéines impliquées dans le métabolisme de l'éthanolamine sont fortement surexprimés chez les souris par rapport à la condition *in vitro* (Log₂FC compris entre 2 et 9). L'éthanolamine est une source de carbone et d'azote pour la bactérie. L'importance de son métabolisme avait déjà été observée en modèle de souris monoxénique pour la souche 630⁶⁵³, mais pas pour la souche R20291⁶⁵², et a été confirmée plus récemment en modèle hamster⁶⁶⁷.

L'acquisition du fer dans l'environnement est vitale pour la majorité des procaryotes. Il a été montré que *C. difficile* héberge dans son génome plusieurs systèmes de transport du fer sous différentes formes : fer ferreux ou ferrique, d'origine organique ou inorganique⁶⁶⁸. Seuls les gènes *feoA* (CD630_15180) et *feoB* (CD630_15170) d'un des trois opérons *feo* présents dans le génome de la souche 630 sont induits *in vivo* par rapport aux conditions *in vitro* (Log₂FC ≈ 4). Cet opéron est impliqué dans l'absorption du fer ferreux. Les autres gènes

impliqués également dans l'acquisition du fer, notamment les transporteurs ABC capables de transporter du fer ferrique (CD630_CD2997-2999) sont réprimés *in vivo*. Cela suggère qu'au cours de l'infection, le fer ferreux est la principale source de fer pour la bactérie.

L'analyse des gènes différentiellement exprimés chez *C. difficile* a également révélé un grand nombre de gènes de sporulation induits *in vivo*. Une centaine de gènes impliqués dans la sporulation chez *C. difficile* sont ainsi induits dont plusieurs gènes dépendant du facteur σ^k impliqués dans la synthèse des couches externes de la spore (cortex, manteau et exospore). Ces résultats suggèrent que la sporulation est rapidement induite au cours de l'infection, permettant ainsi à *C. difficile* de persister dans les intestins de l'hôte et de disséminer dans l'environnement, malgré la réponse immune de l'hôte. Enfin, parmi les gènes différentiellement exprimés, 58 correspondent à des gènes d'ARNnc (dont 19 réprimés et 39 induits) et une dizaine de candidats (dont des ARNnc Hfq dépendant et des *riboswitch*) ont été sélectionnés pour des analyses ultérieures (**Tableau 2**).

Tableau 2 - Candidats ARNnc différentiellement exprimés en condition infectieuse sélectionnés pour des analyses ultérieures. Les ARNnc annotés avec un astérisque (*) correspondent aux ARN identifiés comme partenaires de Hfq lors d'une expérimentation RIP-seq⁶⁰². (IGR = *Intergenic Region*)

Nom de l'ARNnc	Log ₂ FC	p-value ajustée	In vivo	5' start	3' end	Brin	Annotation
CD630_SQ1296	8,359	3,11E-10	UP	2187027	2186776	-	IGR entre CD1880 et CD1882
CD630_SQ995	7,182	3,40E-07	UP	1749408	1749672	+	IGR entre CD1511 et <i>cotB</i>
CD630_SQ1076	6,442	4,43E-18	UP	1851335	1851585	+	ARNas putatif en 3' UTR de <i>map2</i>
CD630_n00930*	4,56	8,72E-06	UP	3196622	3196979	-	IGR entre CD2752 et CD2753
CD630_n00640	3,772	1,02E-09	UP	2180693	2181055	+	IGR entre CD18780 et CD18782
CD630_n00680*	2,921	9,75E-07	UP	2285913	2286311	+	IGR entre CD19801 CD19810
CD630_n00620*	2,057	2,55E-05	UP	1943181	1943450	+	IGR entre CD1667.0 et CD1667.1
CD630_n00830	-6,673	4,27E-19	DOWN	2720238	2720338	+	ARNas CD2351 (<i>grdB</i>)

Parmi les ARNnc sélectionnés, trois ont déjà été identifiés précédemment lors d'une expérience RIP-seq comme étant associé à la protéine Hfq⁶⁶⁹. L'un d'entre eux (CD630_n00620/Cdi1_6/RCd5) est un *riboswitch* de type I se liant au c-di-GMP induit lors de la phase stationnaire de croissance localisé à côté de la séquence d'un activateur transcriptionnel de la famille HxIR. Les deux autres sont localisés dans des régions intergéniques des gènes codant des protéines hypothétiques conservées (CD630_n00930 et CD630_n00620). Deux ARNas ont également été sélectionnés pour des études approfondies. L'un d'entre eux régulé positivement durant l'infection est un ARNas putatif du gène *map2* codant pour une méthionine aminopeptidase (CD630_SQ1076). Le second, réprimé lors de

l'infection, est un ARNas du gène *grdB* codant pour une sous-unité de la glycine réductase. Tous les autres sont situés dans des régions intergénique des gènes codant pour des protéines de fonctions connues ou hypothétiques. L'un d'entre eux par exemple (CD630_SQ995) est localisé à proximité du gène d'une protéine de la membrane externe de la spore, CotB. Un autre (CD630_SQ1296) est localisé à proximité de la séquence codant pour un fragment d'un transporteur ABC. Un candidat (CD630_n00640) est également retrouvé entre un gène d'une protéine putative d'un transposon conjugatif (Tn1549) et de d'un facteur sigma putatif de ce transposon.

Du côté de la souris, le nombre plus important de *reads* s'alignant sur le génome a permis de comparer plusieurs conditions entre elles. De la même manière que pour *C. difficile*, les temps d'infection T28h et T32h ont été regroupés en un seul groupe (MI) dans la mesure où aucune différence statistiquement significative d'expression de gènes n'a été observée entre ces deux conditions. Toutes les PCA issues des analyses différentielles chez la souris sont fournies en annexes (**Figure S6 à S8**). Dans un premier temps, les conditions contrôles (MC pour *Mouse Control*) ont été comparées aux conditions d'infection MI (T28h+T32h) (**Figure 37A**). Cette analyse a mis en évidence 10 011 gènes significativement différentiellement exprimés entre ces deux conditions, 5409 gènes réprimés et 4602 gènes induits dans les conditions d'infection par rapport aux conditions contrôles. Parmi ces gènes différentiellement exprimés, 2579 correspondent à des gènes d'ARNs régulateurs (2449 réprimés et 130 induits) comprenant une très grande majorité de longs ARN non-codants *lncRNA* (plus de 2400). Dans un second temps, le groupe de souris T8h (appelé *early* pour infection précoce) a été comparé avec les groupes combinés de souris infectées T28h et T32h (appelé *late* pour infection tardive) (**Figure 37B**). 3182 gènes significativement différentiellement exprimés ont été détectés lors de la comparaison de ces deux groupes, dont 2087 induits et 1095 réprimés dans les conditions d'infection tardive par rapport aux conditions d'infection précoce. Parmi ces gènes, 593 ARNnc dont 107 induits et 486 réprimés ont été détectés avec comme précédemment une très vaste majorité de *lncRNA*. Enfin, parmi les souris infectées, une souris du groupe T28h et deux du groupe T32h présentaient des symptômes (diarrhée) et un *caecum* très inflammé. Ces souris ont été regroupées (groupe

sick) et comparées aux souris des même points post-infection (T28h et T32h) mais qui ne présentaient aucun symptôme (groupe *healthy*). Cette analyse a révélé 1861 gènes significativement différentiellement exprimés, 1058 induits et 803 réprimés (**Figure 37C**). Parmi ces gènes différentiellement exprimés, 203 sont des gènes d'ARNs régulateurs (144 induits et 59 réprimés). Comme pour l'analyse différentielle précédente, la quasi-totalité de ces ARN sont des lncRNA. Les gènes différentiellement exprimés pour les différentes conditions décrites précédemment ont été classifiés en groupe fonctionnel grâce au programme DAVID ⁶⁷⁰. A partir des clusters fonctionnels des gènes différentiellement exprimés, on peut constater un enrichissement de plusieurs groupes fonctionnels liés à la réponse immunitaire, notamment inflammatoire, et tout particulièrement pour les gènes différentiellement exprimés chez les souris malades (**Figure 37C**). Parmi les ARNnc différentiellement exprimés chez la souris lors de l'ICD, les miRNA minoritaires par rapport aux lncRNA ont été extraits pour chaque modèle comparatif et sont présentés dans le **Tableau 3** ci-dessous. Parmi les 26 miRNA identifiés, 22 sont induits et 4 sont réprimés lors de l'infection à *C. difficile*. Deux miRNA (Mir1938 et Mir3109) ont été retrouvés induits dans les trois analyses différentielles avec un Log₂FC compris entre 2 et 5 mais l'étude de leur fonction n'a jamais été réalisée. Bien que beaucoup de ces microARN ne soient pas encore caractérisés dans la littérature, certains ont été décrits et associés à la réponse inflammatoire. Parmi les miRNA induits chez la souris lors de l'ICD, le Mir29a a été décrit dans mécanismes d'induction de l'apoptose des cellules épithéliales intestinales dans les colites ulcéreuses et a également été observé lors de l'infection par *Salmonella*. Le Mir1192 est décrit comme ayant une action anti-inflammatoire *via* l'inhibition du récepteur à chimiokines CXCR4. Le Mir24-2 est impliqué dans la réparation des dommages intestinaux liés aux maladies inflammatoires intestinales. Le Mir686 joue un rôle pro-inflammatoire en régulant positivement l'expression du ligand CCL7 de la chimiokine 7. Les Mir682 et Mir6538 ont un effet protecteur sur les cellules intestinales endommagées lors d'épisodes ischémiques. Le Mir99a, qui lui est réprimé chez les souris malades infectées avec *C. difficile*, exerce un effet anti-inflammatoire lorsqu'il est exprimé dans les tissus adipeux, en inhibant le TNF- α . L'analyse détaillée de l'ensemble des résultats obtenus dans cette expérience de dual RNA-seq est encore en cours. Ce travail fera l'objet d'une publication en cours de rédaction.

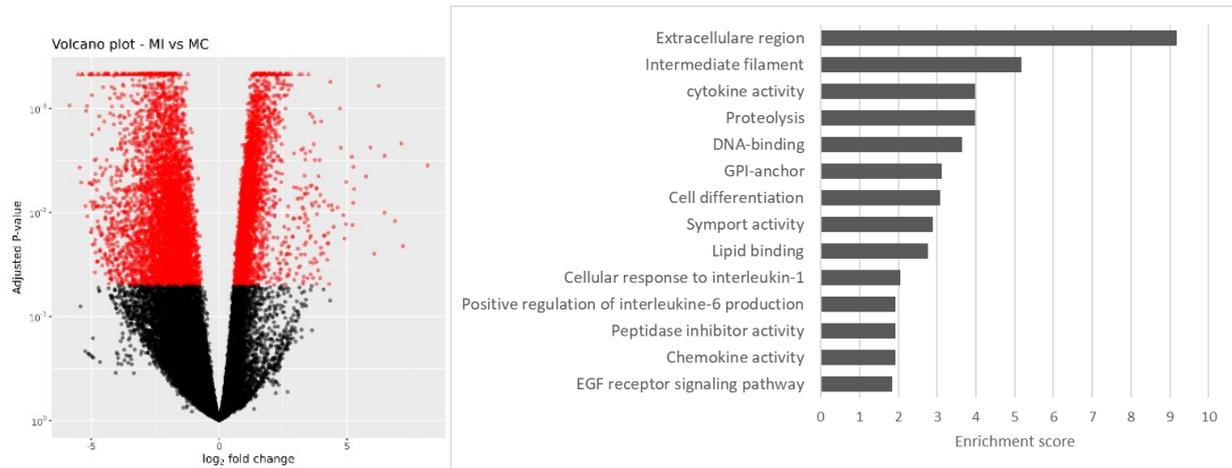
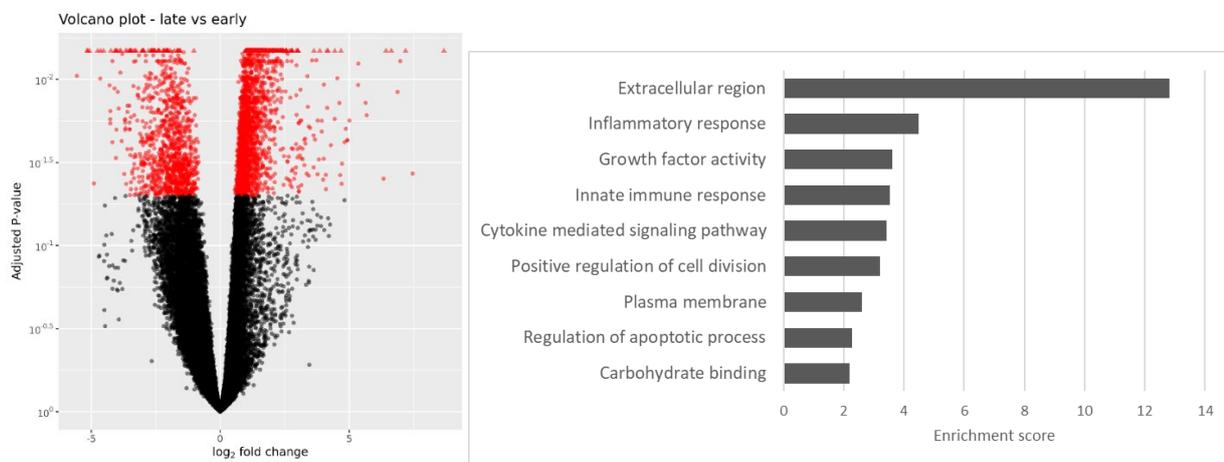
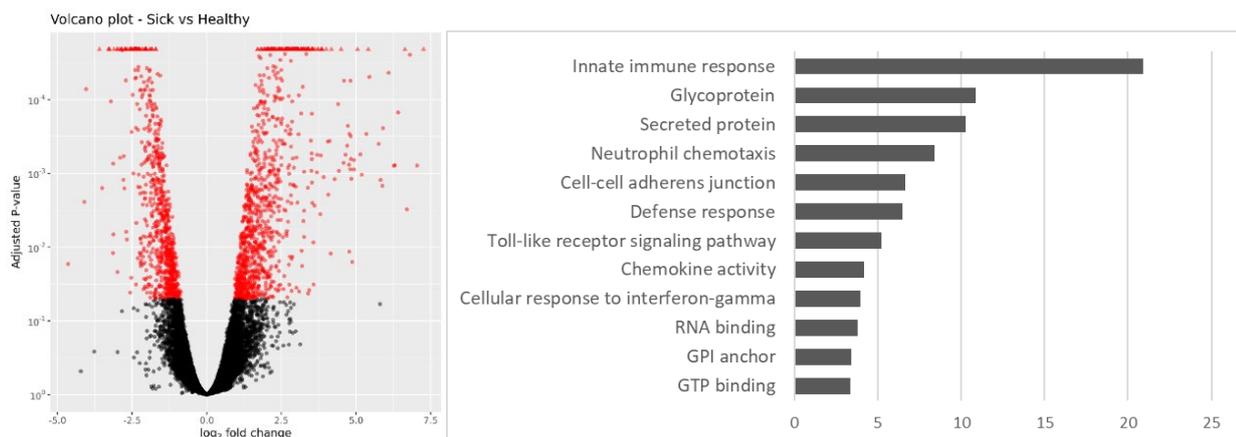
A**B****C**

Figure 37 – Analyse de l’expression différentielle des gènes de la souris au cours d’une ICD. Pour chaque figure, de gauche à droite sont représenté les *Volcano plot* et les groupes de gènes fonctionnels identifiés avec le programme DAVID pour les comparaisons entre (A) les souris non infectées (MC) et souris infectées à T28h et T32h (MI) ; (B) les souris infectées à T8h (early) et à T28h et 32h (late) ; (C) les souris malades (*sick*) et saines (*healthy*) des temps T28h et T32h. La sélection des clusters a été faite manuellement en se référant aux GOterm associés et en sélectionnant seulement les clusters avec un score d’enrichissement supérieur à 1,5. Les paramètres par défaut de DAVID ont été utilisés.

Tableau 3 – Liste des miRNA différentiellement exprimés chez la souris lors de l'ICD. (A) miRNA issus de l'analyse différentielle réalisée entre les souris infectées des groupes T28h/T32h (MI) et les souris non infectées (MC). (B) miRNA issus de l'analyse différentielle réalisée entre les souris infectées à T8h (*early*) et les souris des groupes T28h/T32h (*late*). (C) miRNA issus de l'analyse différentielle réalisée entre les souris malades (*sick*) et les souris saines (*healthy*) des groupes T28h/T32h. Les miRNA communs entre les différentes analyses sont indiqués par un astérisque (*).

Nom du gène	Log2FC	<i>p</i> adj	Annotation	Publications (PMID)
(A) miRNA différentiellement exprimés chez les souris infectées (MI vs MC)				
gene-Mir1938*	5,189	0,00292395	microRNA 1938	Inconnu
gene-Mir3109*	4,106	0,00782967	microRNA 3109	Inconnu
gene-Mir29a	3,975	0,02182115	microRNA 29a	29156760 ; 25674218 ; 23826261
gene-Mir1192	3,578	0,03937017	microRNA 1192	30355939
gene-Mir24-2	3,566	0,03025878	microRNA 24-2	33301881
gene-Mir5123	3,288	0,0459979	microRNA 5123	Inconnu
gene-Mir21a	3,25	0,00896551	microRNA 21a	Inconnu
gene-Mir7677	1,94	0,03568633	microRNA 7677	Inconnu
gene-Mir3113	1,814	0,00389382	microRNA 3113	Inconnu
gene-Mir1893	1,743	0,03852209	microRNA 1893	Inconnu
gene-Mir8094	1,447	0,01879777	microRNA 8094	Inconnu
gene-Mir546	1,149	0,01278035	microRNA 546	Inconnu
gene-Mir692-3	1,129	0,02451075	microRNA 692-3	Inconnu
gene-Mir686	1,056	0,04864586	microRNA 686	32629323
gene-Mir682	1,038	0,03587233	microRNA 682	27124584
gene-Mir1949	-1,682	0,000696305	microRNA 1949	29543730
gene-Mir1843b	-1,262	0,018466735	microRNA 1843b	Inconnu
(B) miRNA différentiellement exprimés chez les souris infectées à T28h et T32h (<i>late</i> vs <i>early</i>)				
gene-Mir1938*	3,036	0,018256316	microRNA 1938	Inconnu
gene-Mir3109*	2,604	0,031136669	microRNA 3109	Inconnu
gene-Mir21a	2,341	0,02655575	microRNA 21a	Inconnu
gene-Mir7677	2,022	0,038392942	microRNA 7677	Inconnu
gene-Mir1949	-1,49	0,007764685	microRNA 1949	29543730
gene-Mir1843b	-1,515	0,009902049	microRNA 1843b	Inconnu
(C) miRNA différentiellement exprimés chez les souris malades (<i>sick</i> vs <i>healthy</i>)				
gene-Mir6538	3,242	0,004433364	microRNA 6538	31789347
gene-Mir1938*	2,527	0,000055343	microRNA 1938	Inconnu
gene-Mir2137	2,301	0,011226193	microRNA 2137	34465878
gene-Mir3109*	1,878	0,019220786	microRNA 3109	Inconnu
gene-Mir6236	1,861	0,001679473	microRNA 6236	32293320 ; 31004989
gene-Mir7672	1,752	0,026098429	microRNA 7672	Inconnu
gene-Mir99ahg	-1,646	0,000359429	Mir99a and Mirlet7c-1	29849090 ; 32793975

I. Construction du mutant de délétion Δhfq

Il existe plusieurs outils génétiques pour la manipulation du génome de *C. difficile*. L'un des plus utilisés est la technologie Clostron basée sur l'utilisation d'introns mobiles de type II modifiés et de marqueur rétro-transposables activés (RAM)^{671,672}. Cependant, bien que cette méthode permette d'éditer presque toutes les régions chromosomiques et que les RAM permettent d'identifier facilement les mutants, elle présente certains inconvénients dont l'insertion d'un gène de résistance à l'erythromycine pouvant engendrer des effets polaires. Une autre limite à cette méthode est la difficulté à trouver un site d'insertion pour les gènes de petite taille. La méthode ACE (*Allele-Coupled Exchange*) est une autre technique d'édition du génome de *C. difficile* très fréquemment utilisée. Cette méthode basée sur l'utilisation d'un vecteur plasmidique semi-suicide portant un marqueur de contre sélection (le gène *codA* ou le gène *pyrE*)^{673,674} implique un événement de double recombinaison entre le plasmide d'édition et le génome et la sélection des clones mutants ayant perdu le plasmide sur un milieu défini sélectif (supplémenté avec la 5-fluorocytosine pour les plasmides portant *codA* ou l'acide fluoro-orotique pour les plasmides portant *pyrE*). Bien que cette méthode soit efficace pour effectuer des mutations ponctuelles, des délétions et des insertions, elle peut être difficile à appliquer dans certaines situations comme l'inactivation de gènes métaboliques pouvant affecter la croissance sur milieu défini. De plus, la difficulté à sélectionner les mutants ayant perdu le plasmide d'édition peut conduire à l'apparition de révertants.

L'approche mise au point (*cf. Publication 2* jointe à la fin du manuscrit⁶⁰⁹) est basée sur l'utilisation d'un vecteur plasmidique contenant une mini-cassette CRISPR artificielle avec des *spacers* ciblant un gène chromosomique⁶⁰⁹. L'ARNcr exprimé sur un vecteur portant la minicassette CRISPR sous le contrôle d'un promoteur natif ou inductible est pris en charge par la machinerie endogène Cas (crRNP) afin de former un complexe qui permettra de reconnaître le *protospacer* associé avec un motif PAM localisé dans la région chromosomique ciblée, entraînant son clivage puis la mort des cellules portant l'allèle

sauvage (**Publication 2 ; FIG 1A**). Le plasmide de mutagenèse possède des séquences homologues flanquant le *protospacer*, permettant la recombinaison homologue et l'échange des allèles mutant et sauvage entre le plasmide et la région chromosomique ciblée, avant que le système CRISPR-Cas ne puisse agir (**Publication 2 ; FIG 1B**). Le complexe crRNP cible ensuite le motif PAM adjacent au *protospacer*, ce qui permet l'élimination du plasmide et la conservation de la version mutée sur le chromosome. Cette technique permet donc de contre-sélectionner les cellules portant l'allèle sauvage et de préserver les cellules mutantes. De plus, cette méthode basée sur un système CRISPR-Cas endogène est plus facile à mettre en place chez les procaryotes que des méthodes utilisant les technologies CRISPR-Cas9 et CRISPR-Cpf1 (Cas12a) exprimant des protéines Cas potentiellement toxiques pour la bactérie. Cette méthode a été utilisée pour la construction d'un mutant de délétion du gène *hfq* codant une protéine chaperon à l'ARN Hfq, qui a par la suite été étudiée pour plusieurs phénotypes *in vitro* et dans un modèle animal pour sa virulence. Une souche complémentée du mutant *via* l'introduction du gène sur un plasmide sous contrôle d'un promoteur inducible par l'anhydrotétracycline, a également été construite et caractérisée. Des analyses par qPCR ont confirmé l'absence d'expression de *hfq* chez les mutants Δhfq des souches CD630E et CDR20291. D'autre part, le niveau de transcrits du gène *hfq* chez le mutant complémenté était 400 à 500 fois plus élevé que celui de la souche sauvage lors de l'analyse qPCR. Cela est dû à la forte induction du gène *hfq* sous le contrôle pTet en présence de 250 ng/mL d'ATc. L'absence de protéine et une forte augmentation de la quantité de protéine ont également été observées par immunoblot, respectivement chez le mutant de délétion et le mutant complémenté (**Publication 2 ; FIG 4**).

II. Étude phénotypique du mutant de délétion *hfq*

De nombreuses études phénotypiques *in vitro* réalisées sur des mutants Δhfq de différentes espèces bactériennes pathogènes ont mis en évidence plusieurs phénotypes impactés par la délétion du gène *hfq* comme une augmentation de la sensibilité à divers stress (stress oxydatif, carence en fer, choc thermique, stress acide, etc.), une motilité réduite et un défaut de croissance (en milieu liquide ou solide) qui est observé chez la quasi-totalité des mutants de délétion⁴⁶⁷ chez les bactéries Gram négatifs, alors que la délétion du gène

hfq chez des bactéries Gram positifs comme *L. monocytogenes*⁶⁷⁵ ou *S. aureus*⁴⁶⁹ n'a pas d'impact sur leur dynamique de croissance. Chez *C. difficile*, une étude antérieure réalisée sur un mutant déplété pour la production de la protéine Hfq par ARNas a mis en évidence plusieurs phénotypes liés à divers processus clés des ICD comme la réponse au stress, la formation de biofilms et la sporulation⁶¹⁰. Afin de pouvoir confirmer les phénotypes observés chez le mutant de déplétion, plusieurs tests phénotypiques ont été réalisés chez le mutant de délétion dans la souche CD630E construite par la méthode CRISPR-Cas endogène⁶⁰⁹.

1. Étude de la sporulation

La capacité à former des spores est un aspect fondamental de la persistance du pathogène dans le tractus intestinal. L'étude précédemment menée par Boudry et *al.* sur le mutant de déplétion dérivé de la souche CD630E pour la protéine Hfq avait mis en évidence un impact de la déplétion de Hfq sur le taux de sporulation⁶¹⁰, se traduisant par une augmentation du taux de sporulation chez *C. difficile* par rapport à la souche sauvage. Les taux de sporulation des souches CD630E WT, CD630E Δ *hfq* et de la souche CD630E Δ *hfq* complémentée par un vecteur d'expression (dérivé du pRPF185) portant le gène *hfq* sous le contrôle d'un promoteur inductible (pTet) ont été comparés (les souches sauvages et mutantes non complémentées porte le vecteur d'expression vide). Ces trois souches ont été cultivées 24h et 48h à 37°C en conditions anaérobies dans un milieu liquide BHIS (*Brain Heart Infusion Supplemented*) supplémenté avec 100 ng/mL d'anhydrotétracycline (ATc) permettant l'induction du gène *hfq* pour la souche complémentée et 7,5 µg/mL de thiamphénicol (Tm) permettant le maintien du plasmide d'expression. Pour chacune des souches, 1 mL de culture a été prélevé à 24h et 48h puis séparé en deux échantillons. Afin de déterminer le nombre total de bactéries, plusieurs dilutions du premier échantillon ont été déposées en gouttes (10 µL par goutte) sur un milieu BHI solide contenant 0,1% de taurocholate (acide biliaire permettant l'induction de la germination des spores de *C. difficile*), 100 ng/mL d'ATc et 15 µg/mL de Tm. Le deuxième échantillon a été incubé pendant 30 min à 65°C afin d'éliminer les cellules végétatives et ne conserver que les spores. Le nombre de spore a ensuite été estimé de la même manière en déposant des gouttes (10 µL)

de plusieurs dilutions de l'échantillon chauffé sur un milieu BHI solide contenant 0,1% de taurocholate. Après 24h et 48h en milieu BHIS (supplémenté avec l'ATc), la souche mutante CD630E Δhfq présentait un taux de sporulation plus élevé que la souche sauvage CD630E. De plus, la complémentation plasmidique du mutant (CD630E Δhfq + p-Hfq) a permis de réverter le taux de sporulation du mutant à un niveau comparable à celui de la souche sauvage (**Figure 38**). Ces résultats sont en adéquation avec ce qui avait été observé précédemment avec le mutant de déplétion et confirment ainsi un rôle de la protéine Hfq dans le contrôle de la sporulation chez *C. difficile*⁶¹⁰.

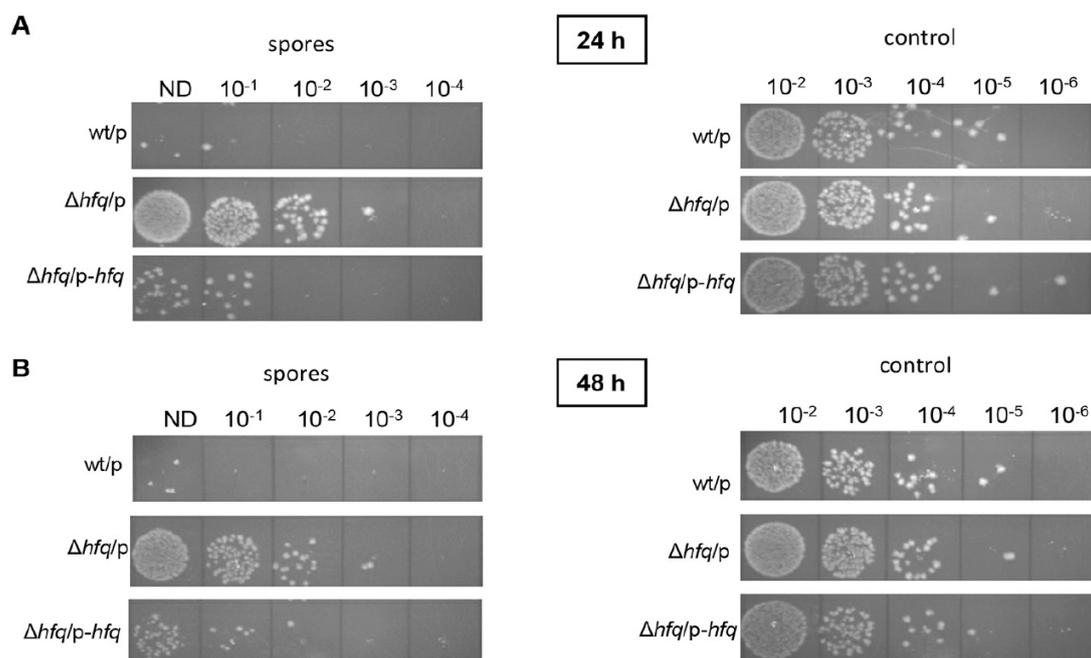


Figure 38 – Comparaison des niveaux de sporulation entre les souches CD630E WT/p, $\Delta hfq/p$ et $\Delta hfq/p-hfq$ de *C. difficile*. Le niveau de sporulation (nombre de spores), comparé à la quantité totale de bactéries comme contrôle, a été estimé après 24h (A) et 48h (B) de croissance en milieu liquide BHIS supplémenté avec de 100 ng/mL d'ATc et 7,5 $\mu\text{g/mL}$ de Tm. Les dilutions utilisées sont indiquées en haut de chaque image (ND = non diluée). Les échantillons pour compter les spores ont été chauffés afin d'éliminer les cellules végétatives, tandis que les échantillons contrôles n'ont pas été chauffés de manière à estimer le nombre total de bactéries. (p = plasmide d'expression vide ; p-*hfq* = plasmide d'expression portant le gène *hfq* sous le contrôle d'un promoteur pTet)⁶⁰⁹

2. Étude de la formation de biofilms

C. difficile est une bactérie capable de former des biofilms mono-espèce¹⁸⁷ ou en association avec un biofilm polymicrobien *in vitro* et plusieurs arguments existent en faveur de la production d'un biofilm *in vivo* qui pourrait contribuer ainsi à sa persistance dans l'intestin de l'hôte et favoriser les rechutes^{178,188,676}. Une étude récente sur les mécanismes de régulation de la formation de biofilms chez *C. difficile* par les ARNnc a mis en évidence

une vingtaine d'ARNnc spécifiquement exprimés durant la formation de biofilms ⁶⁷⁷. La déplétion de la protéine Hfq chez *C. difficile* entraînait une augmentation de la formation de biofilm d'un facteur 2 environ par rapport à la souche sauvage et ce phénotype était réverté par la complémentation ⁶¹⁰. Afin de pouvoir confirmer ce phénotype avec la souche mutante obtenue par la méthode CRISPR-Cas endogène, un test de formation de biofilm a été réalisé dans des conditions similaires à celles de la précédente étude. Les souches sauvages, mutantes et mutantes complémentées ont été cultivées en plaques de polystyrène 24 puits dans 1 mL un milieu BHIS supplémenté avec du glucose (0,1 M) et 50 ng/mL d'Atc (ensemencement au 1/100^{ème} à partir d'une culture de nuit). Après 72h de culture en anaérobiose à 37°C, le surnageant est retiré délicatement, le biofilm adhérent à la plaque est lavé 2 fois par 1 mL de PBS puis séché 10 minutes à 37°C. La biomasse des biofilms produits est alors mesurée par coloration au cristal violet : incubation pendant 30 minutes (37°C) de chaque biofilm avec 1 mL de cristal violet à 0,2%, puis les puits sont à nouveau lavés 2 fois par 1 mL de PBS. Un mL de mélange éthanol/acétone (80/20) est ensuite introduit dans chacun des puits et le biofilm est décroché par grattage. L'absorbance des suspensions ainsi obtenues est alors mesurée à 570 nm et représente la biomasse des biofilms produits par les différentes souches. De manière surprenante aucune augmentation significative de formation de biofilms n'a été observée chez le mutant Δhfq (**Figure 39**).

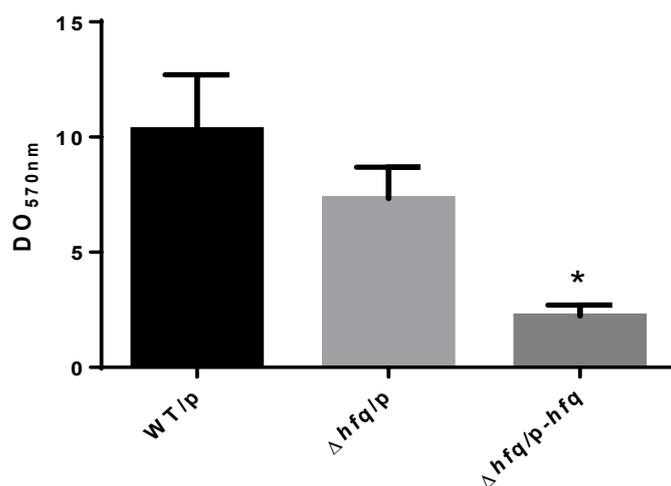


Figure 39 – Formation de biofilm par les souches CD630E WT/p, $\Delta hfq/p$ et $\Delta hfq/p-hfq$ de *C. difficile*. La formation de biofilm a été testé en microplaques dans un milieu BHIS supplémenté avec 0,1 M de glucose, 7,5 µg/mL de Tm et 50 ng/mL d'ATc. Sont représentées les valeurs moyennes ($n = 3$) de la quantification de la formation de biofilm au cristal violet (DO_{570}) et l'erreur standard (* $P \leq 0,05$; ANOVA univariée suivi par un test de Dunnett). (p = plasmide d'expression vide; p-*hfq* = plasmide d'expression portant le gène *hfq* sous le contrôle d'un promoteur pTet)

De plus, la souche complémentée présentait également une diminution significative de la production de biofilm, qui s'explique vraisemblablement par le fort retard de croissance de cette souche, probablement dû à l'effet toxique de la surexpression du gène *hfq* via le pTet (*cf.* les expériences de qPCR et d'immunoblot réalisées après la construction du mutant).

3. Étude de la réponse au stress

Au cours de l'infection, *C. difficile* doit faire face à de multiples stress, incluant la présence de nombreux composés antimicrobiens, des changements brutaux de pH, la présence d'espèces réactives de l'oxygène (ROS) produites durant l'inflammation, une compétition pour l'acquisition du fer avec les autres bactéries¹⁶⁶. Il a été décrit chez d'autres espèces bactériennes un rôle de la protéine Hfq dans l'adaptation à certains de ces facteurs de stress⁶⁷⁸. De plus, lors de la précédente étude sur le mutant de déplétion, la comparaison des transcriptomes de la souche déplétée pour Hfq et de la souche sauvage avait révélé, en outre, une expression différentielle de groupes de gènes pouvant être impliqués dans la réponse au stress chez le mutant déplété (dont *groEL* et *grpE* qui sont induits lors de la réponse à un choc thermique et un choc acide respectivement)⁶¹⁰. Afin d'étudier le rôle de la protéine Hfq dans la réponse au stress chez *C. difficile*, les 3 souches (sauvage, mutante et mutante complémentée) ont été soumises à divers facteurs de stress.

a. Stress oxydatif et carence en fer

Dans un premier temps, des tests de résistance au diamide et dipyridyl ont été réalisés. Le diamide est un agent oxydant des groupes thiols (-SH) entraînant un stress disulfide par lequel une liaison covalente se forme entre deux groupements -SH. Le dipyridyl est un chélateur du fer et d'autres ions métalliques divalents. Les souches ont été testées en milieu solide TY (*Tryptone Yeast Extract*) supplémenté avec de l'ATc (100 ng/mL) et sur lequel a été déposé un disque imbibé soit avec 10 µL de diamide 1M soit avec 10 µL de dipyridyl 0,3M. Le diamètre de la zone d'inhibition a été mesuré après 48h d'incubation à 37°C en conditions anaérobies. Aucune différence entre la souche sauvage et la souche mutante n'a été observée pour les deux composés (**Figure 40**).

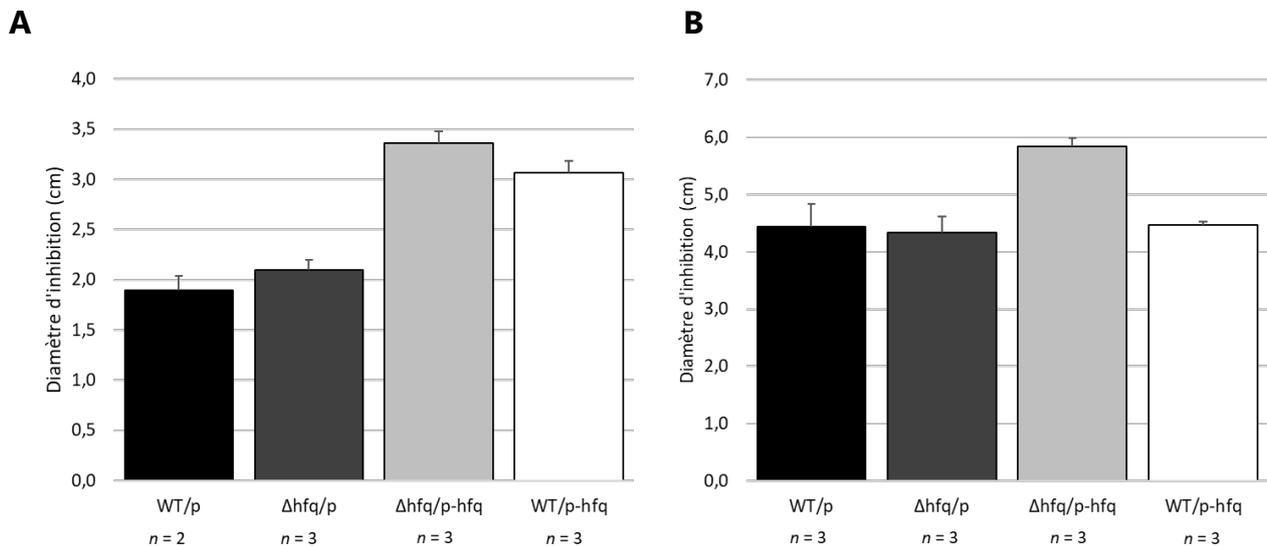


Figure 40 – Sensibilité au stress des souches CD630E WT/p, $\Delta hfq/p$, $\Delta hfq/p-hfq$ et WT/p- hfq de *C. difficile*. Les histogrammes représentent les moyennes des diamètres d'inhibition de croissance des souches CD630E WT/p, $\Delta hfq/p$, $\Delta hfq/p-hfq$ et WT/p- hfq exposées à un disque imbibé avec 10 μ L de diamide 1M (A) ou 10 μ L de dipyridyl 0,3M (B). (p = plasmide d'expression vide ; p- hfq = plasmide d'expression portant le gène hfq sous le contrôle d'un promoteur pTet)

Ces résultats ne correspondent pas à ce qui avait été observé précédemment avec le mutant de déplétion qui présentait une sensibilité accrue (d'environ 30%) aux deux composés par rapport à la souche sauvage⁶¹⁰. Au contraire, une sensibilité légèrement accrue à ces deux composés a été observée pour la souche complétementée (**Figure 40**), mais cet effet est probablement dû à un effet toxique lié à la production en grande quantité de la protéine Hfq dans cette souche.

b. Exposition aux acides biliaires

Les cellules végétatives de *C. difficile* sont également exposées durant l'infection aux acides biliaires primaires et secondaires de l'hôte. Les acides biliaires primaires produits par le foie sont essentiellement composés de l'acide cholique et l'acide chénodésoxycholique conjugués soit avec la taurine soit avec la glycine. Les acides biliaires secondaires sont principalement constitués par l'acide désoxycholique et l'acide lithocholique et sont formés par les bactéries du microbiote intestinal à partir des acides biliaires primaires⁶⁷⁹. Les trois souches étudiées ont été exposés à un mélange d'acides biliaires ainsi qu'à l'acide biliaire primaire cholate et son acide biliaire secondaire dérivé le désoxycholate. Les souches ont été cultivées pendant 4h (jusqu'à obtenir une $DO_{600} \approx 0,4 - 0,5$) dans un milieu riche TY supplémenté en ATc (10 ng/mL) pour l'induction de l'expression de Hfq pour la souche complétementée, puis des dilutions de deux cultures indépendantes ont été déposées en

gouttes sur un milieu solide TY supplémenté avec soit 0,1% d'une mixture d'acides biliaries, soit 0,2% de cholate ou soit 0,025% de déoxycholate. Les boites ont été incubées à 37°C pendant 48h pour le mélange d'acides biliaries et le déoxycholate et 24h pour le cholate. Aucune différence de croissance n'a été observée entre la souche mutante et la souche sauvage que cela soit pour le mélange d'acides biliaries, le cholate ou le déoxycholate (**Figure 41B-D**). La souche complémentée n'a pas non plus présenté de différence avec la souche sauvage.

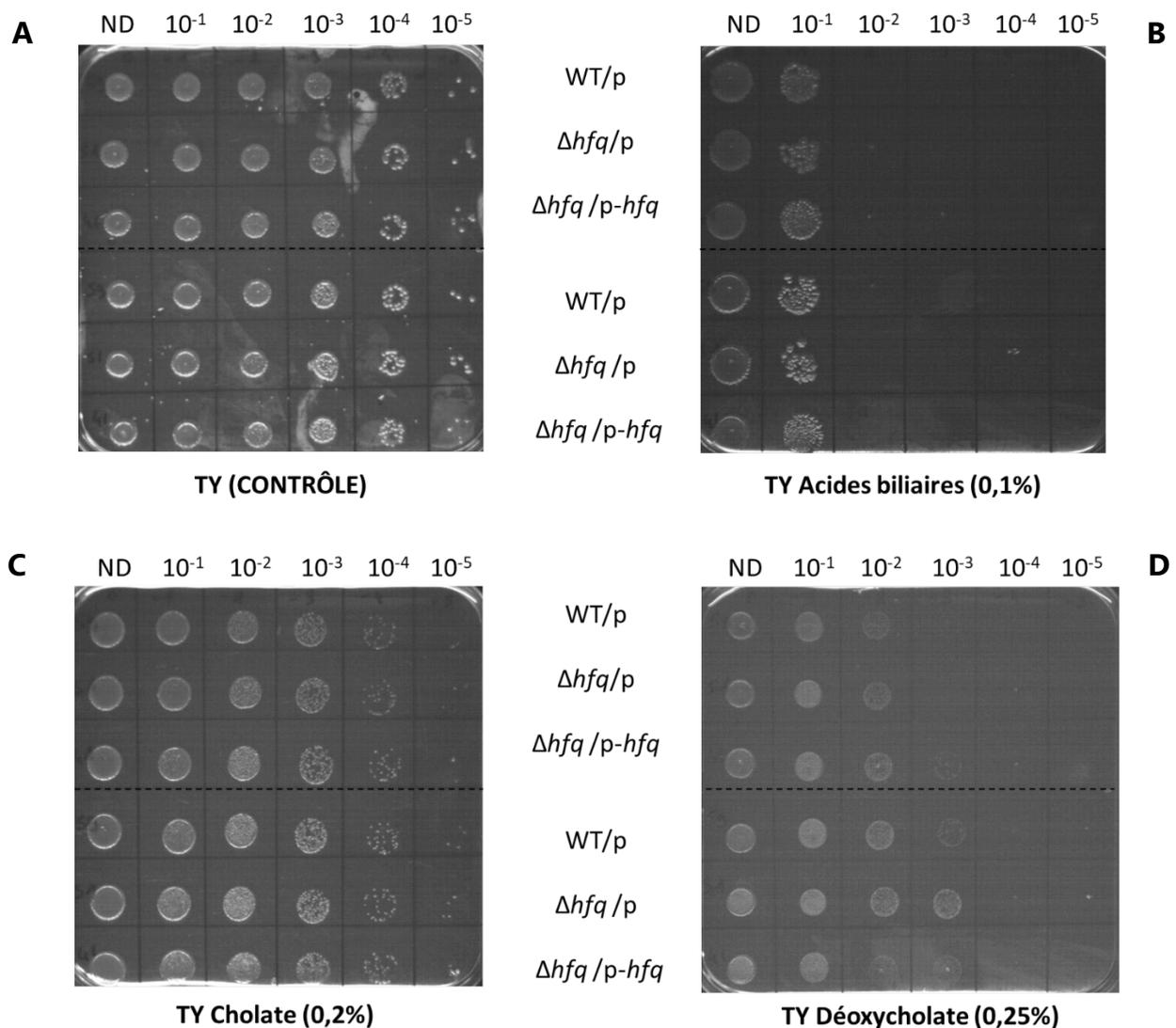


Figure 41 – Tests de sensibilité aux acides biliaries des souches CD630E WT/p, $\Delta hfq/p$ et $\Delta hfq/p-hfq$. La condition contrôle en milieu TY est illustrée en (A). 5 μ L de plusieurs dilutions (indiquées au-dessus des figures ; ND = non dilué) de culture de 4h en TY des souches CD630E WT/p, $\Delta hfq/p$ et $\Delta hfq/p-hfq$ ont été déposés sur des milieux TY supplémentés avec 10 ng/mL d'ATc et (B) un mélange d'acides biliaries 0,1% ; (C) du cholate 0,2% ; (D) du déoxycholate 0,25%. Les duplicats biologiques sont séparés par une ligne pointillée. (p = plasmide d'expression vide ; p-*hfq* = plasmide d'expression portant le gène *hfq* sous le contrôle d'un promoteur pTet)

c. Étude des stress osmotiques et acides

Deux autres facteurs de stress importants auxquels *C. difficile* est confronté dans la lumière intestinale sont les fortes conditions d'osmolarité (équivalentes à 300-400 mM de chlorure de sodium [NaCl])⁶⁸⁰ et les variations importantes de pH (variation entre 5 et 7 dans le colon humain)⁶⁸¹. La capacité du mutant Δhfq à résister à de telles conditions a donc été testée et comparée à celle de la souche sauvage (**Figure 42**).

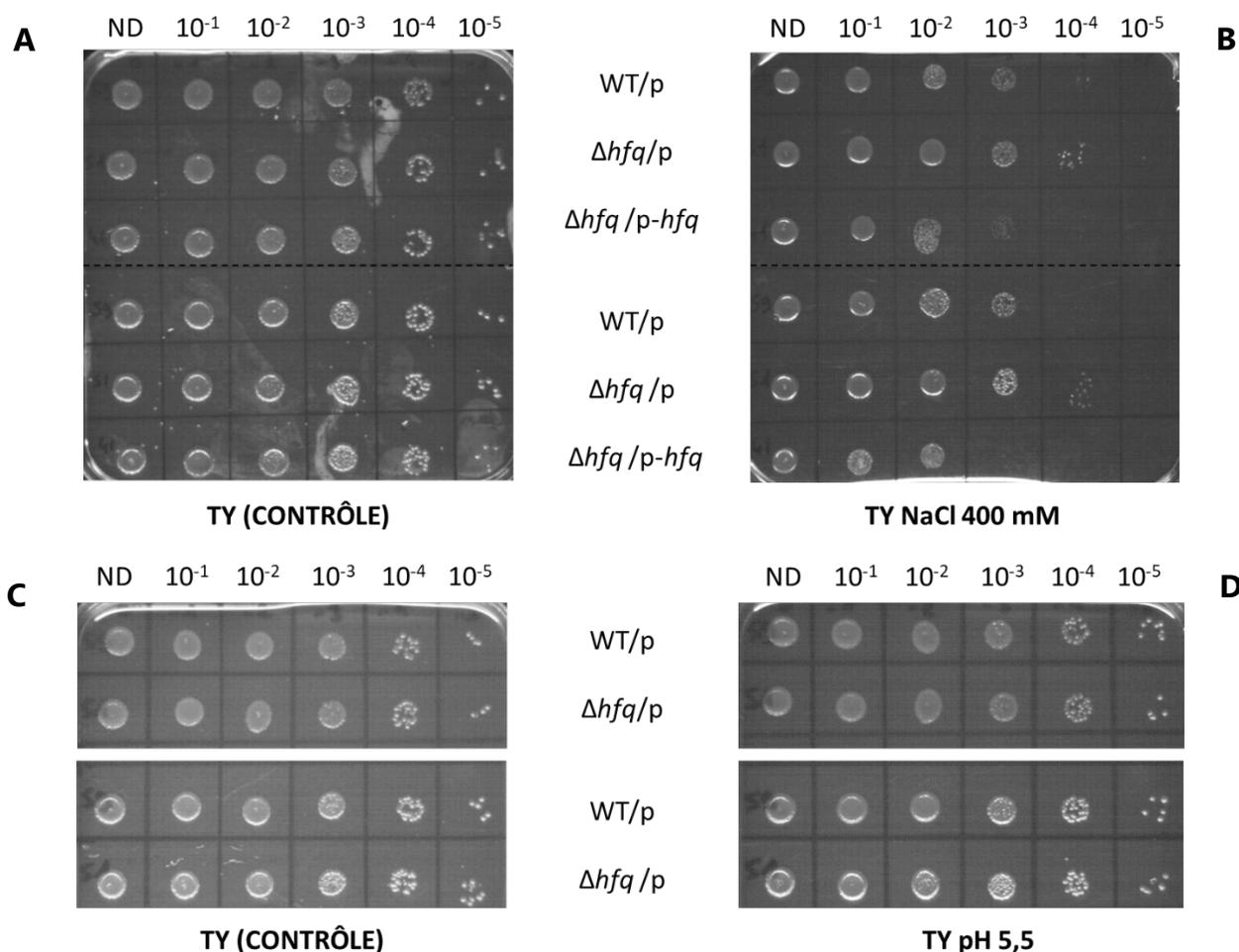


Figure 42 – Tests de sensibilité aux stress osmotiques et acides des souches CD630E WT/p, $\Delta hfq/p$ et $\Delta hfq/p-hfq$. Les conditions contrôles des figures (B) et (D) sont illustrées en (A) et (C) respectivement. 5 μ L de plusieurs dilutions (indiquées au-dessus des figures ; ND = non dilué) de cultures de 4h en TY des souches CD630E WT/p, $\Delta hfq/p$ et $\Delta hfq/p-hfq$ ont été déposés sur des milieux TY supplémentés avec 10 ng/mL d'ATc et (B) du NaCl 400 mM ; (D) un pH de 5,5. La souche mutante complémentée n'est pas représentée pour le test pH. (p = plasmide d'expression vide ; p-*hfq* = plasmide d'expression portant le gène *hfq* sous le contrôle d'un promoteur pTet)

Les souches testées (sauvage, mutante et mutante complémentée) ont été cultivées pendant 4h ($DO_{600} \approx 0,4 - 0,5$) dans un milieu riche TY supplémenté en Atc (10ng/mL) en duplicat biologique. Pour chaque souche, des dilutions des deux cultures indépendantes ont été déposées en gouttes sur un milieu solide TY supplémenté avec du NaCl 400 mM ou sur

un milieu TY à pH 5,5 puis les boîtes ont été incubées à 37°C pendant 24h. La souche mutante semble présenter une légère augmentation de tolérance au NaCl par rapport à la souche sauvage et la complémentation permet de réverter ce phénotype (**Figure 42B**). En revanche aucune différence entre la souche sauvage et la souche mutante n'a été observé en milieu acide (**Figure 42D**).

III. Étude du rôle d'Hfq dans la virulence de *C. difficile*

Outre les phénotypes de déficience de croissance et de sensibilité aux stress accrue des mutants Δhfq , il a également été montré que cette protéine est requise pour le *fitness* et la virulence de nombreuses bactéries pathogènes^{467,678,682}. En effet, les mutants de délétion pour le gène *hfq* sont souvent plus sensibles aux mécanismes de défenses de l'hôte et ont une virulence fortement atténuée dans les modèles animaux, bien qu'il y ait une variation considérable à la fois dans la sévérité et l'étendue des phénotypes⁶⁸². Ces phénotypes de virulence sont cependant bien plus marqués chez les bactéries à Gram négatif que chez les bactéries à Gram positif pour lesquels, hormis *L. monocytogenes*, aucun phénotype de virulence lié à la délétion de *hfq* n'est observé. Afin de tester l'impact de la délétion du gène *hfq* chez *C. difficile*, une expérimentation animale en modèle murin conventionnel (souris femelles C57BL/6) a été réalisée. Le modèle animal utilisé est le modèle murin de Chen et *al.*⁶⁵⁴ déjà utilisé pour l'expérience de Dual RNA seq. Les protocoles ont été approuvés par le Comité d'éthique en expérimentation animale C2EA-26 (n° d'autorisation APAFIS#30300-2021042211153677 v1). Au total, 16 souris conventionnelles femelles C57Bl/6 (Laboratoires Charles River, France), hébergées dans l'animalerie de la faculté de Pharmacie de l'Université Paris-Saclay (n° d'agrément B92-019-01), ont été réparties de manière aléatoire en deux groupes et traitées par le cocktail antibiotique décrit plus haut. Un jour après l'administration IP de clindamycine, les souris du premier groupe ont été infectées par 10^5 spores de la souche sauvage CD630E et les souris du deuxième groupe ont été infectées par 10^5 spores de la souche CD630E Δhfq (gavage oral dans un volume de 0,5 ml). Les souris ont été suivies pendant 29 jours et ont eu accès pendant toute la durée des expériences à de l'eau et de la nourriture stérile *ad libitum* (**Figure 43**).

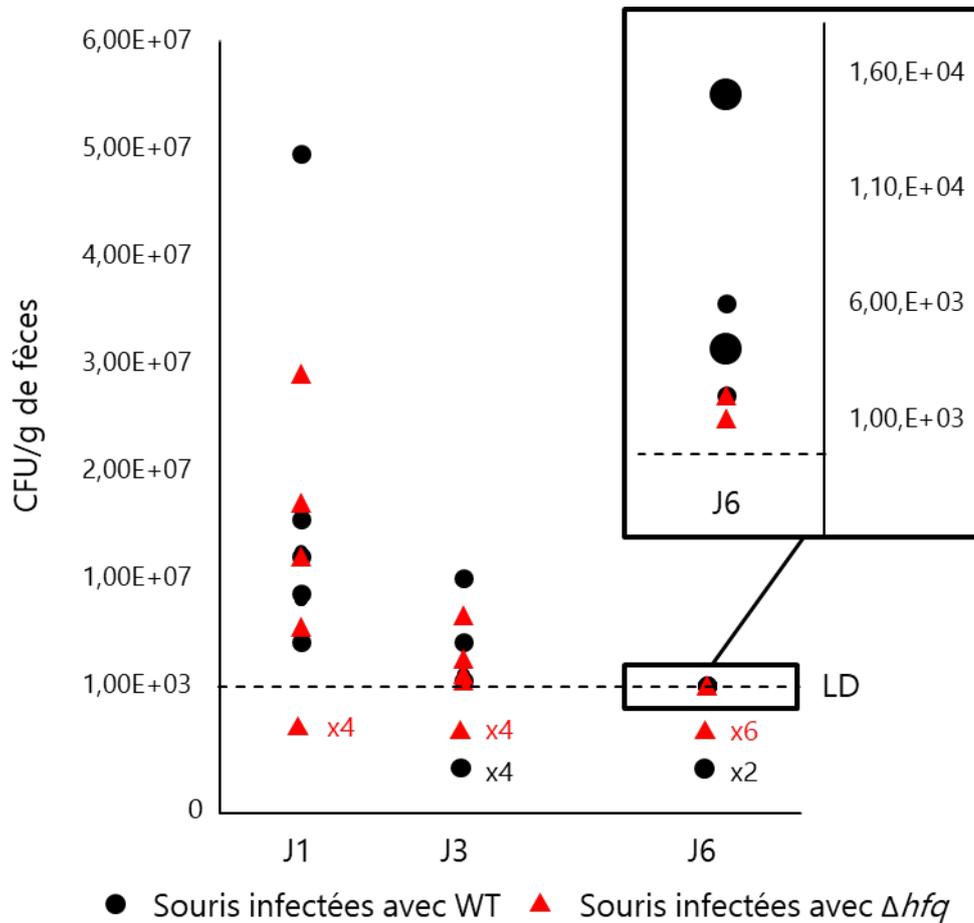


Figure 44 – Evolution de la charge bactérienne de *C. difficile* (formes végétatives) dans les fèces chez les souris infectées avec la souche CD630E WT (noire) et Δhfq (rouge). Les souris infectées avec la souche sauvage sont représentées par un rond noir et celles infectées par la souche mutante Δhfq par un triangle rouge. Les points de diamètres plus importants à J6 représentent deux points qui se superposent. Le seuil de détection est indiqué par un trait discontinu. Les seuils de détection sont 1.10^4 CFU/g de fèces pour les jours 1 et 3, et 1.10^3 CFU/g de fèces pour le jour 6.

Une injection intra-péritonéale (IP) de clindamycine a été réalisée au 27^e jour post-infection afin d'induire une dysbiose susceptible de favoriser la germination de spores persistantes dans le tractus intestinal des souris de manière à comparer le potentiel de rechute des deux souches. Une détection semi-quantitative des cellules végétatives dans les fèces 24h plus tard a montré un regain de colonisation de *C. difficile* de manière équivalente entre les deux groupes de souris infectées.

Partie 3 : Caractérisation d'un ARNnc spécifique des souches de ribotype 027

I. Identification et caractérisation de l'*IGR-1461*

Lors d'une expérience de transcriptomique comparative entre la souche CD630E et la souche hypervirulente de PCR-ribotype 027 CDR20291, plusieurs ARNnc ont été identifiés comme étant spécifiquement exprimés chez la souche CDR20291 par rapport à la souche CD630E (*in vitro*). Parmi ces ARNnc, nous en avons sélectionné un (*IGR-1461*) qui présentait à la fois un fort signal par analyse RNA-seq et Northern-Blot et qui ne présentait aucune équivalence dans la base de données Rfam (*the RNA families database*) (**Figure 45 A et B**). Dans la souche CDR20291, cet ARNnc se trouve dans un prophage (phi027) hautement conservé chez la majorité des souches de PCR-ribotype 027 (**Figure 45 C**)⁵¹.

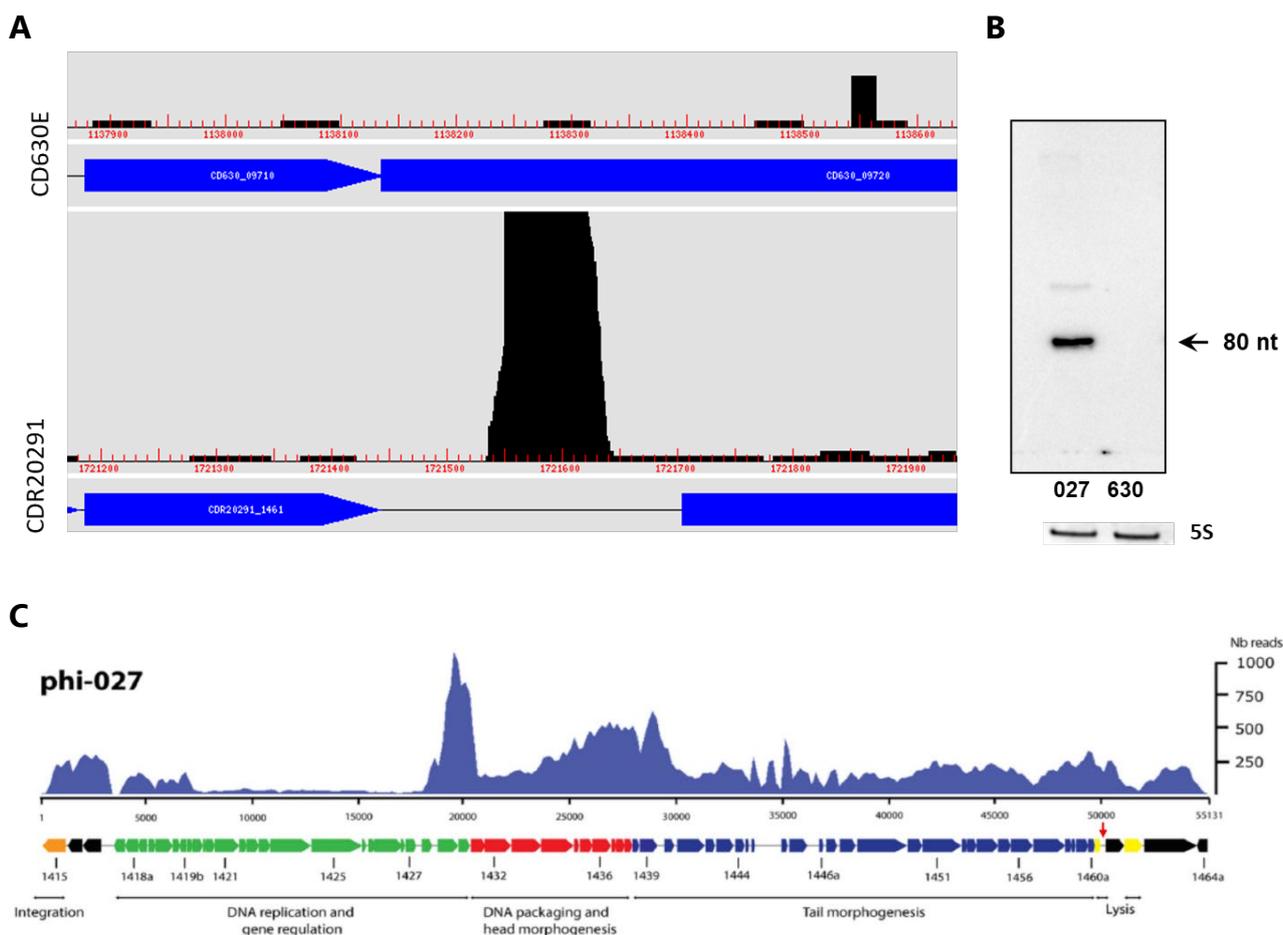


Figure 45 – Identification par RNA-seq de l'*IGR-1461* localisé dans la séquence du prophage phi027. Comparaison des profils (A) RNA-seq et (B) Northern-Blot des souches CD630E et CDR20291 pour l'expression de l'*IGR-1461*. L'ARN 5S sert de contrôle. (C) Profil RNA-seq du prophage phi027 issu d'une autre étude³⁵⁸, spécifique de la souche CDR20291, contenant l'*IGR-1461* (indiqué par une flèche rouge).

50 gènes putatifs (*CDR20291_1415-1465*) ont initialement été annotés dans la séquence du prophage phi027 présent dans le génome de la souche CDR20291, puis 20 autres gènes ont été ajoutés plus récemment après une ré-annotation de cette région prophagique^{51,358}. Une étude de la conservation de l'ARNnc *IGR-1461* par analyse BLAST a confirmé sa conservation majoritairement dans des souches de PCR-ribotype 027 ainsi qu'avec un phage tempéré phiC2 structurellement proche des prophages 1 et 2 de la souche CD630E (**Tableau 4**)⁶⁸³. L'étude de la conservation de la région génomique dans laquelle est situé l'*IGR-1461* a révélé une forte conservation de cette région chez les souches BI et CD196 de PCR-ribotype 027 (**Figure S10**). La région semble également assez bien conservée chez le phage phiC2 et les souches 7032989, 7032994 et 7032985 non toxigènes de PCR-ribotype 010⁶⁸⁴.

Tableau 4 – Étude de la conservation de l'*IGR-1461* par BLAST. Utilisation de la base de données nr/nt avec les paramètres par défaut. Les souches présentées dans le tableau ci-dessous ont été sélectionnées manuellement.

Souches	PCR-Ribotype	Couverture	Identité de séquence
DSM 27640/27638	027	100 %	100 %
BI1	027	100 %	100 %
CF5	017	100 %	100 %
CD196	027	100 %	100 %
Phage ΦC2	-	100 %	100%
Z31	009	100%	100%
M68	017	100 %	99,13 %
CD161	027	100 %	98,26 %

Une prédiction informatique des cibles de cet ARNnc a été réalisée par les algorithmes IntaRNAv2⁶⁸⁵ et RNApredator⁶⁸⁶. Ces algorithmes ont identifié plusieurs cibles potentielles et les quinze meilleurs *hits* communs aux deux prédictions ont été sélectionnés (**Tableau S3**). Il ne semble pas y avoir de catégorie fonctionnelle majoritaire concernant les cibles prédites qui sont majoritairement chromosomiques. Parmi ces cibles, on peut cependant noter la présence de trois gènes codant des protéines hypothétiques du prophage phi027. Ces prédictions suggèrent donc un rôle régulateur de cet ARNnc à la fois sur l'expression de gènes de la bactérie mais également sur certains gènes du prophage phi027. La zone d'interaction prédite de l'ARNnc avec ses cibles putatives semble présenter un consensus

qui correspond à une structure tige boucle identifiée lors de la prédiction de la structure secondaire de l'*IGR-1461* par Mfold (**Figure 46**). En revanche, il ne semble pas y avoir de région ciblée consensus entre les différentes cibles prédites qui peuvent interagir avec l'ARNnc *via* différentes parties de leur séquence.

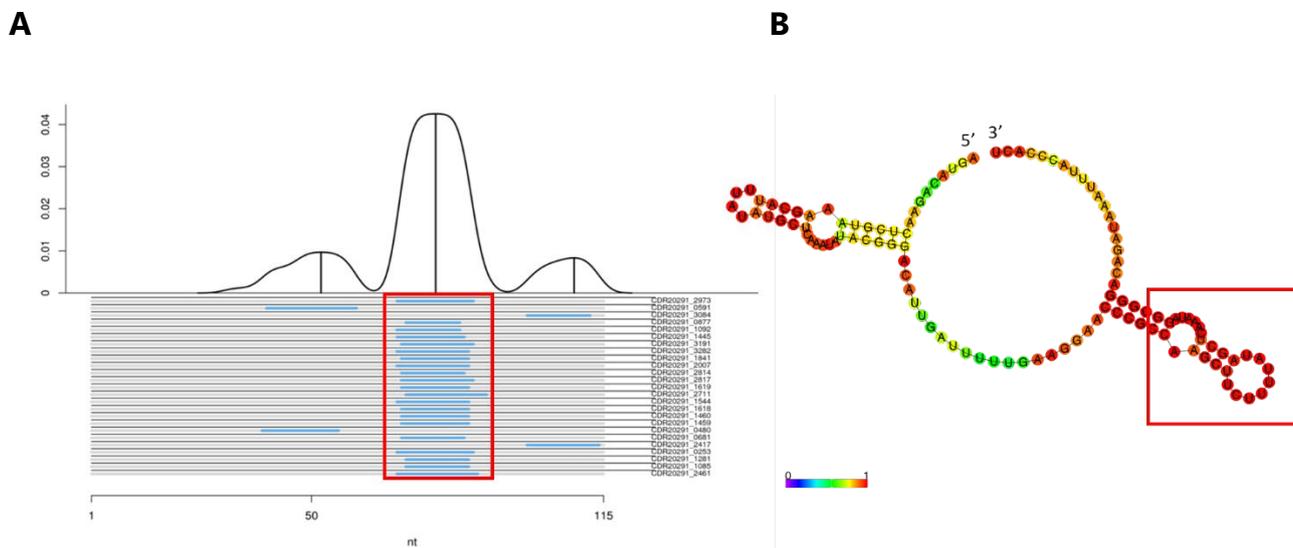


Figure 46 – Prédiction de la région d'interaction de l'*IGR-1461* avec ses cibles prédites. (A) Région de l'*IGR-1461* prédite (IntaRNA) pour l'interaction avec ses cibles potentielles. La *p-value* ajustée est représentée sur l'axe des ordonnées et la position de la séquence de l'*IGR-1461* est indiquée sur l'axe des abscisses. Les régions d'interactions avec les cibles sont représentées par des lignes bleues. (B) La région d'interaction consensus est encadrée en rouge sur la structure secondaire prédite de l'*IGR-1461* (Mfold).

L'*IGR-1461* est situé entre deux ORFs. Le gène *CDR20291_1461* localisé en amont de l'*IGR-1461* code pour une protéine *holin* impliquée dans la dégradation de la paroi cellulaire de l'hôte à la fin du cycle lytique. Bien que sa séquence soit conservée à 91% avec celle du gène *CD630_09710* situé dans le prophage 1 de la souche CD630E, la région contenant l'ARNnc est absente de la souche CD630E. Le gène *CDR20291_1462* situé en aval de l'*IGR-1461* code pour une protéine hypothétique appartenant à la famille Abi (*abortive infection*) déclenchant le système de mort cellulaire altruiste durant l'infection de la cellule par des phages. De la même manière que pour l'*IGR-1461*, la séquence de ce gène est également très conservée chez les souches de PCR-ribotype 027 et chez le phage phiC2. L'analyse par BLAST de la séquence de la protéine codée par ce gène a confirmé son appartenance à la superfamille Abi, et révèle également deux *hits* AbiF-like dans deux souches différentes de *C. difficile*, s'alignant chacun sur les deux moitiés de la protéine CD1462 avec une identité de séquence de 98% et 100% (**Figure S11**). Ce système AbiF identifié chez *L. lactis* permet à la bactérie de résister à l'infection phagique en agissant au niveau de la réplication de l'ADN

du phage ⁶⁸⁷. De plus, l'analyse par BLAST de la séquence de l'*IGR1461* dans la base de données Rfam associe cet ARN un motif abiF conservé qui a récemment été découvert par bio-informatique ⁶⁸⁸. Ces deux éléments suggèrent un lien de corrélation entre le système AbiF et la protéine CD1462 du phage phi027.

Une prédiction de promoteurs (BProm) pour ces deux éléments génétiques n'a permis d'identifier qu'une seule région promotrice en amont de l'*IGR-1461* (**Figure 47**). De plus un terminateur Rho-indépendant a été prédit entre la séquence de l'*IGR-1461* et celle du gène *CD_1462*. Ce schéma structurel associé au profil RNA-seq de ces deux éléments suggère une nature *riboswitch* de l'ARNnc identifié qui régulerait négativement l'expression du gène en aval *CD_1462*. L'identification d'un motif ARN abiF (Rfam : RF03085) de l'*IGR-1461* a permis de prédire la structure secondaire du *riboswitch* présentant trois tiges boucles dont deux aptamères abiF conservés et un terminateur Rho indépendant ⁶⁸⁸. Afin de pouvoir déterminer le rôle régulateur de l'*IGR-1461* sur le gène *CD_1462*, la construction de mutants de délétion a été entreprise.

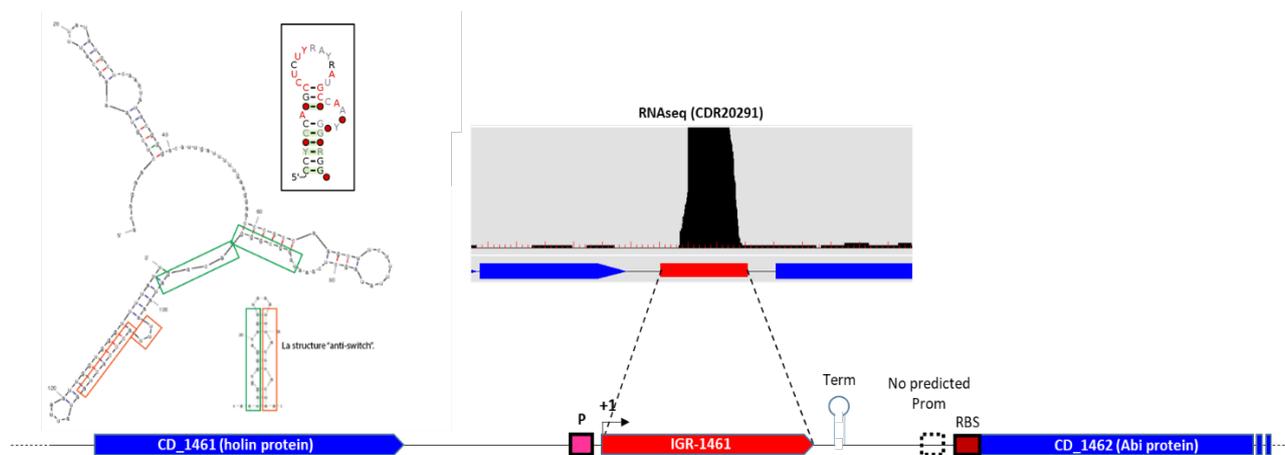


Figure 47 – Organisation génétique de l'*IGR-1461* et du gène *CD_1462*. La prédiction de la structure du *riboswitch* est indiquée sur la partie gauche de la figure. L'aptamère abiF conservé (RF03085) à partir duquel la prédiction a été faite est indiqué au-dessus de la structure du *riboswitch*. Les deux aptamères déduits se trouvent aux positions 7-39 et 59-93. Le terminateur débute à la position 106. La structure anti-switch permettant la libération du terminateur est encadrée en vert et orange. Les profils d'expressions de l'*IGR-1461* et du gène *CD_1462* sont également rappelés dans cette figure.

II. Construction et caractérisation du mutant de délétion

1. Identification des extrémités par 5'-3' RACE

Avant d'entamer la construction des mutants de délétion pour le *riboswitch* et le gène *CD_1462*, l'extrémité 5' précise de l'*IGR-1461* a été déterminée par 5'-RACE (*Rapid*

Amplification of cDNA-ends by polymerase chain reaction). Cette technique permet l'amplification de séquences d'acides nucléiques à partir d'un site interne d'un ARNm circularisé et converti en ADNc dont l'extrémité 5' est inconnue. L'ARNm est d'abord circularisé par traitements successif d'une TAP (Tobacco Acid Pyrophosphatase) et d'une ligase qui permet la ligation des extrémités. L'ARNm circulaire est ensuite converti en ADNc *via* une reverse transcription (RT) et d'un primer spécifique s'hybridant en 3'. L'ADNc circulaire est par la suite amplifié par PCR grâce à des oligonucléotides spécifiques qui s'hybrident en sens inverse à un site interne connu de la séquence. Afin d'augmenter le nombre d'amplicons spécifiques obtenus, une deuxième PCR avec des amorces internes (*Nested PCR*) est également réalisée. La liste des oligonucléotides utilisés pour ces différentes étapes est fournie dans le **Tableau S4**. Le séquençage des amplicons obtenus permet de déterminer la séquence de l'extrémité 5' au niveau du site de ligation des deux extrémités. Le séquençage de 11 amplicons obtenus pour l'*IGR-1461* a permis de déterminer une extrémité 5' au nucléotide près pour cet ARNc. De plus, dans la mesure où aucun promoteur n'a pu être prédit pour le gène *CD_1462*, la même expérience a été réalisée avec son transcrite de manière à déterminer le +1 de transcription de ce gène. Le séquençage des amplicons du 5'-RACE a révélé un +1 de transcription chevauchant la partie 5' de la séquence de l'*IGR-1461*, allant donc dans le sens de l'hypothèse d'une régulation de l'expression du gène *CD_1462* par l'*IGR-1461*. Les séquences des amplicons obtenus sont détaillées en **Tableaux S5 et S6** et leurs positions sur le génome sont illustrées en **Figure S12**.

2. Stratégie de mutagenèse

La construction de trois mutants de délétion a été entreprise afin de pouvoir étudier le système *riboswitch-Abi* dans son ensemble : (1) un mutant de délétion $\Delta IGR-1461$ permettant d'étudier le rôle de l'*IGR-1461* dans la régulation du gène *CD_1462- (abi)* et dans d'autres cibles prédites localisées en *trans*, (2) le mutant de délétion ΔCD_1462 permettant d'étudier le rôle de ce gène lors de l'infection phagique, et (3) le double mutant de délétion $\Delta IGR-1461\Delta CD_1462$ permettant de contrôler la spécificité des effets des délétions simples pour chacun des deux gènes. Pour cela, une stratégie de mutagenèse par échange de couples alléliques d'une version tronquée du (ou des) gène(s) à déléter a été utilisée. Cette

stratégie repose sur l'obtention et la sélection de doubles recombinants entre la version plasmidique tronquée de la cible à déléter flanquée par deux bras d'homologies et la région homologue du génome contenant le ou les gènes à remplacer. Les cellules non mutantes et/ou qui sont toujours porteuses du plasmide de mutagenèse sont ensuite contre-sélectionnées grâce à un système TA de type I de *C. difficile* récemment identifié par le laboratoire⁶¹¹. L'induction de l'expression du gène de la toxine (*via* le promoteur inducible pTet) localisé sur le plasmide permet de ne sélectionner que les cellules double-recombinantes et qui ne possèdent plus le plasmide. L'antitoxine est également associée au gène de la toxine afin de pallier les fuites du promoteur pTet. La carte du plasmide est détaillée en **Figure S13**.

Historiquement, les clostridies, et particulièrement *C. difficile*, sont réputés pour être difficilement manipulables génétiquement par rapport aux autres organismes modèles comme *E. coli* ou *Bacillus subtilis*. En effet, bien qu'il existe de nombreux outils génétiques pour manipuler *C. difficile*, chacun d'entre eux requiert une étape de transfert par conjugaison en utilisant *E. coli* ou *B. subtilis* comme bactérie donneuse⁶⁸⁹. De plus, la souche CDR20291 utilisée pour la construction des mutants étant décrite pour être une souche encore plus difficile à conjuguer que d'autres (notamment la souche CD630E), nous avons dû appliquer le protocole publié par Kirk et Fagan qui inclut une étape de choc thermique à 50°C pendant 15 min⁶⁹⁰. D'après les auteurs, cette étape permettrait d'augmenter l'efficacité de conjugaison d'un facteur 10³. L'utilisation de ce protocole a permis le transfert du plasmide de mutagenèse pour la délétion de l'*IGR-1461* et l'obtention d'un mutant de délétion. La délétion de l'*IGR-1461* a été confirmée par qPCR (**Figure S14**). Les deux autres mutants n'ont cependant pas pu être obtenus. En dépit de multiples tentatives et changement de paramètres expérimentaux (temps de conjugaison, temps et température de choc thermique, matériel, expérimentateur, etc.), aucun autre mutant de délétion n'a pu être obtenu avec ce protocole. Afin de n'exclure aucune piste, nous avons décidé de remplacer notre souche *E. coli* donneuse utilisée en routine dans le laboratoire (HB101/RP4) par une souche récemment publiée et décrite comme augmentant l'efficacité de conjugaison avec la souche CDR20291 d'un facteur 100⁶⁹¹. Le profil de méthylation de la

souche donneuse est connu pour affecter l'efficacité du transfert de l'ADN chez de nombreuses bactéries dont des clostridies. *C. difficile* possède un système de restriction de type IV et, afin de contourner cette barrière potentiellement responsable de la faible efficacité de conjugaison de certaines souches de clostridies (dont la R20291), les auteurs de cet article proposent une nouvelle souche donneuse *E. coli* (sExpress) dépourvue du gène de l'ADN-cytosine méthyltransférase (*Dcm*). Cette nouvelle souche donneuse n'a cependant pas permis d'obtenir de meilleurs résultats puisque sa croissance semble affectée lorsqu'elle porte le plasmide de mutagenèse. L'aspect visqueux des cultures de nuit et une morphologie ronde de la plupart des cellules suggère une lyse cellulaire.

3. Mécanisme de régulation par le riboswitch

L'étude de la régulation de l'expression du gène *CD_1462* par le *riboswitch* a été réalisée dans un premier temps par qPCR avec le mutant $\Delta IGR-1461$. L'expression du gène *CD_1462* chez le mutant de délétion a été comparée à celle de la souche sauvage. Cette analyse a révélé une forte augmentation (d'un facteur 12) du niveau d'expression du gène *CD_1462* chez le mutant $\Delta IGR-1461$ par rapport à la souche sauvage CDR20291 (**Figure 48**).

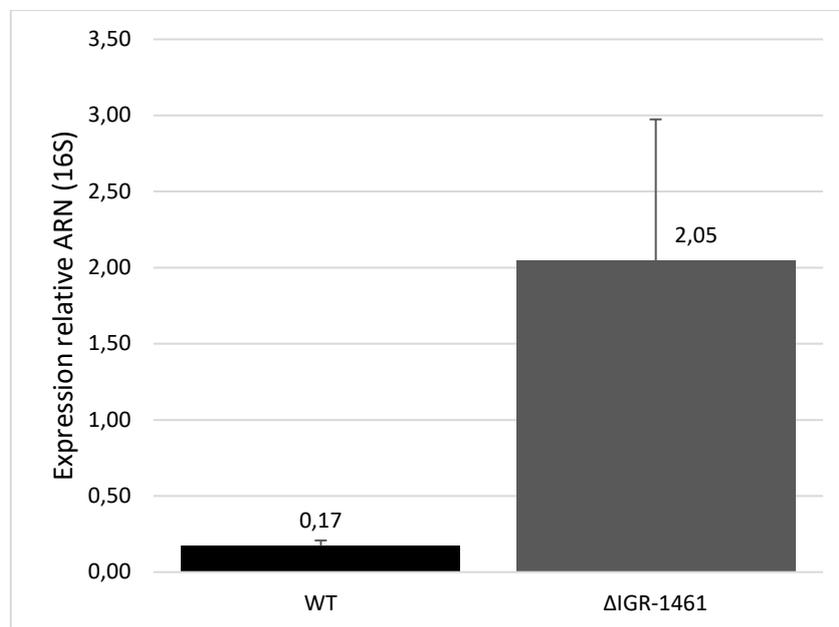


Figure 48 – Analyse par qPCR de l'expression du gènes *CD_1462* chez la souche WT ($n = 2$) et la souche mutante $\Delta IGR-1461$ ($n = 4$).

Ce résultat confirme donc le rôle de régulateur négatif du gène *CD_1462* par le *riboswitch* putatif *IGR-1461*. Dans la majorité des cas, les *riboswitches* agissant au niveau transcriptionnel sont complexés avec un ligand qui induit la formation d'une structure

terminatrice (en *tige-boucle*) dans le domaine régulateur, empêchant ainsi la transcription du gène en aval. Bien que la fixation d'un ligand qui induit la déstructuration de l'aptamère permettant ainsi la transcription soit également possible, nous avons choisi de tester en premier l'hypothèse la plus probable, c'est-à-dire la liaison d'un ligand au *riboswitch IGR-1461* empêchant la transcription du gène en aval *CD_1462*. Afin d'analyser la régulation transcriptionnelle du gène *CD_1462* et l'origine du ligand, plusieurs fusions transcriptionnelles avec le gène rapporteur *phoZ* codant pour une phosphatase alcaline ont été construits dans des plasmides (**Figure 49**).

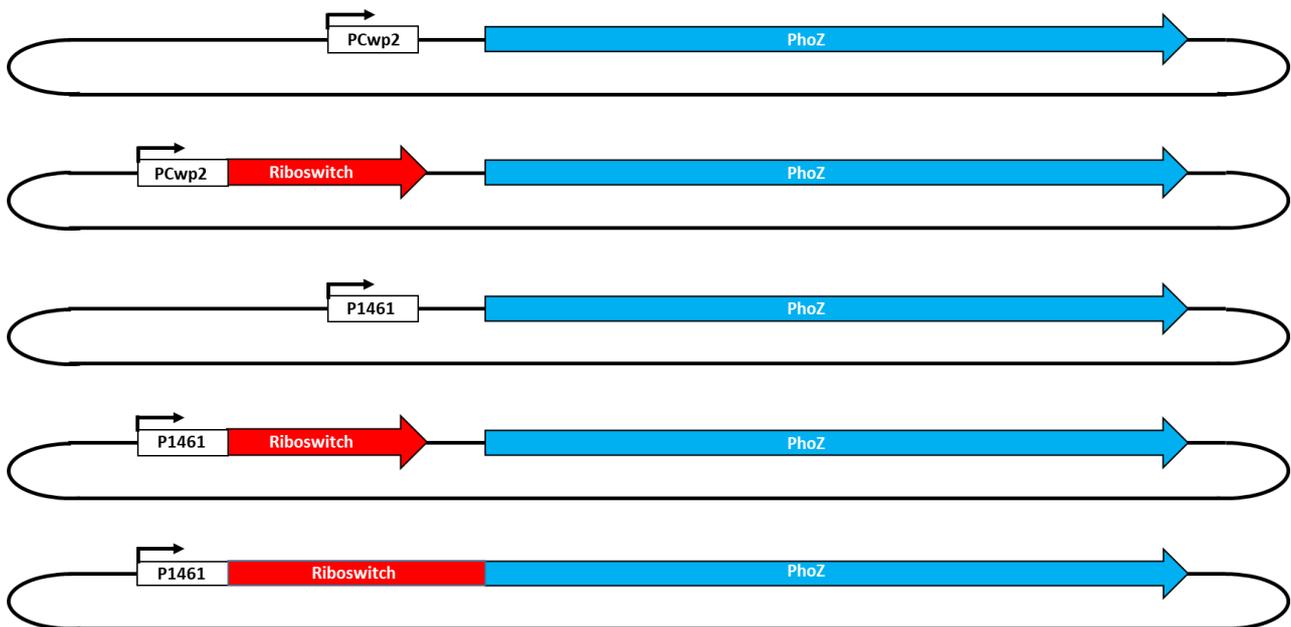


Figure 49 – Représentation schématique des différentes constructions du système rapporteur *phoZ*. Ces plasmides ont été conjugués dans les souches CD630E WT, CDR20291 WT, CDR20291 Δ phi0-27 et CDR20291 Δ IGR-1461. P1461 est le promoteur naturel de l'IGR-1461. PCwp2 est un promoteur constitutif fort utilisé afin de contrôler tout effet de régulation dépendant du promoteur P1461.

Les plasmides construits ont été transférés par conjugaison dans différentes souches de *C. difficile* afin de tester plusieurs hypothèses. Il est intéressant de noter que si l'utilisation de la souche donneuse *E. coli* sExpress *Dcm*- n'a pas été fructueuse pour les plasmides de mutagenèse, elle en revanche a permis de faire passer l'ensemble des systèmes rapporteurs dans la totalité des souches de *C. difficile*, ce qui n'a pas été possible avec la souche HB101/RP4. Les souches qui ont été conjuguées avec les différents plasmides sont : la souche CD630E (WT), la souche CDR20291 (WT), la souche CDR20291 Δ IGR-1461 et une souche CDR20291 Δ phi027 récemment construite par nos collaborateurs Canadiens de l'université de Sherbrooke au cours d'une étude sur le rôle du prophage phi027 dans le

contrôle de la production des toxines TcdA et TcdB dans les souches de ribotype 027 (en cours de publication). La souche CD630E permet de tester l'hypothèse de la spécificité du ligand à la souche CR20291 et doit donc être comparée avec cette dernière. La souche CDR20291 $\Delta\phi027$, délétée pour la région prophagique $\phi027$ entière, permet de vérifier si le ligand est produit spécifiquement par le prophage afin de contrôler l'expression de son système Abi. La souche CDR20291 $\Delta IGR-1461$ quant à elle est un contrôle interne permettant de s'assurer que l'effet observé n'est dû qu'au *riboswitch* cloné sur le plasmide et non au *riboswitch* endogène. Pour chaque système étudié, l'unité enzymatique (UE) de la phosphatase alcaline (PA) a été déduit à partir de la DO_{420} d'un lysat de culture cellulaire en phase exponentielle exposé à du para-nitrophénylphosphate (PNP). Pour chaque échantillon, la DO_{420} obtenue est normalisée par le temps d'apparition de l'indicateur coloré après ajout du réactif PNP (t_{min}), la DO_{600} de la culture cellulaire et le volume cellulaire prélevé (V_{cell}) selon la formule suivante :

$$UE = \frac{(DO_{420} - (1,75 \times DO_{550})) \times 1000}{t_{min} \times DO_{600} \times V_{cell}}$$

Les résultats des UE obtenus pour chaque construction dans les différentes souches sont présentés en **Figure 50**. Le premier système étudié correspond au *riboswitch* sous le contrôle de son promoteur natif cloné en amont du gène *phoZ* (P1461+*riboswitch*). Le signal détecté grâce à cette construction a été comparé au signal de la PA seule sous le contrôle du promoteur natif du *riboswitch* (P1461). Cette construction ne semble pas être fonctionnelle dans la mesure où aucun signal n'a été détecté dans aucune des souches testées. Dans la mesure où un signal est obtenu pour la construction P1461+*riboswitch*, on peut supposer que la partie avale du promoteur P1461 ait été coupée trop tôt lors de la construction du plasmide. La région promotrice initialement clonée a donc été rallongée d'une vingtaine de nucléotide en aval et clonée à nouveau dans le système rapporteur. Cette construction récemment obtenue est en cours d'analyse et devrait permettre d'obtenir de nouveaux résultats très prochainement.

Afin de contrôler tout éventuel effet dépendant du promoteur P1461, un système *riboswitch-phoZ* sous le contrôle d'un promoteur fort constitutif (PCWP2) a été construit.

On constate, de la même manière que ce qui avait été observé lors de la qPCR du gène *CD_1462* chez le mutant $\Delta IGR-1461$ et chez le WT, une très forte diminution de l'expression de la PA en présence du *riboswitch* (PCWP2+*riboswitch*) par rapport au promoteur seul (PCWP2). Cela suggère un mécanisme de régulation du gène *CD_1462* par l'*IGR-1461* en *cis* et renforce donc l'hypothèse de la nature *riboswitch* de cet ANRnc. Ces résultats associés à ceux de la qPCR permettent également de supposer que la tendance attendue pour le système rapporteur avec le P1461 seul devrait être comparable à celle observée avec le promoteur PCWP2. Si l'on compare le signal obtenu entre les différentes souches, on constate que ce dernier est diminué quelle que soit la souche dans laquelle se trouve le *riboswitch*. Il semble donc a priori que le ligand ne soit ni spécifique de la souche CDR20291 ni du phage phi027. La différence de signal observée entre la souche CD630E et les 3 autres souches pourrait être dû à une différence de lyse. Cette hypothèse doit cependant encore être vérifiée en titrant les extraits protéiques obtenus pour chaque souche, puis en utilisant ces valeurs pour normaliser les UE de la PA.

Chez *Lactococcus lactis*, il existe un mécanisme de régulation traductionnelle du gène de défense *abiD1* contre l'infection du phage bIL66⁶⁹². Ce phage produit une endonucléase RuvC-like essentielle permettant de résoudre les structures d'ADN en branche générées durant la réplication³³². Il a été montré que lors de l'infection phagique, le transcrite du gène phagique *orf1* (codant pour la protéine RuvC-like) se lie à une structure secondaire localisée en amont du gène *abiD1* permettant la libération du RBS masqué par cette structure et induisant ainsi la traduction du transcrite *abiD1*. Afin de pouvoir tester ce modèle de régulation pour le gène *CD_1462* chez *C. difficile*, un système rapporteur de fusion traductionnelle a été réalisé entre l'*IGR-1461* et le gène *phoZ*. Ainsi, contrairement aux constructions précédentes, le *riboswitch* suivi du RBS du gène *CD_1462* ont été fusionnés au gène *phoZ* dépourvu de son RBS. Si l'on compare les résultats de la fusion traductionnelle avec ceux obtenus pour la construction P1461+*riboswitch*, on constate une baisse importante d'UE de la PA pour les souches CD630E, CDR20291 $\Delta IGR-1461$ et CDR20291 Δ phi027. Cela semble donc suggérer un mécanisme de régulation supplémentaire au niveau traductionnel qui permettrait de verrouiller la production de la protéine codée par le gène

CD_1462. Les résultats de la fusion traductionnelle dans la souche CDR20291 n'ont pas encore été obtenus. En effet, le transfert du système rapporteur portant la fusion traductionnelle dans cette souche a été plus compliqué à réaliser que pour les autres souches et n'a été obtenu que très récemment. Les tests sont en cours de réalisation et devraient pouvoir donner des résultats très prochainement.

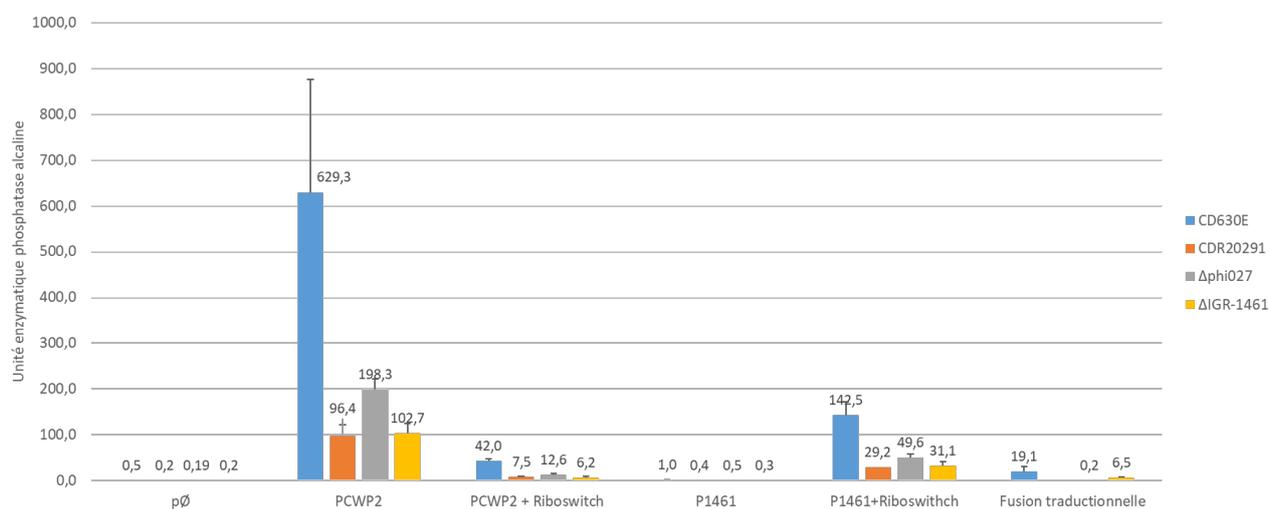


Figure 50 – Étude du mécanisme de régulation de l'IGR-1461 dans un système rapporteur utilisant le gène de la phosphatase alcaline. Les valeurs moyennes ($n = 4$) des UE obtenues pour chaque construction sont représentées sur ce graphique. Les constructions correspondantes sont représentées en Figure 49. Les résultats de la construction de la fusion traductionnelle dans la souche CDR20291 n'ont pas encore été obtenus et ne sont donc pas représentés ici.

III. Impact de l'expression de la protéine Abi sur la croissance

Les systèmes Abi permettent d'interférer avec l'infection phagique en agissant à différentes étapes du cycle infectieux du phage, aboutissant ainsi à la mort de la cellule infectée afin de prévenir la propagation des particules phagiques au sein de la communauté bactérienne. Afin de pouvoir vérifier l'appartenance de la protéine codée par le gène *CD_1462* à la famille Abi, des tests de croissance cellulaire ont été réalisés sur des cellules surexprimant la protéine CD1462. Pour cela le gène *CD_1462* a été cloné dans un vecteur d'expression (dérivé du pRPF185) et placé sous le contrôle du promoteur inductible fort pTet (souche CDR20291/p-*CD_1462*). Le vecteur d'expression vide a également été transféré dans 2 souches contrôles, la souche WT CDR20291 (CDR2091/p) et la souche mutante délétée pour l'ARNnc *IGR-1461* (CDR20291Δ*IGR-1461*/p) qui, comme vu précédemment, exprime fortement le gène *CD_1462* (**Figure 48**). Un suivi de culture en plaque en milieu riche TY supplémenté avec de l'ATc (50 ng/μL) de ces 3 souches a été réalisé sur une durée de 23h

(avec mesure de DO_{600} toute les 30 min) (**Figure 51**). On constate un décalage de la courbe de croissance de la souche surexprimant la protéine CD1462 (R20291 + pTet Abi) par rapport à la souche sauvage R20291 montrant donc un impact de cette protéine sur la croissance cellulaire. La reprise de croissance de la culture cellulaire pour finalement atteindre la phase stationnaire pourrait s'expliquer par la réduction de la stabilité de l'ATc à 37°C qui diminuerait donc avec le temps l'induction de la protéine CD1462 permettant ainsi aux cellules de se diviser à nouveau. De plus, bien que le niveau d'expression du gène *CD_1462* soit plus élevé dans la souche CDR20291 Δ IGR-1461/p que la souche CDR20291/p, aucune différence de croissance n'est observée entre ces deux souches (**Figure 51**).

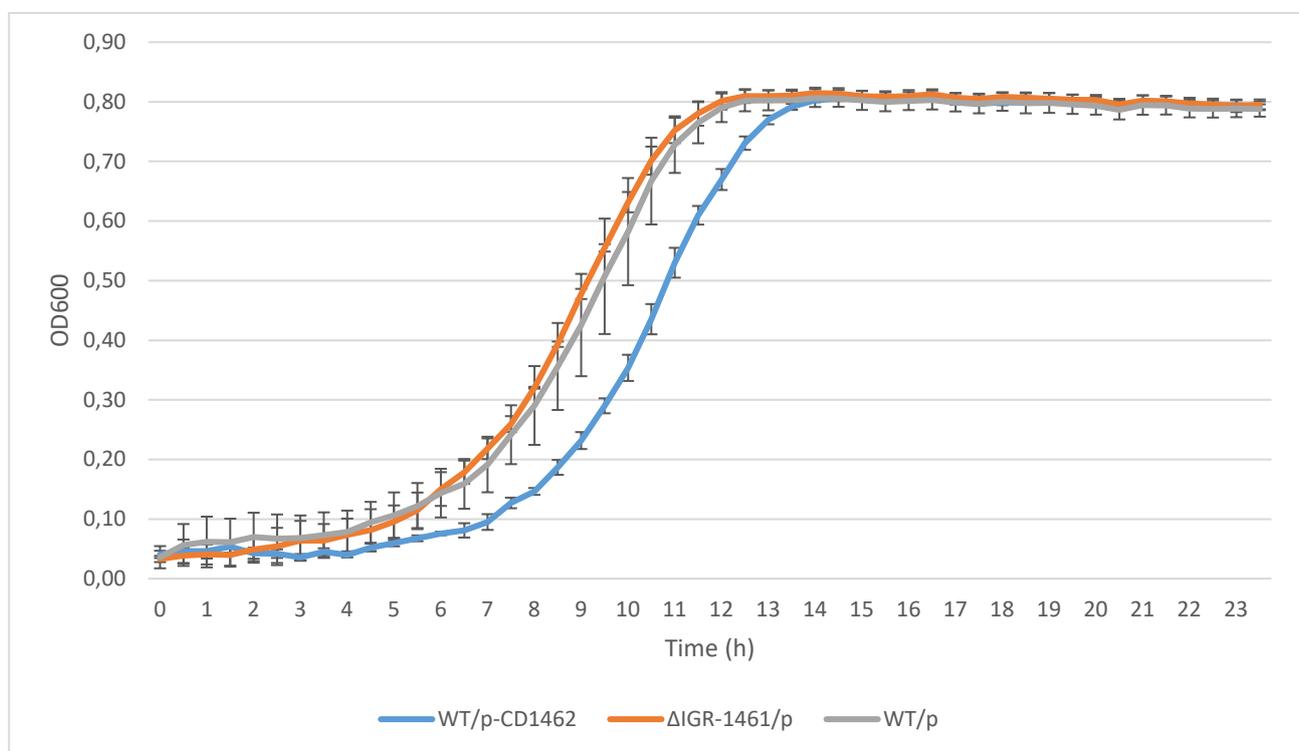


Figure 51 – Courbes de croissances des souches CDR20291/p, CDR20291 Δ IGR-1461/p et CDR20291/p-*CD_1462*. Les trois souches CDR20291/p, CDR20291 Δ IGR-1461/p et CDR20291/p-*CD_1462* ont été cultivées en microplaque 24 puits (scellée) pendant 23h dans du milieu TY supplémenté avec 7,5 μ g/mL de Tm et 50 ng/mL d'ATc. Les valeurs des DO_{600} ont été mesurées automatiquement toutes les 30 min. Les valeurs moyennes des DO_{600} ($n = 3$ par souche) des trois souches ont été utilisées pour l'obtention des courbes de croissances (/p pour le plasmide vide ; /p-*CD_1462* pour le plasmide d'expression du gène *CD_1462*).

Discussion et perspectives

C. difficile est un entéropathogène majeur associé aux soins chez l'adulte dans les pays industrialisés. Il est responsable de diarrhées modérées à sévères associées majoritairement à la prise d'un traitement antibiotique et peut entraîner des inflammations importantes de la paroi intestinale pouvant se manifester dans les formes les plus sévères par des colites pseudomembraneuses, des mégacôlons toxiques associées à une mortalité non négligeable. Ce projet de thèse fait suite à la récente identification par RNA-seq d'un grand nombre d'ARNnc chez *C. difficile* (plus de 250) incluant diverses classes (riborégulateurs agissant en *trans* et en *cis*, *riboswitch*, ARN CRISPR). Il est maintenant acquis que, chez de nombreuses bactéries pathogènes, les ARNnc jouent un rôle clé dans les processus de régulation de la virulence. Plusieurs approches expérimentales ont été mises en œuvre durant ce projet de thèse afin de caractériser des mécanismes de régulation originaux de la virulence par des ARNnc chez *C. difficile*.

I. Identification des ARN exprimés lors de l'infection

La compréhension de la pathogenèse des bactéries nécessite de connaître les gènes qui sont exprimés et leurs mécanismes de régulation durant l'infection. Tandis que le RNA-seq est devenu aujourd'hui une méthode de routine utilisée dans l'analyse de l'expression des gènes chez les pathogènes, d'autres techniques dérivant du RNA-seq ont émergé durant ces dernières années permettant d'aller au-delà de la simple analyse des profils ARNm, comme l'identification des réseaux de régulation post-transcriptionnelle contrôlés par les petits ARN et les protéines associées. Parmi ces approches, la technique dite de dual RNA-seq permettant d'étudier simultanément le transcriptome d'un pathogène et de son hôte durant l'infection a été développée il y a environ une dizaine d'années⁶⁴⁶. L'intérêt majeur de cette technique est de pouvoir analyser la variation de l'expression des gènes à différents temps dans deux organismes appartenant à deux domaines du vivant différents, et ce avec une profondeur et un niveau de précision élevés. Depuis son développement, la technique de dual RNA-seq a été appliquée à plusieurs bactéries pathogènes afin de pouvoir analyser simultanément les profils d'expression de leurs gènes et de ceux de l'hôte au cours de l'infection^{646,647}. Dans le cas des ICD, plusieurs études de transcriptomique ont déjà été réalisées sur des modèles *in vivo*^{120,649–653}, mais aucune étude de dual RNA-seq n'a été

décrite jusqu'à présent. Un protocole dual RNA-seq appliqué à *C. difficile* a donc été mis en place et optimisé. Afin de mimer les conditions infectieuses de *C. difficile* chez l'Homme, un modèle murin conventionnel a été utilisé conformément à ce qui avait été décrit dans la littérature ⁶⁵⁴. Une expérience de suivi clinique réalisée en amont a permis d'identifier trois points de cinétique infectieuse pour l'expérience de dual RNA-seq : un temps précoce à 8h correspondant au plateau de colonisation de la bactérie chez l'hôte, sans aucun signe clinique associé, et deux temps tardifs 28h et 32h permettant correspondant au début des signes cliniques associées à une réponse immunitaire tout en évitant le développement de symptômes trop forts qui, en raison des diarrhées, empêchent la récupération de matériel bactérien en quantité suffisante pour l'extraction des ARN. Les lyses du *caecum* et de son contenu ont dû être réalisées séparément pour des raisons techniques. En effet, lorsque le *caecum* et le contenu caecal sont placés dans le même tube de lyse, une mauvaise répartition des phases organiques et aqueuses a été observée après centrifugation des lysats cellulaires. L'apport excessif d'acides biliaires avec le matériel intestinal entraîne une basification du pH du phénol entraînant ainsi une inversion de la phase aqueuse contenant les ARN et de la phase phénolique. Ce problème a pu être résolu lorsque le *caecum* et son contenu ont été lysés séparément. Une lyse simultanée du *caecum* et des bactéries du microbiote pourrait être envisagée à condition de diviser au moins par deux la quantité de matériel lysé et d'effectuer plusieurs cycles afin de lyser la totalité des échantillons. Néanmoins, la séparation des échantillons du *caecum* et de son contenu pour la purification des ARNs permet de réajuster la proportion des ARN eucaryotes et procaryotes pour la suite de l'expérience. La purification par la méthode TRIzol a été choisie puisque c'est une méthode décrite comme particulièrement efficace pour l'extraction d'ARN à partir de tissus riches en RNAses, mais aussi pour l'isolement de petits ARN ⁶⁹³. Cette méthode a également été largement utilisée dans les études de dual RNA-seq publiées ⁶⁴⁷. Dans la plupart des études de dual RNA-seq décrites, la fraction de *reads* bactériens obtenus après séquençage est très faible ⁶⁴⁷. Ainsi, afin de pouvoir enrichir artificiellement les échantillons à séquencer en ARN procaryotes, les ARN extrait du *caecum* et du contenu caecal ont été purifiés séparément puis mélangés avec un ratio en quantité d'ARN totaux de deux ARN eucaryotes pour un ARN procaryote afin d'assurer une couverture maximale à la fois pour le génome de la souris et pour le génome

de la bactérie. Dans la mesure où la quantité des ARN eucaryotes purifiés était environ 10 fois supérieure à celle des ARN procaryotes, la totalité des échantillons procaryotes a été utilisée pour environ 1/5 des échantillons eucaryotes.

Le séquençage des échantillons ARN a permis d'obtenir en moyenne 100 millions de *reads* par échantillon. Afin de pouvoir maximiser le nombre de *reads* pouvant s'aligner sur le génome de *C. difficile*, la taille des *reads* conservés lors du nettoyage des données brutes a été abaissée à 10 nucléotides (au lieu de 30 habituellement). Cela a permis de récupérer entre 1 et 5 millions de *reads* totaux par échantillons. Les *reads* ne s'alignant pas par paires sur leur génome de référence ont été éliminés par la suite. Lors des expériences de dual-RNA-seq, les ARN des deux espèces étudiées sont séparés *in silico* lors de l'étape de *mapping* durant laquelle les *reads* sont alignés avec les deux génomes de références séquentiellement ou avec un génome concaténé (c'est-à-dire contenant les deux génomes dans un même fichier). Cette dernière option a été utilisée puisqu'elle demande moins de temps de calcul et permet au programme d'alignement utilisé de choisir le génome sur lequel le *read* s'aligne le mieux. La fraction de *reads* s'alignant de manière équivalente sur les deux génomes représente environ 3% des *reads* totaux et peut ainsi être éliminée⁶⁹⁴. L'utilisation successive de deux programmes d'alignement a permis de cartographier après la seconde étape entre 2000 et 16 000 *reads* supplémentaires par échantillon sur le génome de *C. difficile*. Cependant, la proportion de *reads* s'alignant sur le génome de *C. difficile* par échantillon est comprise entre 0,1% et 0,3% contre 20% et 70% pour la souris. Bien que cette faible proportion ne soit pas idéale pour l'identification de transcrits faiblement exprimés et la caractérisation d'effets subtils, les données obtenues sont tout de même analysables et interprétables. En effet, étant donné que le contenu total en ARN des cellules de mammifère dépasse largement celui d'un pathogène bactérien, la part de *reads* bactériens dans une expérience de dual-RNA-seq est généralement minoritaire. Néanmoins, environ 25 millions de *reads* totaux sont suffisant pour réaliser le profil de toutes les classes majeures d'ARNnc, incluant les ARN régulateurs eucaryotes (miRNA et lncRNA) et les ARNnc bactériens⁴⁰⁹. En revanche, l'absence presque totale de *reads* de *C. difficile* au temps T8h n'a pas permis d'exploiter ce temps dans les analyses. L'hypothèse d'un défaut de colonisation par *C.*

difficile à ce temps est exclue dans la mesure où les dénombrements de cellules végétatives de *C. difficile* sur milieu sélectif ont permis d'estimer le nombre de CFU/g de fèces de l'ordre de 10⁸. Bien que la fraction de *reads* pour *C. difficile* soit faible, cela n'est pas le cas pour les bactéries commensales du microbiote. En effet, les échantillons comportent entre 30% et 80% de *reads* d'espèces microbiennes constituant le microbiote intestinal. Ces *reads* ont été analysés par kraken2⁶⁶⁰, un logiciel basé sur une approche en *k*-mer permettant de réaliser une classification taxonomique de données métagénomiques. Cette analyse a mis en évidence un regroupement cohérent des différents échantillons non infectés et infectés entre eux à l'exception d'un échantillon issu du groupe de souris infectées à T32h (S5) qui possède un profil plus proche des souris non infectées que des souris infectées. On peut également noter un regroupement des souris infectées à T8h avec les souris non infectées, ce qui corrèle avec le nombre faible de *reads* de *C. difficile* obtenus pour ce groupe. La proportion des groupes microbiens ne peut pas être quantifiée facilement avec les données de méta-transcriptomique que nous avons obtenues à l'inverse des données de métagénomique basées sur le séquençage des fragments des gènes des ARNr 16S ou le séquençage des génomes. On peut cependant noter parmi les différentes espèces microbiennes de tous les échantillons une forte proportion du phage de *Proteus mirabilis* VB_Pmis_Isfahan récemment identifié. Il s'agit d'un phage lytique appartenant à la famille des *Siphoviridae* et regroupé dans le même clade que les phages vB_SenS_Sasha, 9NA et FLS_SP-06 de *Salmonella*⁶⁹⁵. Bien qu'il semble actif transcriptionnellement, la bactérie *P. mirabilis* est absente des espèces identifiées, suggérant donc l'infection d'une autre bactérie par ce phage.

Étant donné la trop faible proportion de *reads* de *C. difficile* au temps T8h, ce point a été exclu de l'analyse différentielle. De plus, la comparaison entre les temps T28h et T32h n'a montré aucune différence statistiquement significative entre ces deux points qui ont donc été regroupés en une seule condition (*Mouse Infected*; MI) qui a été comparée avec la condition *in vitro*. Lors d'analyses ultérieures, des temps moins rapprochés (comme 24h et 32h) pourront être considérés. La comparaison des profils d'expression des gènes de *C. difficile* entre des conditions *in vitro* et *in vivo* a permis de révéler environ 1372 gènes

significativement différentiellement exprimés dans les conditions *in vivo*. La répartition en nombre de gènes induits et réprimés est homogène avec environ 700 gènes induits et 600 gènes réprimés. Afin d'obtenir un premier aperçu des groupes de gènes avec des profils d'expression similaires, des analyses d'enrichissement basées sur un vocabulaire structuré et des voies ou fonctions qui sont significativement enrichies dans les gènes différentiellement exprimés ont été réalisées. En effet, les gènes corégulés présentent souvent des fonctions similaires ou partagent la même voie de régulation⁶⁹⁶. Une première analyse a été effectuée avec le programme DAVID. Cependant les noms des catégories fonctionnelles obtenues ne correspondaient pas avec les gènes référencés dans le cluster. Le programme COG, pour lequel la dernière version du génome de *C. difficile* avait récemment été ajoutée, a été utilisé dans un second temps⁶⁹⁷. Ce programme, initialement créé pour la classification des différentes familles de protéines, est largement utilisé aujourd'hui pour l'analyse fonctionnelle des gènes. Cette analyse a révélé une proportion élevée (25%) de gènes impliqués dans le transport et le métabolisme différentiellement exprimés *in vivo* par rapport aux conditions *in vitro*. Ce résultat reflète l'activité constante de *C. difficile* dans un milieu dont les nutriments sont en renouvellement permanent contrairement aux conditions *in vitro*. Bien que ces résultats donnent un premier aperçu fonctionnel des familles de gènes exprimés au cours de l'infection, l'étape complexe d'analyse fonctionnelle requiert d'effectuer plusieurs tests avec des programmes différents afin d'obtenir une synthèse interprétable et reflétant le plus possible une réalité biologique. L'analyse individuelle des gènes différentiellement exprimés a révélé une induction des gènes des toxines TcdA et TcdB (résultat confirmé par qPCR) ainsi que du gène codant pour l'adhésine CwpV impliquée dans l'auto-agrégation de *C. difficile* mais également dans la résistance aux phages^{356,665}. Des gènes de l'assemblage flagellaire comme *fliQ*, *flhA* et *motA* ont été détectés comme significativement induits dans les conditions *in vivo*. De nombreux gènes impliqués dans le transport et le métabolisme des sucres étaient réprimés, suggérant une sélection du type de glucide par *C. difficile* lors de l'infection. Un nombre important de gènes de sporulation étaient également induits. Seul un nombre restreint de transporteurs du fer ferreux (*feoA* et *feoB*) sont induits lors de l'infection, suggérant une préférence pour l'acquisition du fer sous cette forme. Tous ces résultats sont très cohérents avec les résultats d'une analyse

transcriptomique par *microarray* lors d'une infection de souris axéniques par la souche CD630⁶⁵³, ce qui valide notre modèle et notre protocole. En complément, environ 40 ARNnc sont induits lors de l'infection, et parmi eux se trouvent des ARNas, des ARN dépendant et indépendant de la protéine Hfq et des *riboswitch*. Parmi ces ARNnc, une dizaine des candidats les plus différentiellement exprimés (induits et réprimés) ont été sélectionnés. L'expression différentielle de ces ARNnc devra être vérifiée par qPCR et Northern-Blot, puis ils devront être caractérisés informatiquement et phénotypiquement grâce à des mutants de délétion. Nous rechercherons les cibles des ARNnc par des approches *in silico* et expérimentales, en particulier par la méthode MAPS (*MS2-affinity purification*) pour identifier des cibles des ARNnc codés *en trans*. En parallèle de l'analyse de nos données, une réanalyse de jeux de données RNA-seq issus de trois études différentes déjà publiés sur un modèle infectieux de *C. difficile* *in vivo* est en cours^{120,649,650}. En effet, les auteurs de ces études ne s'étaient pas intéressés aux ARNnc et la réanalyse de ces données avec les ARNnc sera comparée avec nos données de dual RNA-seq afin de pouvoir les confirmer ou apporter de nouveaux éléments. De plus, une étape de *data mining* avec le programme APERO⁶⁹⁸ qui permettra de détecter de nouveaux ARNnc spécifiquement exprimés dans des conditions *in vivo* à partir de données *paired-end* est également en cours.

Du côté de la souris, plusieurs conditions ont pu être analysées et comparées en elles. De la même manière que pour *C. difficile*, les profils d'expression des souris appartenant aux groupes T28h et T32h ne présentaient aucune différence significative et ont donc été regroupés. La comparaison des profils d'expression des gènes entre des souris infectées (T28h + T32h) et des souris non infectées a permis d'identifier 10 000 gènes différentiellement exprimés. Les profils d'expression des gènes de trois souris qui présentaient des symptômes caractéristiques des ICD (diarrhée, *caecum* inflammatoire et avec peu de contenu) à T28h et T32h ont été comparées avec les profils d'expression de souris ne présentant pas de symptôme aux temps T28h et T32h. Environ 1000 gènes différentiellement exprimés ont été détectés. Enfin, la comparaison des gènes des souris infectées du groupe T8h avec les souris des groupes T28h et T32h mélangés a mis en évidence environ 3000 gènes différentiellement exprimés. L'analyse fonctionnelle des gènes

différentiellement exprimés dans les différentes conditions comparées a été réalisée avec le programme DAVID. L'utilisation de DAVID pour la souris s'est avérée plus robuste que pour *C. difficile*. Cependant, comme commenté précédemment, d'autres analyses fonctionnelles devront être réalisées afin d'obtenir un profil le plus exact possible des groupes de gènes différentiellement exprimés lors de l'infection à *C. difficile*. Ces analyses ont permis d'enrichir un nombre important de clusters impliqués dans la réponse inflammatoire notamment chez les souris qui présentaient des symptômes. La comparaison des profils d'expression des gènes entre les différentes conditions testées a également permis de mettre en évidence un grand nombre d'ARNs régulateurs différentiellement exprimés. Une très large majorité de ces ARNnc était composée de lncRNA. Bien que ces ARNnc soient encore globalement très mal caractérisés, tout un champ de la littérature est en train de se développer sur leur rôle dans le contrôle de l'immunité inné et adaptative dans la réponse à l'infection par des bactéries pathogènes ⁶⁹⁹. Une analyse individuelle approfondie devra être réalisée afin d'identifier les lncRNA pouvant être impliqués dans la réponse à l'infection par *C. difficile*. En plus des lncRNA, une vingtaine de miRNA ont été identifiés. Beaucoup de ces miRNA n'ont jamais été décrits dans une publication. Cependant, un peu moins d'une dizaine de ces miRNA ont déjà été décrits et impliqués dans la réponse immunitaire et pourront faire l'objet d'analyses plus approfondies. Deux étaient notamment significativement induits dans les trois analyses différentielles et devront être étudiés plus attentivement.

Ce projet a permis la mise au point d'un protocole dual RNA-seq chez *C. difficile* qui sera valorisée très prochainement dans une publication (en cours de rédaction) et qui pourra être appliqué à d'autres projets de dual RNA-seq au cours de l'infection à *C. difficile*. Les ARNnc étant très conservés entre les différentes souches de *C. difficile*, les données obtenues avec la souche CD630E pourront être comparées avec différentes souches dont la souche hypervirulente CDR20291 ou des souches non toxigènes. Des optimisations pourront également être apportées au protocole développé. Il pourra notamment être envisagé de tester des temps d'infection tardifs plus espacés, comme un temps 24h et 32h afin de pouvoir suivre la dynamique d'expression des gènes tout au long du processus infectieux.

II. Rôle de la protéine Hfq dans la virulence

Si le rôle de la protéine Hfq dans la virulence des pathogènes à Gram négatif est bien caractérisé, il reste cependant encore largement débattu chez les pathogènes à Gram positif pour lesquelles la délétion du gène *hfq* ne semble avoir un impact sur la virulence que chez *L. monocytogenes*^{467,675}. De nombreux phénotypes sont également associés aux mutants Δhfq comme une sensibilité accrue à de nombreux stress, une augmentation de la formation de biofilms, une mobilité réduite et une croissance ralentie. Une étude antérieure d'un mutant de déplétion de Hfq avait mis en évidence le rôle pléiotrope de cette protéine chez *C. difficile*, ce qui constitue une caractéristique unique parmi les bactéries à Gram positif. Depuis, des mutants de délétion du gène *hfq* ont été obtenus dans différentes souches de *C. difficile* grâce à une méthode utilisant le système CRISPR-Cas endogène de la bactérie⁶⁰⁹, et publiés (publication dans laquelle je figure en 2^{ème} auteur). Nous avons réalisé l'analyse phénotypique du mutant Δhfq obtenu pour comparer l'effet de la délétion du gène *hfq* avec l'effet de la déplétion en Hfq. Dans un premier temps, divers phénotypes associés à la déplétion de cette protéine chez *C. difficile* lors d'une étude antérieure ont été testés⁶¹⁰. Parmi les phénotypes testés, seule l'augmentation de la sporulation chez le mutant Δhfq a pu être confirmée, l'altération de sporulation étant le phénotype le plus marqué de la souche déplétée pour Hfq qui avait également été confirmé par l'analyse transcriptomique. Les autres phénotypes testés n'ont révélé que peu ou pas de différence significative entre le mutant Δhfq et la souche sauvage. L'effet sur la croissance de la souche déplétée en Hfq (avec l'utilisation de fortes concentrations en ATc pour maintenir l'expression de l'ARN antisens) pourrait en partie être à l'origine des phénotypes observés précédemment. Un éventuel effet non spécifique dû à la surexpression de l'ARNs ciblant l'ARNm de *hfq* n'est pas exclu. Cela pourrait donc expliquer les différences observées entre les phénotypes du mutant de déplétion et du mutant Δhfq . La souche mutante Δhfq complémentée, présente une réduction considérable de croissance, vraisemblablement liée à l'effet toxique dû à la surexpression du gène *hfq* sous contrôle d'un promoteur inductible pTet. Ce phénomène avait déjà été observé avec le mutant de déplétion pour lequel une concentration de 500 ng/mL d'ATc avait été utilisée pour l'induction du gène *hfq*. Bien qu'une concentration plus faible d'ATc (100 ng/mL) ait été utilisée dans notre travail, cela ne semble pas être suffisant

pour éviter cet effet toxique. Effectivement, des essais récents ont montré que seules une concentration d'ATc de 10 ng/mL permettait d'obtenir une croissance normale associée à une induction suffisante du gène *hfq* sous le contrôle du pTet. Parmi les stress testés, une résistance modérée au stress osmotique (400 mM de NaCl) a été observée chez le mutant par rapport à la souche sauvage. Cette caractéristique particulière semble s'écarter des observations généralement faites chez les mutants Δhfq des bactéries à Gram négatif mais également de ce qui avait été décrit chez *L. monocytogenes*^{467,675}. Les tests de sensibilité à une forte osmolarité chez les mutants Δhfq décrits dans la littérature sont souvent réalisés en milieu liquide. Cette expérience devra donc être réalisée en milieu liquide afin de valider le phénotype observé dans des conditions de culture différentes. Nous avons également analysé l'expression de quelques gènes associés à la virulence dont les gènes des toxines *tcdA* et *tcdB*, le gène de l'adhésine FbpA, le gène de la piline majeure du *pilus* de type IV *pilA*, et les gènes de la protéase CD630_28300 et de l'adhésine CD630_28310. Seul le gène *CD630_28300* présentait une expression différente, plus élevée, chez le mutant par rapport à la souche sauvage et la souche mutante complémentée (**Figure S15**). Le gène *CD630_28300* code pour une zinc-métalloprotéase appelée Zmp1 qui est capable de cliver plusieurs protéines de l'hôte (comme la protéine de choc thermique HSP90 β , les anticorps IgA2, le fibrinogène ou la fibronectine), ainsi que le domaine de liaison de l'adhésine CD630_28310 au collagène, ce qui entraîne une séparation entre l'adhésine et le collagène^{146,147}. Ce résultat suggère donc que le mutant pourrait présenter une diminution de sa capacité à adhérer à la paroi intestinale de l'hôte.

Afin de tester l'impact de la délétion du gène *hfq* sur les capacités de colonisation du tube digestif et la virulence de *C. difficile* nous avons réalisé des essais *in vivo* sur modèle murin avec le mutant Δhfq et la souche sauvage. Les souris C57Bl/6 ont été traitées avec un cocktail d'antibiotiques permettant d'induire une dysbiose favorisant l'implantation de *C. difficile*⁶⁵⁴. Deux groupes de 8 souris chacun ont été infectées avec, d'une part la souche sauvage CD630E et, d'autre part, la souche CD630E Δhfq et leur évolution clinique ainsi que la charge bactérienne (formes végétatives) dans les fèces ont été suivies. Aucune des souris des 2 groupes n'a présenté de symptôme d'ICD au cours de cette expérience. A J1 et J6 un

nombre plus important (4 souris à J1 ; 6 souris à J6) de souris infectées par la souche mutante présentaient une valeur de numération des cellules végétatives inférieure au seuil de détection par rapport aux souris infectées par la souche sauvage. Ces résultats traduisent un défaut de persistance des formes végétatives de la souche mutante par rapport à la souche sauvage, éliminée moins rapidement. Dans la mesure où plusieurs souris des deux groupes présentaient une numération sous le seuil de détection, cette expérience devra être reproduite afin de confirmer le phénotype observé. La colonisation du colon est une étape majeure du cycle infectieux de *C. difficile* durant laquelle plusieurs facteurs bactériens sont impliqués comme des protéines de surface pouvant adhérer aux entérocytes. Nous avons déjà montré *in vitro* qu'une surexpression du gène *CD630_28300* était observée dans le mutant Δhfq , pouvant aboutir à une diminution d'adhésion de la bactérie à la muqueuse épithéliale. Mais il serait intéressant d'étudier l'expression de gènes codant d'autres facteurs impliqués dans la colonisation, notamment des facteurs de résistance aux défenses de l'hôte. Toutefois, ce résultat doit être nuancé car l'injection IP de clindamycine au 27^e jour s'est traduit par une détection similaire de formes végétatives dans les 2 groupes de souris. Dans la mesure où le mutant Δhfq présente une augmentation de sa capacité à sporuler *in vitro*, par rapport à la souche sauvage, l'élimination plus rapide des formes végétatives du mutant pourrait être contre-balançée par l'augmentation de la quantité de spores produites. Pour évaluer ce phénotype de sporulation *in vivo*, la quantité de spores de *C. difficile* présentes dans les fèces des 2 groupes de souris (conservés à -80°C) sera évaluée très prochainement. À plus long terme, il serait intéressant d'étudier les ARNnc associés à Hfq (identifiés lors d'une précédente expérience RIP-seq) et ciblant les gènes impliqués dans la colonisation. En revanche, il est difficile de conclure sur l'implication de l'Hfq dans la virulence. Dans la littérature, le hamster est plus souvent décrit comme modèle de virulence pour les infections à *C. difficile*. Nous avons cependant décidé d'utiliser un modèle murin plus simple à manipuler et beaucoup moins sensible que le hamster. En effet, ces animaux développent des infections fulgurantes et meurent très rapidement. Une nouvelle expérimentation animale avec un modèle hamster pourrait toutefois être envisagée pour évaluer l'impact de la délétion du gène *hfq* sur la virulence de *C. difficile*.

III. Caractérisation du *riboswitch IGR-1461*

Nous avons élargi les analyses des ARNnc à d'autres souches de *C. difficile*, en particulier la souche hypervirulente de PCR-ribotype 027 R20291. L'émergence puis la dissémination de ces souches épidémiques et hypervirulentes depuis 2002 ont été associées à l'augmentation de la proportion de formes sévères d'infections à *C. difficile* de manière préoccupante en Amérique du Nord puis en Europe. Néanmoins, les bases moléculaires de l'hypervirulence de ces souches restent largement inconnues. Les ARNnc peuvent contribuer à l'adaptation des souches hypervirulentes aux conditions chez l'hôte. De manière intéressante, des données de transcriptomique comparative obtenues par le laboratoire (non publiées) entre la souche CD630 et la souche hypervirulente CDR20291 a mis en évidence plusieurs ARNnc spécifiquement exprimés dans la souche CDR20291. L'un de ces ARNnc présentant un signal fort en RNA-seq et Northern-blot et sans équivalent Rfam, a été sélectionné pour une caractérisation approfondie. Cet ARNnc (*IGR-1461*) est localisé dans le génome du prophage phi027 qui est commun et spécifique à la quasi-totalité des souches de PCR-ribotype 027 chez *C. difficile* et absent de la souche CD630. Une analyse par BLAST dans les bases de données a mis en évidence la conservation de cet ARNnc chez différentes souches de ribotype 027 et chez un autre phage de *C. difficile*, le phage phiC2. De plus, chez les souches de ribotype 027 et chez le phage phiC2, les gènes adjacents sont également bien conservés. La prédiction *in silico* des cibles de ce petit ARN a révélé plusieurs cibles potentielles de fonctions diverses dont des gènes appartenant au prophage phi027, suggérant donc une régulation croisée entre le prophage phi027 et le génome de la bactérie, ainsi qu'une auto-régulation de l'expression des gènes du prophage. Aucune région de fixation commune n'a pu être prédite entre les différentes cibles ARNm et l'*IGR-1461*, mais il semblerait qu'une région unique structurée en tige-boucle soit prédite pour l'interaction de cet ARNnc avec ses cibles. L'interaction avec les cibles prédites pourra être développée ultérieurement lors des prochaines étapes de caractérisation de cet ARNnc par des méthodes de co-immunoprécipitation comme la méthode MAPS (*MS2-affinity purification*) couplée à du RNA-seq, dont le protocole chez *C. difficile* est en cours de développement⁷⁰⁰. L'étude de la structure génétique de l'ARNnc et du gène localisé en aval (*CD_1462*) nous a permis de supposer une régulation de ce gène par l'*IGR-1461* via un mécanisme en *cis* de

type *riboswitch*. En effet, une seule région promotrice a été prédite en amont de la séquence de l'*IGR-1461* et aucune pour le gène *CD_1462* situé en aval. De plus, une expérience 5'-RACE a permis de localiser le +1 de transcription du gène *CD_1462* dans la séquence de l'*IGR-1461* au même niveau que celui de l'ARNnc IGR-1461. Une prédiction de la structure secondaire de ce *riboswitch* a mis en évidence trois tiges-boucles dont deux aptamères conservés de types *abiF* et un terminateur transcriptionnel Rho indépendant. Le gène *CD_1462* est annoté comme codant pour une protéine appartenant à la famille des protéines *Abi* impliquées dans l'abortion de l'infection phagique par des mécanismes d'induction de la mort (ou du ralentissement physiologique) de la cellule infectée. L'analyse par BLAST de la séquence protéique a suggéré une homologie avec une protéine *AbiF-like* retrouvée chez deux autres souches de *C. difficile*. De plus, la séquence du gène de cette protéine est identique à 99% avec celle codant pour une protéine *AbiF* retrouvée chez le phage tempéré *phiC2* de *C. difficile*⁶⁸³.

Afin de pouvoir caractériser le mécanisme de régulation du *riboswitch* sur le gène *CD_1462*, un mutant de délétion de l'*IGR-1461* a été construit. Une expression du gène *CD_1462* 12 fois supérieure chez le mutant Δ *IGR-1461* par rapport à la souche sauvage confirme la régulation transcriptionnelle négative de l'expression du gène *CD_1462* par l'*IGR-1461*. La construction d'un système rapporteur avec le *riboswitch* sous le contrôle d'un promoteur constitutif fort cloné en amont du gène *phoZ* a également permis de confirmer une répression transcriptionnelle exercée par ce *riboswitch* sur le gène localisé en aval. De plus, une construction de fusion traductionnelle avec le *riboswitch* et le gène *phoZ* a mis en évidence une diminution plus marquée de la production de phosphatase alcaline suggérant donc un mécanisme de régulation à la fois au niveau transcriptionnel et traductionnel qui permettrait une répression totale du gène *CD_1462* hors de ses conditions d'induction. Cette hypothèse devra cependant être confirmée par une construction *HA-tag* pour visualiser la protéine CD1462 et déterminer sa localisation cellulaire chez le mutant Δ *IGR-1461*. D'autre part, d'après les résultats des expériences de phosphatase alcaline, le *riboswitch* semble fonctionner dans plusieurs contextes génétiques (*CD630*, *CDR20291* et *CDR20291* Δ *phi027*) excluant donc une production spécifique du ligand par la souche *CDR20291*, ainsi qu'une

production du ligand par le phage phi027. Bien qu'un niveau plus élevé de phosphatase alcaline soit observé dans la souche CD630E par rapport aux autres souches, cela ne permet pas de conclure à un effet directement lié au mécanisme du *riboswitch*. Cette augmentation pourrait en effet être due à une lyse plus importante de la souche CD630 par rapport aux autres souches testées, une normalisation des valeurs d'UE sur la quantité de protéines totales pour chaque échantillon doit être réalisée avant de conclure. Afin de pouvoir identifier l'origine et la nature du ligand, des tests similaires ont été réalisés à partir de culture dans milieu contrôlé (CDMM) permettant ainsi de tester l'hypothèse de la présence d'un ligand métabolite dans un milieu de culture riche (**Figure S16**). Ces tests n'ont pas révélé de différences de régulation du gène de la phosphatase alcaline par le *riboswitch* par rapport à des conditions de culture en milieu riche, excluant donc la présence du ligand dans le milieu de culture. Cependant, bien que dans la majorité des cas, la fixation du ligand à l'aptamère induise la formation d'une structure de terminateur dans le domaine régulateur, le cas de figure inverse est possible où la fixation du ligand à l'aptamère permettrait de lever l'inhibition de la transcription. Etant donné la nature hypothétique de la protéine CD1462 impliquée potentiellement dans l'infection phagique abortive, la piste d'un ligand d'origine phagique devra être étudiée en premier.

Afin de confirmer la nature Abi de la protéine CD1462, des tests de croissance d'une souche surexprimant le gène *CD_1462* sous le contrôle d'un promoteur inductible ont été réalisés. La surexpression de cette protéine se traduit par un ralentissement de la croissance bactérienne lors de sa surexpression. Cette caractéristique va donc dans le sens de l'hypothèse de la nature Abi de la protéine CD1462. On observe cependant une reprise de la croissance bactérienne après quelques heures de culture qui pourrait être due à une dégradation progressive de l'ATc dans le milieu de culture réduisant progressivement l'induction de l'expression du gène *CD_1462*. En effet, bien que dans la plupart des cas, l'activation des systèmes Abi aboutissent à la mort de la cellule, leur effet peut être réversible et peut dans certains cas permettre à la bactérie d'activer d'autres systèmes de défense contre les phages^{333,334}. D'autre part, une dynamique de croissance du mutant *ΔIGR-1461* similaire à celle de la souche sauvage a été observée et ce malgré un niveau de transcrit plus

important du gène *CD_1462* chez ce mutant. Il est possible qu'une structure secondaire masquant le RBS du gène *CD_1462* puisse empêcher la production de la protéine CD1462 chez le mutant $\Delta IGR-1461$ comme suggéré lors des expériences de PA. Cette hypothèse pourra être vérifiée en contrôlant le niveau de la protéine CD1462 par une construction *HA-tag* chez le mutant $\Delta IGR-1461$ et chez un mutant ΔCD_1462 . Dans la mesure où la construction du mutant ΔCD_1462 avec le système de mutagenèse toxine-antitoxine n'a pas fonctionné, l'utilisation du système CRISPR-Cas endogène utilisé pour la construction du mutant Δhfq pourra être envisagée⁶⁰⁹. Bien que la nature Abi de la protéine CD1462 reste à confirmer, la présence d'une protéine Abi dans le génome d'un prophage est surprenante. En effet, ces systèmes de défense sont la plupart du temps portés sur des plasmides conférant à la bactérie la capacité de résister à l'infection phagique⁶⁸⁷. L'avantage conféré par la présence d'un tel système sur la séquence d'un prophage est encore mal compris mais une des raisons pourrait être d'empêcher la surinfection de la bactérie par d'autres phages qui seraient sensibles au système Abi codé par le prophage. Des travaux récents ont notamment montré que, contrairement à la souche sauvage, le phage phi027 est capable d'infecter un mutant $\Delta phi027$. Bien que le prophage phi027 comporte dans son génome un système de défense CRISPR, la présence du système de défense Abi pourrait être également liée à la résistance des souches portant le prophage dans leur génome au phage phi027. La surexpression du gène *CD_1462* sur un plasmide dans une souche mutante $\Delta phi027$ permettrait de pouvoir déterminer le rôle de la protéine CD1462 lors de l'infection. De plus, l'expression du système complet (*riboswitch+CD_1462*) sur un plasmide chez le mutant $\Delta phi027$ permettrait d'étudier le mécanisme d'activation de ce système lors de l'infection phagique. Cependant, une telle construction s'est avérée complexe à obtenir. En effet, il semblerait que le système complet sous le contrôle de son promoteur naturel soit toxique chez *E. coli* ce qui entraîne des remaniements de séquence du plasmide empêchant d'obtenir la construction complète. Bien qu'il soit possible que des fuites du promoteur P1461 entraîne une faible production de la protéine CD1462 toxique chez *E. coli*, il n'est pas invraisemblable que le ligand ne soit produit que par des souches de *C. difficile* et que le gène *CD_1462* soit alors exprimé dans un système hétérologue. Des expériences d'expression du système *riboswitch-CD_1462* chez *B. subtilis* sont en train d'être réalisées par nos collaborateurs de

l'Université de Sherbrooke afin de tester cette hypothèse. Bien que le système complet sous le contrôle de son promoteur naturel n'ait pour l'instant pas pu être obtenu, un système *riboswitch-CD_1462* sous le contrôle d'un promoteur inductible pTet déjà construit permettra tout de même d'étudier le mécanisme d'activation du gène *CD_1462* lors de l'infection phagique. Le mécanisme d'activation du gène *CD_1462* pourra également être étudié grâce à un système rapporteur *phoZ* de manière à comparer le niveau de production de PA à différents temps d'une infection phagique. D'autre part, la comparaison des systèmes rapporteurs du *riboswitch* fusionnés avec le gène *phoZ* avec et sans le RBS du gène *CD_1462* pourra permettre d'étudier le mécanisme de régulation traductionnelle révélé par les expériences de PA. On pourrait en effet envisager un mécanisme de régulation traductionnelle similaire à celui du système AbiD1 décrit chez *L. lactis* avec la fixation d'une protéine phagique au RBS permettant de lever l'inhibition traductionnelle⁶⁹².

Ce projet a permis l'identification d'un système Abi putatif et de son *riboswitch* associé ce qui constitue une caractéristique unique pour *C. difficile* chez qui un tel système n'avait jamais été identifié auparavant. Des études approfondies de ce système devront être réalisées afin d'étudier le rôle de la protéine Abi putative dans la réponse au stress et la défense contre l'infection phagique, par comparaison avec les travaux réalisés chez *L. lactis*^{332,692}. Son mécanisme d'action devra être également caractérisé *via* l'identification de ses partenaires grâce à des expériences de *Pull-down* et sa localisation cellulaire.

IV. Conclusion

Les travaux effectués dans le cadre de cette thèse ont permis d'apporter de nouveaux éléments de connaissance sur les mécanismes de régulation par ARNnc chez *C. difficile*. La mise en place d'un protocole de dual RNA-seq dans un modèle infectieux de *C. difficile* a permis d'obtenir des nombreuses données préliminaires sur la dynamique d'expression des ARNnc dont l'analyse plus détaillée permettra certainement d'identifier de nouveaux éléments impliqués dans le contrôle des processus de virulence de *C. difficile* et dans la défense de l'hôte. La caractérisation d'un nouveau mutant de délétion pour le gène de la protéine ARN chaperon Hfq a révélé certains phénotypes intéressants du mutant, notamment une augmentation de la sporulation, une résistance modérée à une forte osmolarité,

et une diminution de fitness global du mutant en modèle souris. Enfin, l'identification d'un nouveau riboswitch présent spécifiquement chez les souches hypervirulentes de PCR-ribotype 027 a permis de mettre en évidence un mécanisme de régulation original à deux niveaux (transcriptionnel et traductionnel) d'une protéine Abi-like associée au système de défense par mort altruiste lors de l'infection phagique, ce qui constitue une caractéristique unique puisqu'un tel système n'avait encore jamais été caractérisé chez *C. difficile*. Ces résultats apportent donc des éléments nouveaux sur l'étendue des mécanismes de régulation par ARNnc dans la physiopathologie de *C. difficile* et permettront de développer de nouveaux projets de recherche afin d'approfondir la caractérisation de ces mécanismes.

Annexes

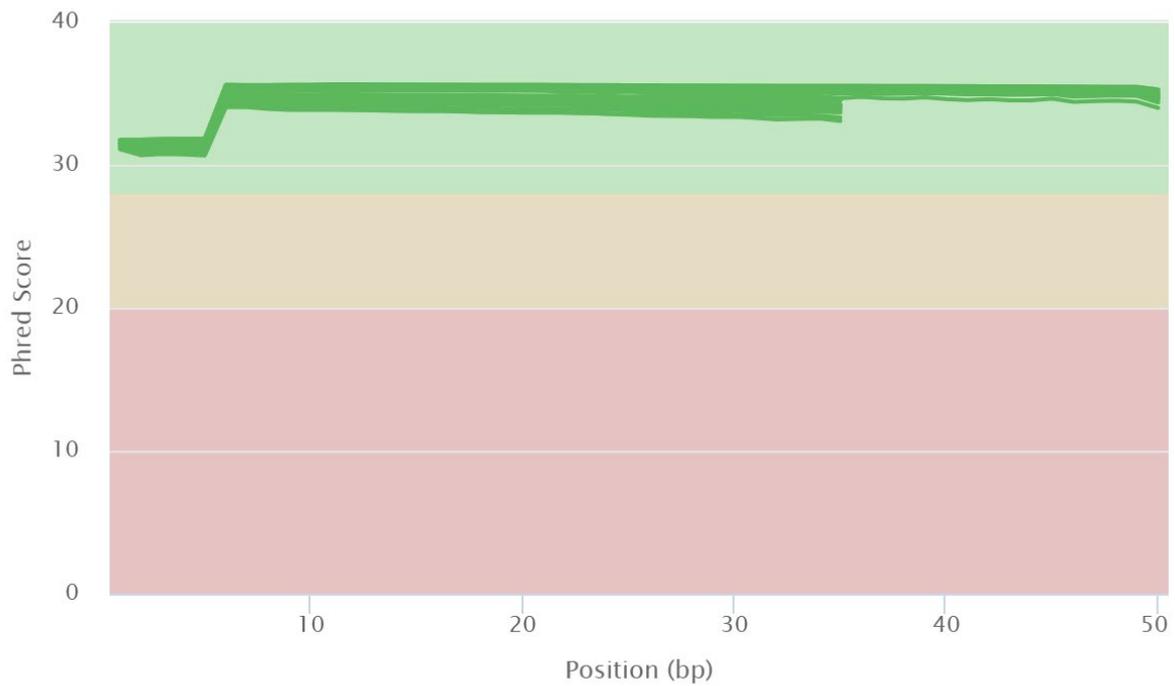


Figure S1 – Score de qualité Phred des *reads* (PE 50-35 nt) obtenus après séquençage.

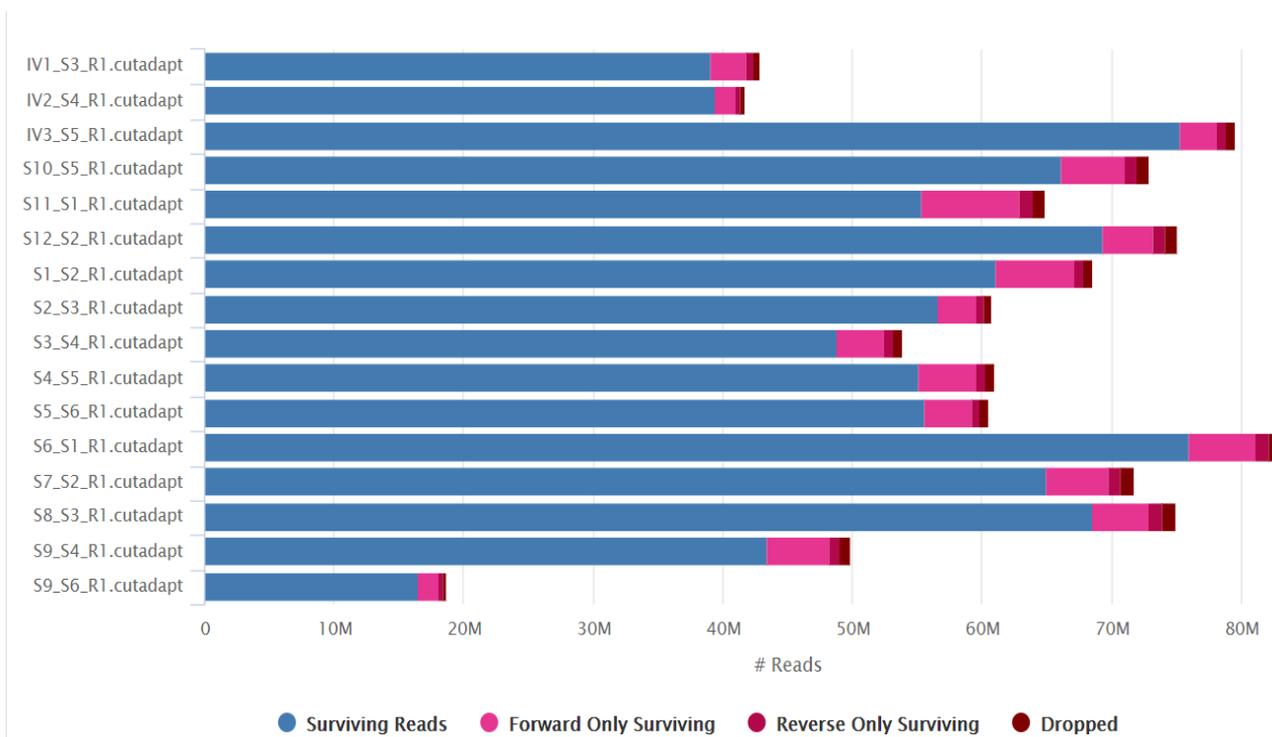


Figure S2 – Nombre de *reads* conservés (bleu) après nettoyage par Trimmomatic.

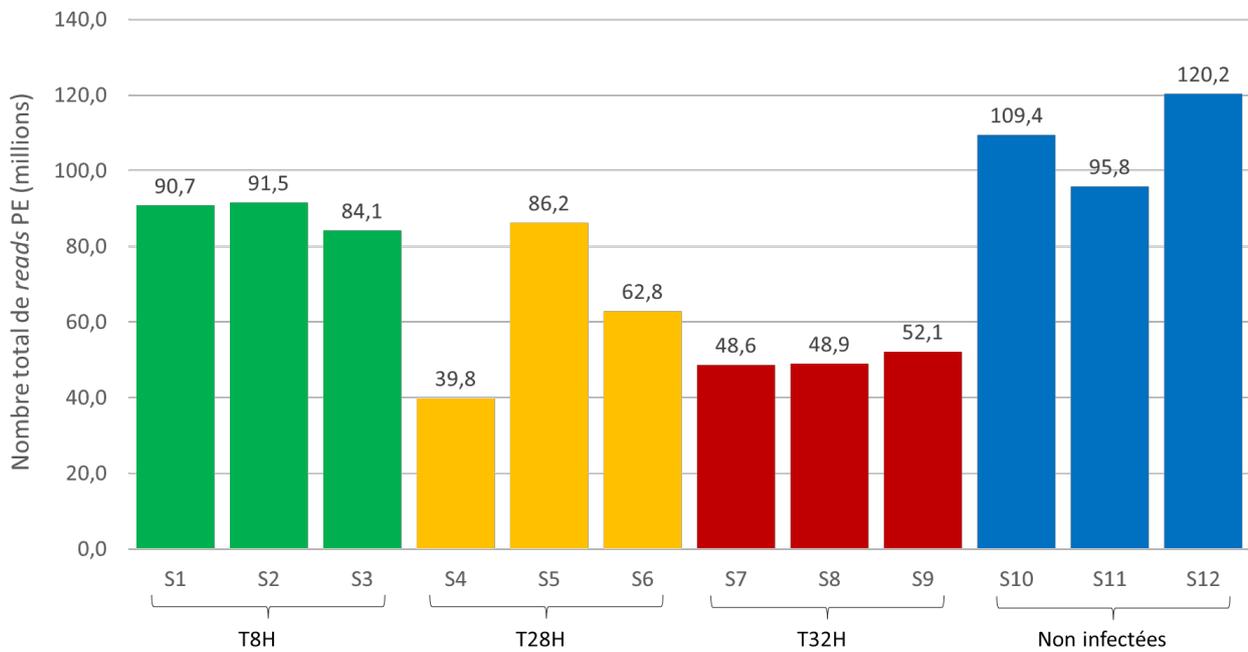


Figure S3 – Nombre de *reads* ne s'alignant ni sur le génome de *C. difficile* ni sur celui de la souris.

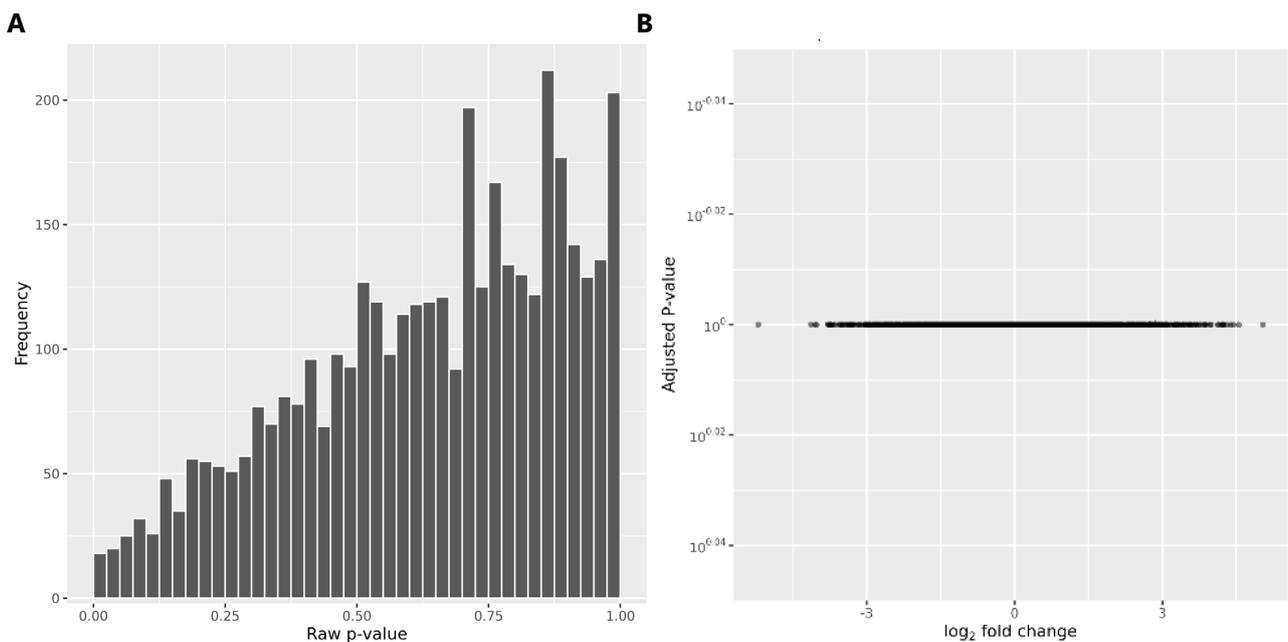


Figure S4 – Analyse différentielle de l'expression des gènes entre les groupes de souris infectées à 28h (MI28h) et à 32h (MI32h). (A) Les distributions des *p-values* brutes sont calculées par le test statistique pour la ou les comparaison(s) effectuée(s). Cette distribution devrait être un mélange d'une distribution uniforme sur [0,1] et d'un pic autour de 0 correspondant aux caractéristiques différentiellement exprimées. (B) *Volcano plot* de la comparaison effectuée entre les groupes MI28h et MI32h.

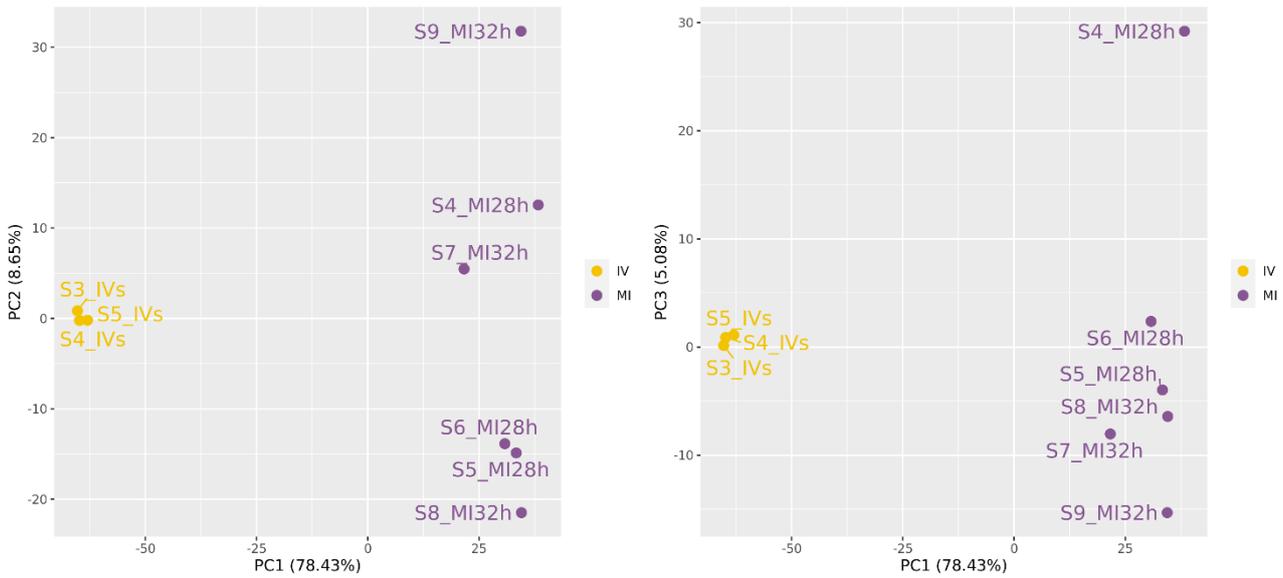


Figure S5 – Analyse en Composante Principale de l’analyse d’expression différentielle des gènes de *C. difficile* entre les groupes IV (*in vitro*) et MI (*mouse infected*). Sur cette figure, l’axe PC1 doit permettre de séparer les échantillons issus des différentes conditions comparées, signifiant que la variabilité biologique est la principale source de variation dans les données analysées.

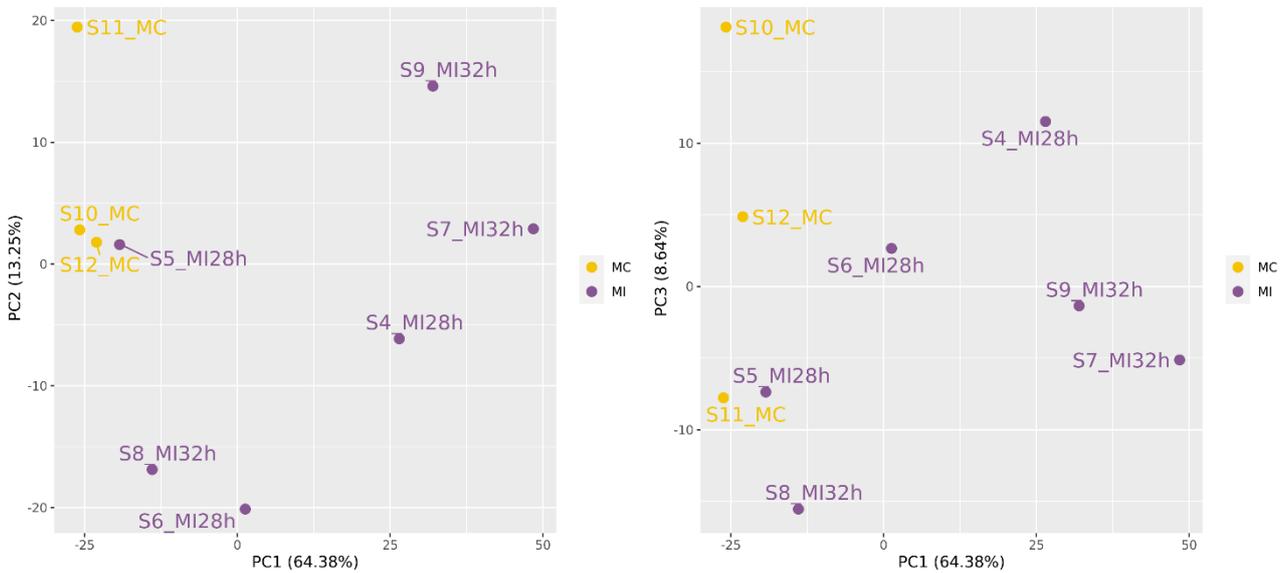


Figure S6 – Analyse en Composante Principale de l’analyse d’expression différentielle des gènes murins entre les groupes IV (*in vitro*) et MI (*mouse infected*).

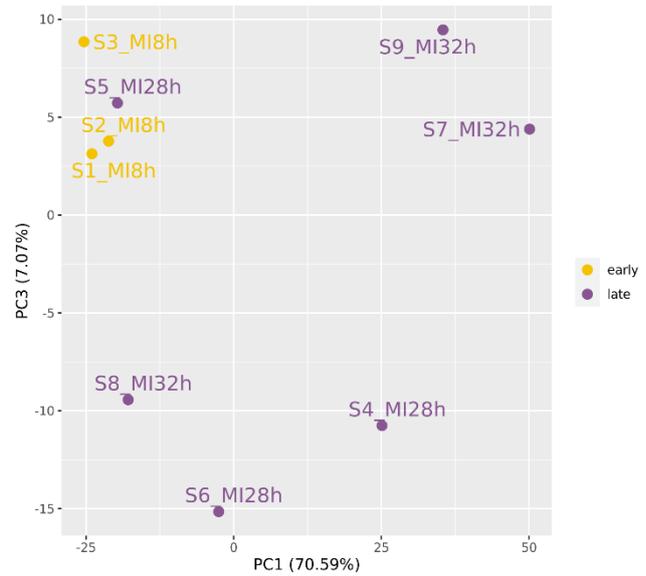
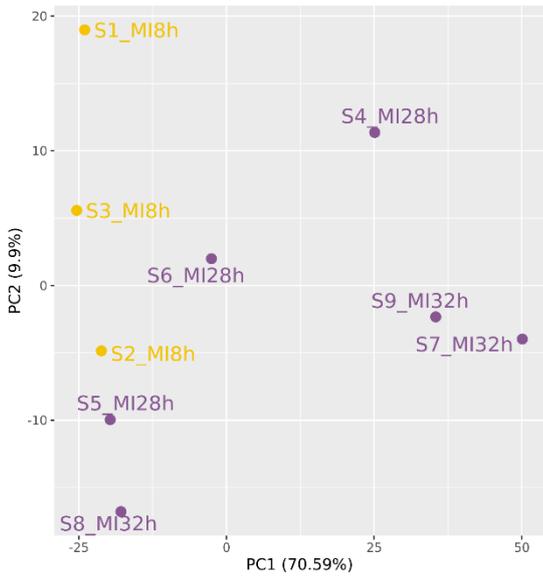


Figure S7 – Analyse en Composante Principale de l’analyse d’expression différentielle des gènes murins entre les groupes IV et MI.

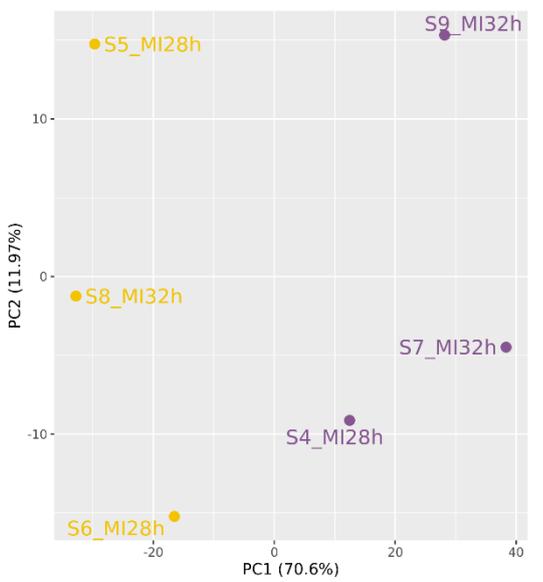


Figure S8 – Analyse en Composante Principale de l’analyse d’expression différentielle des gènes murins entre les groupes IV et MI.

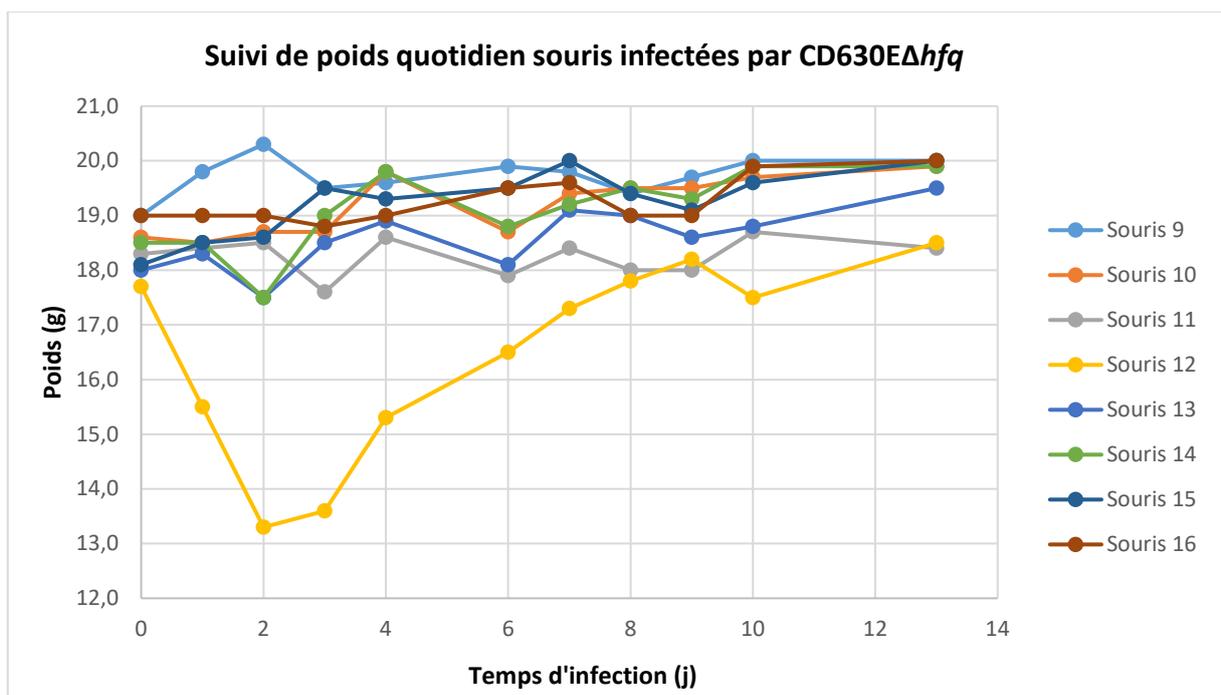
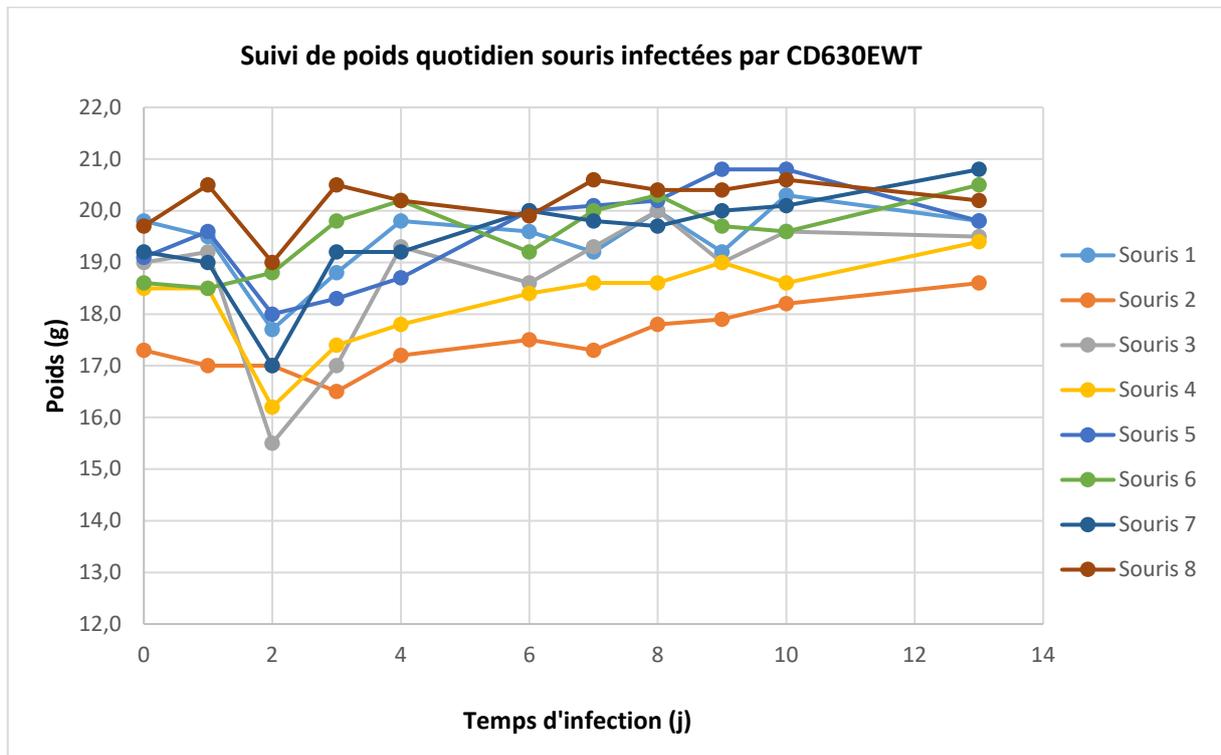


Figure S9 – Variation du poids entre les souris infectées (A) par la souche CDC630E WT et (B) Δhfq .

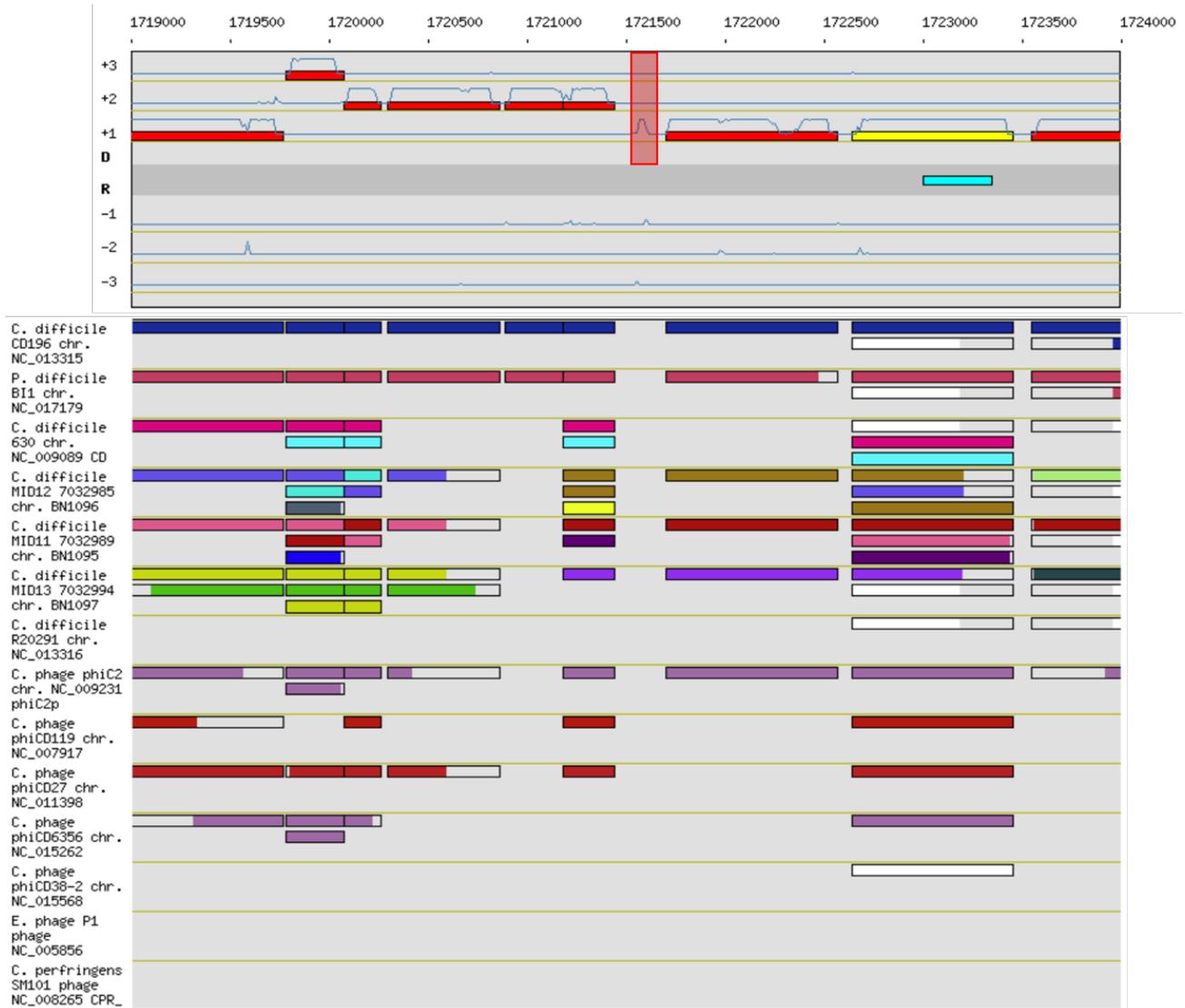


Figure S10 – Étude de la conservation de la région dans laquelle se trouve l'IGR-1461. Les régions parfaitement conservées sont indiquées en couleur tandis que les régions partiellement conservées sont indiquées en blanc. La région de la souche CDR20291 est indiquée dans l'encadré en haut de la figure. L'IGR-1461 est repéré par un encart rouge. L'analyse comparative a été faite avec l'outil MaGe du Génoscope (<http://www.genoscope.cns.fr/agc/mage>)

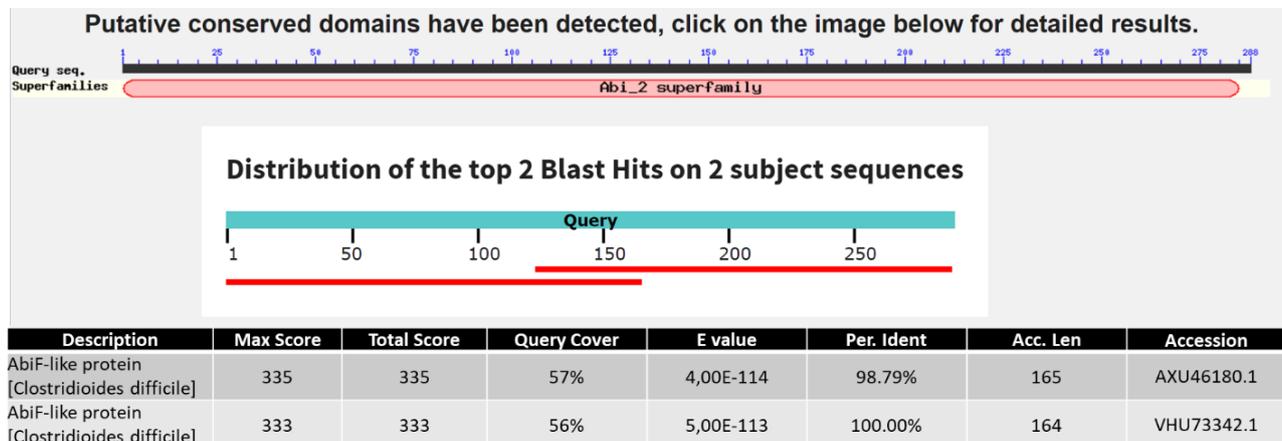


Figure S11 – Étude de la conservation de la protéine CD1462 par analyse BLAST.

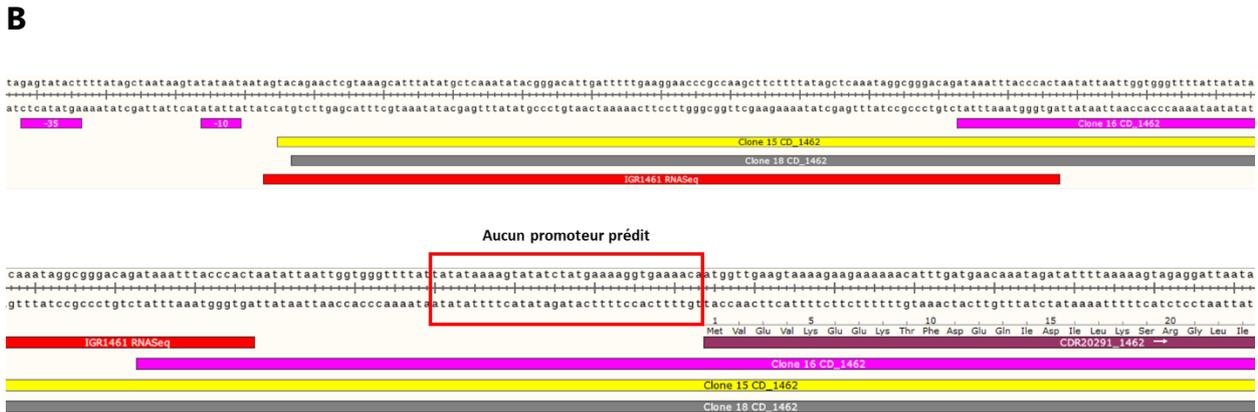


Figure S12 – Séquences déduites de l’expérience 5’ RACE pour (A) l’*IGR-1461* et (B) le gène *CD_1462*.

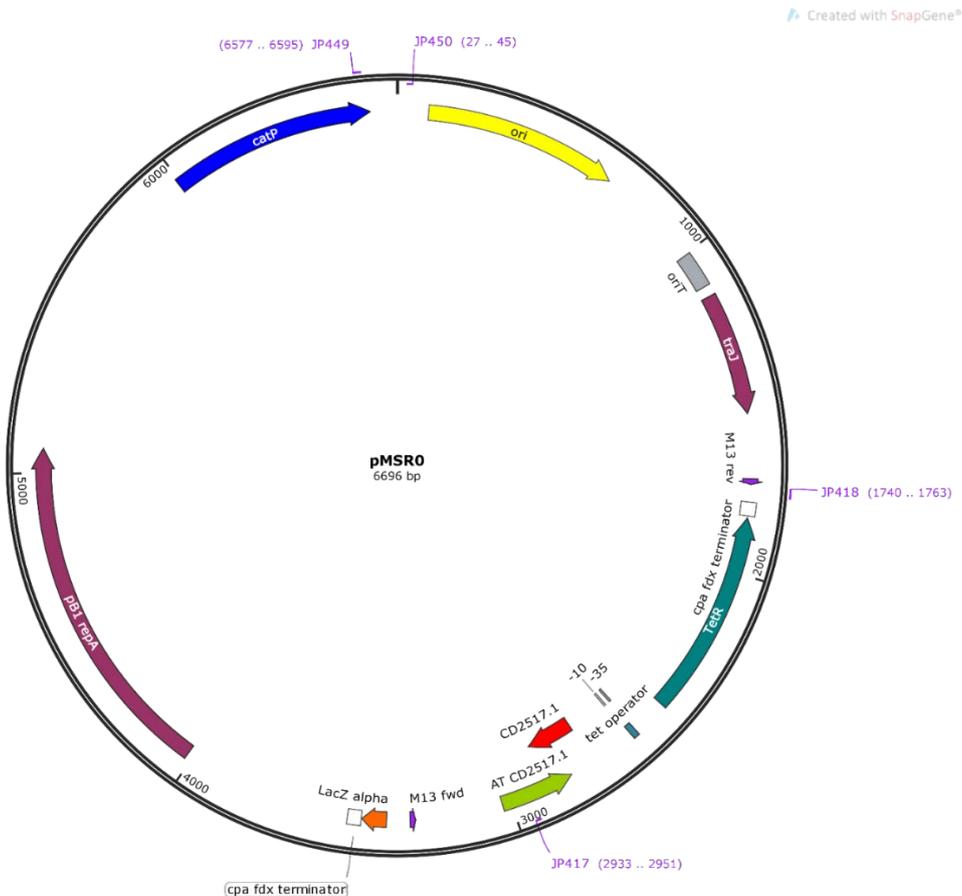


Figure S13 – Carte du plasmide de mutagenèse **pMSR0**. Les gènes de la toxine *CD2517.1* et de son antitoxine sont indiqués en rouge et vert respectivement. Le site de clonage se trouve après la *tet operator*.

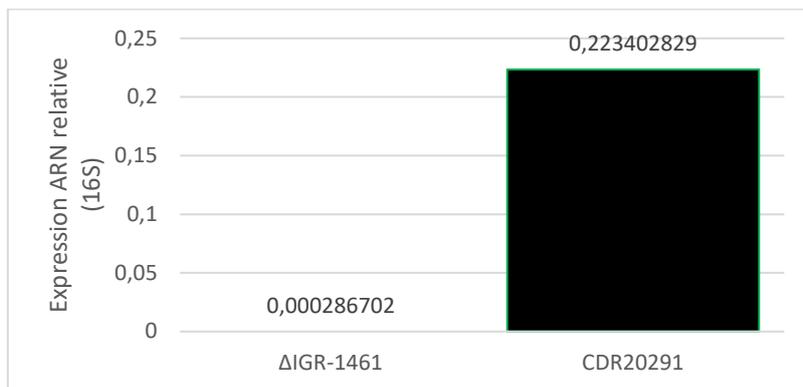


Figure S14 – Confirmation du mutant $\Delta IGR-1461$ par qPCR. Sont représentées les expressions relatives de l'*IGR-1461* chez le mutant $\Delta IGR-1461$ ($n = 3$) et la souche WT CDR20291.

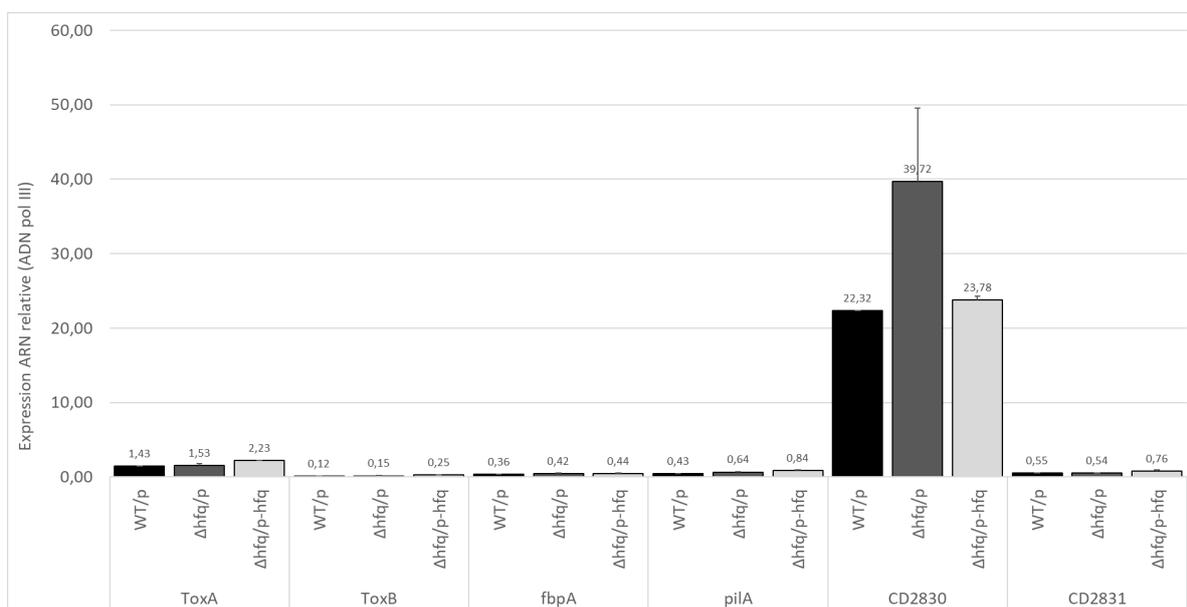


Figure S15 – Analyse de l'expression de gènes associés à la virulence chez les souches CD630E WT, Δhfq et Δhfq complétée. ($n = 3$)

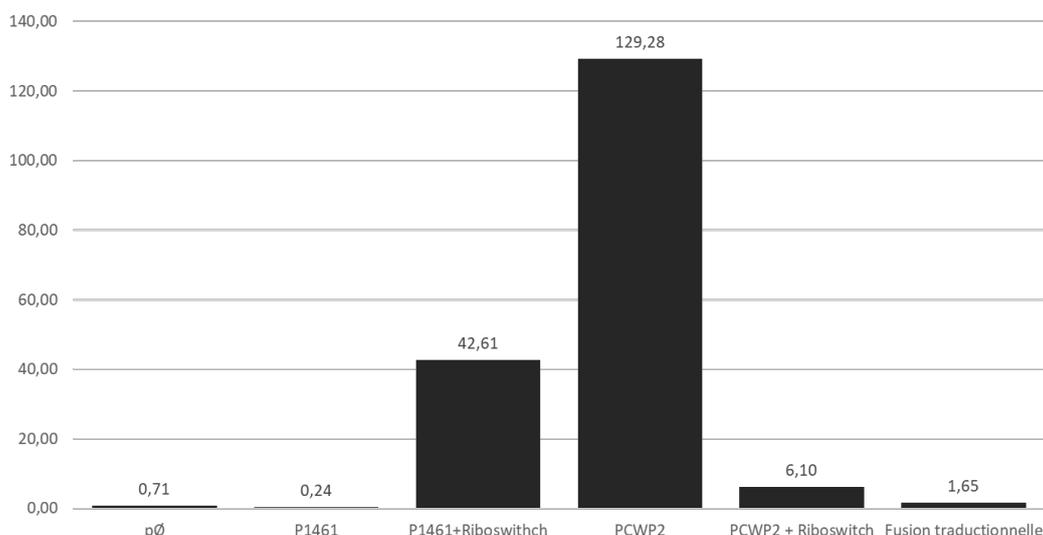


Figure S16 – Expérience de phosphatase alcaline réalisée dans la souche CD630E cultivée en milieu minimum (CDMM).

Tableau S1 – Dénombrement des cellules végétatives de *C. difficile* pour chaque souris infectée.

Souris	Temps	Dilutions étalées (100 µL)						UFC/g	Moyenne
		1,00E-02	1,00E-02	1,00E-03	1,00E-03	1,00E-04	1,00E-04		
1	8 h	124	120	13	20	3	0	1,46E+08	1,63E+08
2	8h	186	200	27	23	2	3	2,31E+08	
3	8h	85	95	10	9	1	2	1,12E+08	
4	28 h	201	234	26	18	0	1	1,63E+08	1,41E+08
5	28h	139	135	8	20	2	1	1,42E+08	
6	28h	80	75	11	14	2	1	1,18E+08	
7	32 h	6	11	0	1	0	0	4,50E+06	9,67E+07
8	32 h	180	213	43	25	3	1	2,46E+08	
9	32 h	54	56	6	7	0	0	4,00E+07	

Tableau S2 – Dosage Nanodrop des ARN totaux procaryotes et eucaryotes obtenus.

#	Infection	T(h)	Souris	ARN	Concentration ARN (ng/µL)	A260	A280	260/280	260/230
1	OUI	8	S1	EUCARYOTES	844,1	21,101	10,59	1,99	1,32
2		8	S2		287,6	7,189	3,665	1,96	1,07
3		8	S3		740,8	18,521	9,499	1,95	1,47
4		28	S4		1127,4	28,185	14,306	1,97	1,59
5		28	S5		1056,8	26,42	13,237	2	1,8
6		28	S6		1499,6	37,49	18,921	1,98	1,54
7		32	S7		1014	25,349	12,599	2,01	1,13
8		32	S8		1263,3	31,581	15,895	1,99	1,67
9		32	S9		1754,2	43,855	22,057	1,99	1,58
10	NON	8	S10		626,2	15,654	7,788	2,01	1,8
11		8	S11		667,3	16,683	8,476	1,97	1,63
12		8	S12		1167,5	29,186	14,738	1,98	1,32
19	OUI	8	S1	PROCARYOTES	118,2	2,956	1,625	1,82	0,29
20		8	S2		153,6	3,84	2,176	1,76	0,24
21		8	S3		80,4	2,01	1,14	1,76	0,21
22		28	S4		153,4	3,835	1,955	1,96	0,25
23		28	S5		118,6	2,964	1,607	1,84	0,28
24		28	S6		180,7	4,518	2,291	1,97	0,37
25		32	S7		99,1	2,479	1,319	1,88	0,36
26		32	S8		182,3	4,557	2,659	1,71	0,62
27		32	S9		62,8	1,57	0,919	1,71	0,19
28	NON	8	S10		191,6	2,758	1,438	1,92	0,54
29		8	S11		114,7	2,869	1,505	1,91	0,64
30		8	S12		114,6	2,864	1,515	1,89	0,46

Tableau S3 – Prédiction des cibles ARNm de l'IFR-1461. Les cibles présentées ci-dessous ont été prédites par deux algorithmes différents (IntaRNA et RNAPredator). Les 15 meilleures cibles (énergie de liaison < -10 kJ/mol) prédites par les deux algorithmes ont été sélectionnées.

Target Gene	Annotation
CDR20291_0253	flagellar assembly protein (fliH)
CDR20291_0253	flagellar assembly protein CDR20291
CDR20291_0480	single peptidase I
CDR20291_0591	two-component response regulator
CDR20291_0937	PTS system; IId component CDR20291
CDR20291_1085	elongation factor P (efp)
CDR20291_1281	ribonuclease
CDR20291_1445	phiCD027 protein
CDR20291_1459	Hypothetical protein phage phiCD027
CDR20291_1460	Hypothetical protein phage phiCD027
CDR20291_1544	putative dipeptidase CDR20291
CDR20291_1832	two component sensor histidine kinase CDR20291
CDR20291_1841	propanediol/ethanolamine utilization protein
CDR20291_2461	ribulose-phosphate 3-epimerase
CDR20291_3084	aspartate-semialdehyde dehydrogenase (asd)

Tableau S4 – Séquence des amorces oligonucléotides utilisés lors de l'expérience de 5'-RACE.

Référence	Séquences	Gène amplifié
Oligos pour RT		
OS740	5'-GGGGATCCAGTGGGTAAATTTATCTGTCC-3'	<i>IGR_1461</i>
OS736	5'-AGCGCATTGATTCTGACCT-3'	<i>CD_1462</i>
Oligos utilisés pour la PCR		
OS468	5'-CGGGACATTGATTTTTGAAG-3'	<i>IGR_1461</i>
OS733	5'-CATATAAATGCTTTACGAGTTCTG-3'	
OS737	5'-GTTTTTCTTCTTTACTTCAACCA-3'	<i>CD_1462</i>
OS738	5'-GGGTTGCAACGGAAATAATGACT-3'	
Oligos utilisés pour la Nested-PCR		
OS733	5'-CATATAAATGCTTTACGAGTTCTG-3'	<i>IGR_1461</i>
OS734	5'-AGGAACCCGCCAAGCTTCTTT-3'	
OS737	5'-GTTTTTCTTCTTTACTTCAACCA-3'	<i>CD_1462</i>
OS735	5'-TTCAAACATGTTGCCAGAAGA-3'	

Tableau S5 – Séquences des amplicons de l’IGR-1461 obtenues lors de l’expérience de 5’ RACE. La séquence complète déduite est indiquée sous le tableau. Le terminateur Rho indépendant est souligné.

#	Séquences de l’IGR_1461
1	CAGAACTCGTAAAGCATTATATGCTCAAATATACGGGACATTGATTTTTGAAGGAACCCGCCAAGCTTCTTTTATAGCTC AAATAGGCGGGACAGATAA
2	ACAGAACTCGTAAAGCATTATATGCTCAAATATACGGGACATTGATTTTTGAAGGAACCCGCCAAGCTTCTTTTATAGCT CAAATAGGCGGGACA
3	ACAGAACTCGTAAAGCATTATATGCTCAAATATACGGGACATTGATTTTTGAAGGAACCCGCCAAGCTTCTTTTATAGCT CAAATAGG
4	ACAGAACTCGTAAAGCATTATATGCTCAAATATACGGGACATTGATTTTTGAAGGAACCCGCCAAGCTTCTTTTATAGCT CAAATAGGCGGGACAG
5	ACAGAACTCGTAAAGCATTATATGCTCAAATATACGGGACATTGATTTTTGAAGGAACCCGCCAAGCTTCTTTTATAGCT CAAATAGGCGGGACAG
6	CAGAACTCGTAAAGCATTATATGCTCAAATATACGGGACATTGATTTTTGAAGGAACCCGCCAAGCTTCTTTTATAGCTC AAATAGGCGGGACAGA
7	ACAGAACTCGTAAAGCATTATATGCTCAAATATACGGGACATTGATTTTTGAAGGAACCCGCCAAGCTTCTTTTATAGCT CAAATAGG
8	CAGAACTCGTAAAGCATTATATGCTCAAATATACGGGACATTGATTTTTGAAGGAACCCGCCAAGCTTCTTTTATAGCTC AAATAGGCGGGACAGA
9	TACAGAACTCGTAAAGCATTATATGCTCAAATATACGGGACATTGATTTTTGAAGGAACCCGCCAAGCTTCTTTTATAGC TCAAATAGGCGGGACAGATAAATTTACCCACTAA
10	ACTCGTAAAGCATTATATGCTCAAATATACGGGACATTGATTTTTGAAGGAACCCGCCAAGCTTCTTTTATAGCTCAAAT AGGCGGGACAGA
11	ACTCGTAAAGCATTATATGCTCAAATATACGGGACATTGATTTTTGAAGGAACCCGCCAAGCTTCTTTTATAGCTCAAAT AGGCGGGACAGA

Séquence complète de l’IGR-1461 déduite à partir des séquences ci-dessus :

TACAGAACTCGTAAAGCATTATATGCTCAAATATACGGGACATTGATTTTTGAAGGAACCCGCCAAGCTTCTTTTATA
GCTCAAATAGGCGGGACAGATAAATTTACCCACTAATATTAATTGGTGGG

Tableau S6 - Séquences des amplicons de du gène CD_1462 obtenues lors de l’expérience de 5’ RACE.

#	Séquences de CD_1462
1	TACAGAACTCGTAAAGCATTATATGCTCAAATATACGGGACATTGATTTTTGAAGGAACCCGCCAAGCTTCTTTTATAGCTC AAATAGGCGGGACAGATAAATTTACCCACTAATATTAATTTGGTGGGTTTTATTATATAAAAAGTATATCTATGAAAAGGTGAAA ACAATGGTTGAAGTAAAAGAAGAAAAACATTTGATGAACAAATAGATATTTTAAAAAGTAGAGGATTAATAAATAATGATAA AGAAGATGCTAAATTTGTATTAAGTAATGTCAATTATATAGGTTTACAGCATATCTTCTAAGTTTTAAGAATGATGATGGCT CATATAAAGAAGGAACACGTTTGAAGAAGTTTATGATATATATAGGTTTAAATAAGGAATTTAGGATATTTAACAGATTTG TTAGGGAGCATAGAAATAGCATTTAGGACATACATTGCATATACATTAGCAATTAACATGGTGCTTGTGGATATCTAGAAAAG GGAGAGTTTCAAAGATGAAAAATCTATATTAATTTTTTACAGCATTAGAGAGAGAAAAAGTAATAATTCAGATAAGCTTT TTATTATACACCATAAAGAAAAATATGAAGGAAAACCTCCATTTGGGTTGCAACGGAAATAATGACTTTTTGGTATGTTGTCA AAATTAATTCAAACATGTTGCCAGAAGATACTAGATACATAAAAAATAATTTGTGTAGAGTGAATACTTTATTAGTTAAATC TTGGTTACAATCATTAAACACAGGTCAGAAATCAATGCGCTCATTATGGTAGAATATATAA
2	ATAAATTTACCCACTAATATTAATTTGGTGGGTTTTATTATATAAAAAGTATATCTATGAAAAGGTGAAAACAATGGTTGAAGTA AAAGAAGAAAAACATTTGATGAACAAATAGATATTTTAAAAAGTAGAGGATTAATAAATAAATGATAAAGAAGATGCTAAAT TGTATTAAGTAATGCTCAATTATTATAGGTTTACAGCATATCTTCTAAGTTTTAAGAATGATGATGGCTCATATAAAGAAGGAA CTACGTTTGAAGAAGTTTATGATATATATAGGTTTAAATAAGGAATTTAGGATATTTAACAGATTTGTTAGGGAGCATAGAA ATAGCATTTAGGACATACATTGCATATACATTAGCAATTAACATGGTGCTTGTGGATATCTAGAAAAGGAGAGTTTCAAAGA TGAAAAATCTATATTAATTTTTTGCAGCATTAGAGAGAGAAAAAGTAATAATTCAGATAAGCTTTTTATTATACACCATA AAGAAAAATATGAAGGAAAACCTCCCATTTGGGTTGCAACGGAAATAATGACTTTTTGGTATGTTGTCAAAATTTATATTCAAC ATGTTGCCAGAAGATACTAGATACATAAAAAATAATTTGTGTAGAGTGAATACTTTATTAGTTAAATCTTGGTTACAATCATT AACACAGGTCAGAAATCAATGCGCTCATTATGGTAGAATATATAATAATAATTTCCGTATATAACAATAAAAAACGAATATA AAAAGTAACTTGGATAATAAAAAGATATTTTCTTATATACTTGCATGAAGCATTGACTATGGATAAATTAATTTGGAAAT AGTTTTTTTATAAAACCTCAAAGTTAATTAATGATTAATAAATCTATAGACTTAAAGCTTATTGGTTTTCCATAAATTTG GATAGAGATATTGGCTAAATAAAAATAGTTACTTTAAGAAGTTTATAAACACTTACTATATGTAAGTGTTTTTTTAT
3	CAGAACTCGTAAAGCATTATATGCTCAAATATACGGGACATTGATTTTTGAAGGAACCCGCCAAGCTTCTTTTATAGCTCAA ATAGGCGGGACAGATAAATTTACCCACTAATATTAATTTGGTGGGTTTTATTATATAAAAAGTATATCTATGAAAAGGTGAAAAC AATGGTTGAAGTAAAAGAAGAAAAACATTTGATGAACAAATAGATATTTTAAAAAGTAGAGGATTAATAAATAAATGATAAAG AAGATGCTAAATTTGTATTAAGTAATGCTCAATTATTATAGGTTTACAGCATATCTTCTAAGTTTTAAGAATGATGATGGCTCA TATAAAGAAGGAACACGTTTGAAGAAGTTTATGATATATATAGGTTTAAATAAGGAATTTAGGATATTTAACAGATTTGTT AGGGAGCATAGAAATAGCATTTAGGACATACATTGCATATACATTAGCAATTAACATGGTGCTTGTGGATATCTAGAAAAGG AGAGTTCAAAGATGAAAAATCTATATTAATTTTTGACAGCATTAGAGAGAGAAAAAGTAATAATTCAGATAAGCTTTTTT ATTATACACCATAAAGAAAAATATGAAGGAAAACCTCCCATTTGGGTTGCAACGGAAATAATGACTTTTTGGTATGTTGTCAA ATTATATTCAAACATGTTGCCAGAAGATACTAGATACATAAAAAATAATTTGTGTAGAGTGAATACTTTATTAGTTAAATCTT GGTTACAATCATTAAACACAGGTCAGAAATCAATG

Publication 1



Clostridioides difficile – phage relationship the RNA way

Victor Kreis¹ and Olga Soutourina^{1,2}

Clostridioides difficile (formerly *Clostridium difficile*)-associated diarrhea is currently the most frequently occurring nosocomial diarrhea worldwide. During its infection cycle this pathogen needs to survive in phage-rich gut communities. Recent data strongly suggest that regulatory RNAs control gene expression in *C. difficile* and many of these RNAs appear to modulate *C. difficile*-phage interactions. Of the 200 regulatory RNAs identified by deep sequencing and targeted approaches, many function as antitoxins within type I toxin-antitoxin modules and CRISPR RNAs for anti-phage defenses. In this review, we discuss recent insights into the role of RNAs in modulating interactions between *C. difficile* and phages in light of intriguing data in other prokaryotes.

Addresses

¹ Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91198, Gif-sur-Yvette, France

² Institut Universitaire de France (IUF), Paris, France

Corresponding author:

Soutourina, Olga (olga.soutourina@i2bc.paris-saclay.fr)

Current Opinion in Microbiology 2022, 66:1–10

This review comes from a themed issue on **Cell regulation**

Edited by **Aimee Shen** and **Rita Tamayo**

<https://doi.org/10.1016/j.mib.2021.11.012>

1369-5274/© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Bacteriophages (also called phages) are the most abundant and diversified genetic entities in microbial world. They have a great impact on bacterial physiology and evolution, and they also shape the composition of bacterial communities. Phages together with other mobile genetic elements (MGE) participate in horizontal gene transfer (HGT), which contributes to the acquisition of new adaptive traits important for survival, virulence and antibiotic resistance of bacterial pathogens. Upon infection, phages can adopt either lytic or lysogenic cycles [1]. The infection by virulent or lytic phages leads to the lysis of host bacterial cell at the end of the phage replication cycle. Depending on the conditions, temperate phages can opt for lytic or lysogenic cycle inducing the integration of the phage genome into the host bacterial chromosome to become prophage (Figure 1) [1]. Importantly, toxins of several most dangerous bacterial pathogens

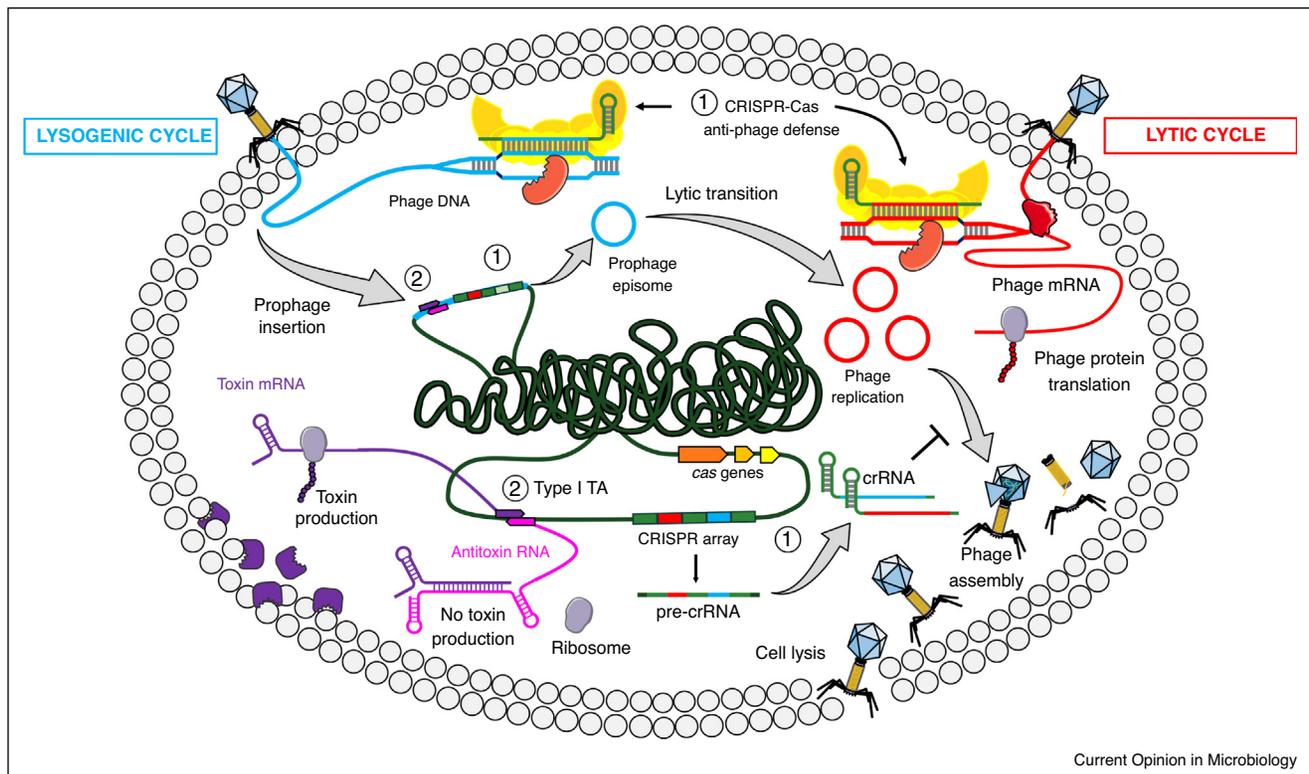
including Shiga-toxin producing *Escherichia coli*, *Clostridium botulinum*, *Corynebacterium diphtheriae* and *Vibrio cholerae* are encoded inside prophages [1].

Bacteria and phages develop sophisticated strategies to co-evolve in various environments including the human gut. To cope with the pressure from their genetic invaders, bacteria have developed efficient defense systems including innate and adaptive immunity such as restriction-modification (RM) and CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR-associated) systems (Box 1) and abortive infection mechanisms (for review see Ref. [2]). Through the constant arms race, phages on their side have developed anti-defense strategies to escape from bacterial host immunity during their life cycle [3].

RNAs in bacteria–phage interactions

Recent findings highlight the importance of non-coding RNAs (ncRNAs) in modulating bacteria–phage interactions [4*,5,6]. CRISPR RNAs interact with double-stranded DNA or RNA to cleave foreign nucleic acids as a defense mechanism, while other RNAs affect the interaction of bacteria with phages or MGEs at the post-transcriptional level (Figure 1, summarized in Ref. [4*]). Phage-encoded RNAs affect bacterial gene expression and can regulate the lysis-lysogeny decision during temperate phage infections, and bacterial RNAs in turn reshape phage gene expression [4*,7]. Recent work in enteric pathogens emphasizes some of the themes in RNA regulation of bacteria–phage and bacteria–MGE interactions [5]. For example, in *Salmonella enterica*, this regulatory network includes both core genome-encoded and horizontally acquired small RNAs (sRNAs) [5,8–11]. In the enterohemorrhagic *E. coli* (EHEC), a recent paper showed that the sRNA, StxS, which is generated when transcription of the gene encoding Shiga toxin prematurely terminates in a prophage, acts *in trans* to control the host genes by activating the expression of the general stress response sigma factor, RpoS [12]. An additional layer has been recently included into these cross-regulations by identification of anti-sRNAs encoded by prophages in pathogenic *E. coli* [13]. Such ‘sponge’ RNAs target sRNAs to interfere with their activity. For example, the EHEC phage-derived AsxR and AgvB sRNAs antagonize core genome-encoded sRNAs to impact virulence gene expression [13]. Transcriptomic studies continue to detect ncRNAs encoded by phages and prophages of other pathogens, for example, sRNA1 and sRNA2 of PAK_P3 phage of *Pseudomonas aeruginosa* [14,15]. Host bacterial sRNAs may also participate in the regulation of anti-phage CRISPR-Cas system, for example, pant463

Figure 1



Summary of RNA-based mechanisms in *C. difficile*-phage interactions.

The lysogenic and lytic cycles are respectively represented in blue and red. Bacterial genome is green colored, except for CRISPR-Cas and type I toxin-antitoxin (TA) system components. (1) In order to face the phage invasion, bacteria use CRISPR-Cas anti-phage defense system. CRISPR array is transcribed into pre-crRNA, which is then matured in crRNAs to finally guide Cas effector (yellow) to complementary sequence where Cas nuclease (orange) will cut phage DNA. (2) Among TA systems implicated in bacteria-phage interplay, type I TA module encodes a small hydrophobic toxic protein (purple) and an antisense RNA (pink) acting as an antitoxin to alter the toxin mRNA stability and/or translation. These systems are encoded both inside prophages and outside MGEs in the *C. difficile* genome.

and PhrS sRNAs that activate and pant391 sRNA that represses CRISPR expression in *P. aeruginosa* [16].

Clostridioides difficile

C. difficile is a Gram-positive, spore-forming, strictly anaerobic bacterium found in environment and in mammalian intestinal tract. This human enteropathogen is a leading cause of nosocomial diarrhea in adults associated generally with antibiotic treatment. Among the most alarming aspects of *C. difficile*-associated infections are the increasing incidence of community-acquired infections, emergence of antibiotic resistance and hypervirulence, high rates of recurrent cases, and increasing severity of disease associated with standard treatment failure [17]. All these worrying points motivate the basic and applied research to better understand the mechanisms controlling the *C. difficile*'s infection cycle to facilitate the development of alternative therapeutic strategies.

C. difficile usually produces two toxins, TcdA and TcdB, that are major virulence factors encoded inside the 19.6-

kb pathogenicity locus (PaLoc) together with regulatory components TcdR, TcdC and TcdE [17]. Some *C. difficile* strains produce a binary toxin, *C. difficile* transferase (CDT), encoded in a 6.2-kb locus. Additional factors that help colonization have been identified like adhesins, pili, and flagella [18].

To survive inside the host environment, *C. difficile* has developed sophisticated adaptation strategies including efficient defense systems to cope with the presence of abundant and diverse phage invaders. High specificity of phage targeting makes phage therapy an interesting alternative to existing treatments of *C. difficile* infections (reviewed in Ref. [19]). Harnessing endogenous CRISPR-Cas and toxin-antitoxin (TA) systems as novel therapeutic strategies against *C. difficile* infections open interesting perspectives for alternative and synergistic treatments [20,21,22,23,24]. The knowledge of these systems also allowed the development of efficient genome editing strategies for *C. difficile* [20,22].

Box 1 CRISPR-Cas and TA systems in prokaryotes

The CRISPR-Cas prokaryotic adaptive immunity systems provide defense against foreign nucleic acids [6]. CRISPR loci are arranged in arrays of direct repeats separated by variable sequences called spacers that match viral or plasmid DNA and have been acquired during prior encounters with MGE in the 'adaptation' step [25]. CRISPR arrays are transcribed as single RNA transcripts (pre-crRNA) that are processed to generate small CRISPR RNAs (crRNAs). In complex with Cas effector proteins, these crRNAs interfere with phage infection and plasmid transfer by recognizing foreign nucleic acids that leads to their destruction during an 'interference' process (Figure 1). Either a single Cas protein effector in class 2 (type II, V, and VI) or a multi-subunit Cas protein effector complex in class 1 (type I, III, and IV) systems together with the mature crRNA achieves the interference step [26].

Toxin-antitoxin (TA) modules encode two-component systems consisting of a stable 'toxin' and an unstable 'antitoxin' to repress the function or expression of the toxin (see Ref. [27] for review). The overexpression of toxin either kills the cells or confers growth stasis. TA systems have been initially discovered on plasmids where they confer stability of maintenance through post-segregational killing [28]. TA systems are encoded on bacterial and archaeal chromosomes including pathogens, but their function remains largely unclear. Among their suggested functions are prophage maintenance, hypothetical chromosomal region stabilization outside the MGE, prevention of phage infection, stress response and persist formation [27,29,30]. TA systems are classified into six types depending on the nature and the action of the antitoxin that can be either a protein or an sRNA [27]. In type I TA, the antitoxin is a small antisense RNA that neutralizes toxin mRNA by inhibiting its translation and/or promoting its degradation [27,30] (Figure 1). Type I toxins are usually small hydrophobic proteins less than 60 amino acids in length that contain a potential transmembrane domain and charged amino acids at the C-terminus [31].

The purpose of this review is to give an overview of the RNAs contributing to the interactions of *C. difficile* with phages focusing on CRISPR-Cas and TA systems (Box 1) in light of intriguing recent data in other prokaryotes. We will first describe what is known about the contribution of phages to *C. difficile* physiology and then highlight the recent findings on the emerging roles of RNAs in *C. difficile*-phage interactions. Open questions that arise from these first observations and future directions with potential roles of additional classes of regulatory RNAs will be discussed. This increasing knowledge will provide essential basis for future development of new molecular tools as well as promising diagnostic and therapeutic applications.

Relationship between *C. difficile* and its phages

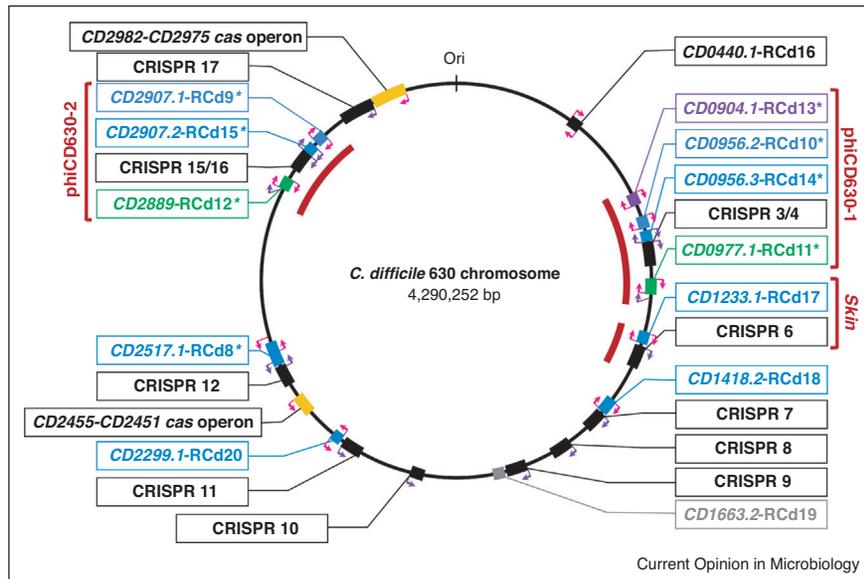
Phages infect almost every bacterial species, and prophage regions are found in most of the bacterial genomes. *C. difficile* is not an exception. During its infection cycle, *C. difficile* survives within phage-rich gut communities and is therefore expected to possess efficient systems to control the genetic exchanges favored in such environments. Indeed, the genome of *C. difficile* is highly mobile and mosaic, with 11% of strain 630's genome possibly originating from MGEs. These genetic exchanges likely

facilitate the acquisition of adaptive traits for gastrointestinal growth [32]. This acquisition needs to be balanced by the efficient defense against abundant foreign genetic elements such as phages and plasmids. All *C. difficile* phages isolated and characterized to date are considered temperate meaning that their genome can be integrated into the host bacterial genome as a prophage [19**]. Prophages are highly enriched in *C. difficile* genomes based on extensive genome sequencing [33]. Each *C. difficile* strain carries usually between 1 and 3 and up to 6 prophages in addition to phage-related genomic regions (see Ref. [19**] for review). Some of prophages including large phage genomes of more than 130 kb can be stably maintained as extrachromosomal DNA in *C. difficile* [33]. The outcome of phage interactions with a susceptible *C. difficile* strain is likely related to the availability of suitable receptors, superinfection exclusion mechanisms provided by endogenous prophages expressing proteins that block phage binding to bacterial surface receptor or DNA injection, and the presence of efficient defense systems including CRISPR-Cas [19**,21,34].

C. difficile phages also contribute to virulence-associated phenotypes. For example, a locus encoding the Cdt toxin has been identified in prophage phiSemix9P1 in some *C. difficile* strains [35], and the pathogenicity locus (PaLoc) shares features with phages, since it encodes a phage-like holin, TcdE [19**]. *C. difficile* phages phiCD119 and phiCD38-2 have been shown to influence negatively or positively the toxin production in this pathogen (reviewed in [19**]). For instance, the RepR regulator expressed from the phiCD119 prophage binds to *tcdR* promoter leading to repression of toxin gene expression dependent on TcdR toxin-specific sigma factor [36]. The genomes of *C. difficile* prophages encode genes that could interfere with bacterial gene regulations affecting surface proteins, antibiotic resistance and quorum sensing (QS) [19**,33]. For example, phiCDHM1 and related prophages carry Agr-like QS genes [37] that may influence important virulence-associated phenotypes in *C. difficile*. In *Staphylococcus aureus*, QS system includes a key ncRNA, RNAIII, as a central player in the regulation of virulence [38], whether an sRNA participates in QS regulations in *C. difficile* remains an open question. Transcriptome analysis during lysogeny revealed active prophage gene expression and potential crosstalk between resident phi027 and introduced phiCD38-2 prophage and also an impact on host gene expression in R20291 strain [39]. For example, phiCD38-2 upregulated several genes involved in DNA replication and gene regulation encoded within phi027.

Suppression of anti-phage defense systems like CRISPR-Cas, RM, and type II TA has been reported during JD032 phage infection of *C. difficile* ribotype 078 strain, TW11 [40]. Thus, in addition to the reprogramming of host bacteria metabolism during phage lytic cycle, this

Figure 2



Genomic map of type I TA and CRISPR-Cas components encoded in *C. difficile* strain 630 genome.

The location of type I TA modules is shown in blue for loci adjacent to CRISPR arrays, in green for c-di-GMP responsive riboswitch-associated loci, and in purple for additional prophage-associated TA loci. The prophage position is indicated using red lines for phiCD630-1, phiCD630-2, and skin element. * indicates TA modules with detailed characterization [22*,24,52*]. CRISPR arrays are numbered according to CRISPRdb (<https://crispr.i2BC.paris-saclay.fr/crispr/>) and previous publications [21,34,46,51,52*]. The CD1663.2-RCd19 locus associated with CRISPR 9 array but encoding a small protein with a divergent sequence is indicated in grey. « Ori » indicates the origin of replication. The purple arrow and the pink arrow indicate the position of transcriptional start sites associated with Sigma A-dependent or Sigma A-dependent and Sigma B-dependent promoters, respectively. Cas operons are indicated in yellow.

RNA-seq analysis also revealed the massive downregulation of several defense systems to ensure the effective phage infection. The prophage can also influence the relationship of host bacteria with other infecting phages for example by carrying superinfection exclusion genes that provide phage immunity [10]. In *C. difficile*, prophages encode components of CRISPR-Cas defense systems, with highly expressed and functional CRISPR 3/4 and CRISPR 15/16 arrays in phiCD630-1 and phiCD630-2 prophages of 630 strain (Figure 2), two transcriptionally active CRISPR arrays in phi027 prophage of epidemic strains of 027 ribotype and a CRISPR array with *cas3* gene in phiCD211-like phages [33,34,37,39].

Regulatory RNAs in *C. difficile* interplay with phages

Recent work has led to the discovery that RNAs modulate *C. difficile*-phage interactions. This advance has been facilitated by combining *in silico* analyses, RNA-seq, and genome-wide promoter mapping, which identified more than 200 ncRNAs in this pathogen [41*,42,43*,44,45*,46] (summarized in Ref. [44]). In the following sections we will give an overview of regulatory RNAs identified in *C. difficile* that could contribute to its interplay with phages (Figure 2) (Box 2).

CRISPR RNAs

CRISPR RNAs (Box 1) were among the most abundant RNAs detected in *C. difficile* by deep sequencing [46]. An active *C. difficile* type I-B CRISPR-Cas system is in many

Box 2 *C. difficile* RNA nomenclature

In the paper describing the first experimental genome-wide identification of ncRNAs in *C. difficile* [46] we introduced the RCd names for selected RNAs studied in more detail similar to SR, Rsa, Rli sRNA names in *Bacillus subtilis*, *S. aureus* and *Listeria monocytogenes*, respectively [47]. In addition to SQ RNAs initially predicted *in silico* and confirmed by RNA-seq, general CD630_n annotation has been established according to the genomic position in 630Δ*erm* strain [46] that was then adopted by BioCyc Genome database (<https://biocyc.org>) and MAGE Microscope Microbial Genome Annotation and Analysis Platform (<https://mage.genoscope.cns.fr>). The common or Rfam database names are generally used for riboswitches and conserved housekeeping RNAs [48]. Genome-wide *in silico* and high throughput transcriptomics and interactomics analyses introduced alternative names for potential ncRNAs in *C. difficile*, that is, sCD, SQ, CDIF630nc or nc [42,43*,46,49]. We suggest to adopt RCd nomenclature for future detailed characterization of ncRNAs in this pathogen. Table 1 describes the ncRNAs RCd1-RCd20 studied in *C. difficile* emphasizing their position related to prophages as well as their potential role in *C. difficile*-phage interplay, as well as CRISPR 3 – CRISPR 16 RNAs of CRISPR-Cas system [21,34,46].

Table 1

List of regulatory RNAs studied in *C. difficile* and their potential implication in interplay with phages

RNA ID [Ref] ^a	Soutourina <i>et al.</i> , PLoS Genetics [46] alternative name	Fuchs <i>et al.</i> , PNAS [43*] alternative name	5' start	3' end	Strand	Function/comments	Association	Prophage location
RCd1 [46] ^b	CD630_n00660	CDIF630nc_070	2 199 358	2 199 543	+	ncRNA IGR	–	–
RCd2 [46]	CD630_n00030	CDIF630nc_003	240 878	241 078	–	<i>cis</i> -antisense (5' UTR)	–	–
RCd3 [46]	CD630_n00410	CDIF630_01210	1 259 900	1 260 092	+	ncRNA IGR, 6S RNA	–	–
RCd4 [46]	CD630_n00210	–	655 072	655 670	+	ncRNA IGR	–	–
RCd5 [46]	CD630_n00680	–	2 285 913	2 286 352	+	ncRNA IGR	<i>cdi1_6</i>	–
RCd6 [46]	CD630_n00170	CDIF630nc_012	560 200	560 340	–	ncRNA IGR	–	–
RCd7 [71]	CD630_n00040	CDIF630nc_004	280 449	280 676	+	ncRNA IGR	–	–
RCd8 [52*]	CD630_SQ1781	CDIF630nc_081	2 908 031	2 908 258	+	Antitoxin type I	Toxin CD2517.1	–
RCd9 [52*,22*]	CD630_n01000	CDIF630nc_096	3 398 302	3 398 435	+	Antitoxin type I (long form)	Toxin CD2907.1	phiCD630-2
RCd10 [52*,22*]	CD630_n00370	CDIF630nc_032	1 124 339	1 124 041	–	Antitoxin type I (short form)	Toxin CD0956.2	phiCD630-1
RCd11 [22*]	CD630_n00390	CDIF630nc_035	1 142 439	1 142 289	–	Antitoxin type I (long form)	Toxin CD0977.1 <i>cdi1_5</i>	phiCD630-1
RCd12 [22*]	CD630_n00980	CDIF630nc_093	3 377 979	3 380 352	+	Antitoxin type I (short form)	Toxin CD2889 <i>cdi1_4</i>	phiCD630-2
RCd13 [22*]	AS ^c CD630_n00350 ^d	–	1 089 726	1 089 627	–	Antitoxin type I	Toxin CD0904.1	phiCD630-1
RCd14 [22*]	AS <i>CD0965.3</i>	CDIF630nc_033	1 124 789	1 124 691	–	Antitoxin type I	Toxin CD0956.3	phiCD630-1
RCd15 [22*]	AS <i>CD2907.2</i>	CDIF630nc_098	3 397 852	3 397 977	+	Antitoxin type I	Toxin CD02907.2	phiCD630-2
RCd16 ^f [24]	CD630_n00150	CDIF630nc_011	523 347	523 449	–	Antitoxin type I	Toxin CD0440.1	–
RCd17 ^f [52*]	SQ808	CDIF630nc_042	1 434 082	1 434 312	–	Antitoxin type I	Toxin CD1233.1	<i>skin</i> element
RCd18 ^f [52*]	CD630_n00500	CDIF630nc_049	1 644 858	1 644 958	–	Antitoxin type I	Toxin CD1418.2	–
RCd19 ^f [52*]	CD630_n00610	CDIF630nc_062	1 937 370	1 937 470	–	Antitoxin type I ^e	Toxin CD1663.2 ^e	–
RCd20 ^f [52*]	SQ1641	CDIF630nc_076	2 665 093	2 665 330	+	Antitoxin type I	Toxin CD2299.1	–
CRISPR 3/4	CD630_n00380	–	1 124 818	1 125 820	+	CRISPR array	RCd10	phiCD630-1
CRISPR 6	CD630_n00460	–	1 434 587	1 435 639	+	CRISPR array	RCd17	<i>skin</i> element
CRISPR 7	CD630_n00510	–	1 645 024	1 646 018	+	CRISPR array	RCd18	–
CRISPR 8	CD630_n00560	–	1 756 109	1 756 833	+	CRISPR array	–	–
CRISPR 9	CD630_n00600	–	1 935 311	1 936 495	+	CRISPR array	RCd19	–
CRISPR 10	CD630_n00690	–	2 298 940	2 298 609	–	CRISPR array	–	–
CRISPR 11	CD630_n00790	–	2 664 038	2 662 916	–	CRISPR array	RCd20	–
CRISPR 12	CD630_n00860	–	2 908 058	2 907 464	–	CRISPR array	RCd8	–
CRISPR 15/16	CD630_n00990	–	3 397 912	3 396 821	–	CRISPR array	RCd9	phiCD630-2
CRISPR 17	CD630_n01010	–	3 456 904	3 455 483	–	CRISPR array	–	–

^a RCd nomenclature and CRISPR RNA identified in *C. difficile* strain 630. Type I TA modules have been included into T1TA data base (<https://d-lab.arna.cnrs.fr/t1tadb>).

^b The ID RNA named sCD2200 in Ref. [42].

^c «AS» means antisense RNA.

^d *CD0904.1* gene was previously annotated CD630_n00350.

^e This module has a divergent sequence.

^f New names introduced in this review.

aspects unique with unusually large set of CRISPR arrays and *cas* operons, markedly enriched compared with other species, providing large defense capacity for *C. difficile* adaptation inside the host [21,34,44,46]. Indeed, genome sequencing, RNA-seq and bioinformatics analysis identified a high number of CRISPR arrays, 8.5 on average per *C. difficile* genome and the presence of up to three *cas*

operons in the majority of sequenced strains [21,32,34,37,46,50]. Another striking feature of *C. difficile* CRISPR-Cas system is the location of several CRISPR arrays inside the prophages (Table 1) (Figure 2) [34,37]. The expression of all CRISPR arrays has been observed in *C. difficile* 630 Δ *erm* and R20291 strains, the crRNAs from prophage-encoded CRISPR arrays are the most

abundant and the most active for interference [34,46,51]. The stable processed crRNAs from phage-associated arrays as the functional units of CRISPR-Cas system have been constitutively detected by Northern blot throughout the growth, while the full-length pre-crRNA and intermediate transcripts were more abundant during exponential growth as compared to the onset of the stationary phase [46]. Similar expression profiles from exponential to stationary phase have been observed by qRT-PCR for all CRISPR arrays in *C. difficile* 630 Δ erm strain [51]. The analysis of the CRISPR system targeting revealed that more than one-third of all analyzed *C. difficile* CRISPR spacers matched the *Clostridium* phages [34]. Given the limited number of available Clostridial phage genomes these proportions could even be underestimated [19**]. Like many bacteria, *C. difficile* incorporates phage-derived sequences into CRISPR arrays as memories of past encounters with a given phage [21,34,37]. Interestingly, several *C. difficile* CRISPR spacers simultaneously target the related phage genomes [34]. Such multiple targeting might provide *C. difficile* with an efficient and economical defense against several related phages by the same crRNA thus increasing its fitness inside the phage-rich gut communities. Experimental evidence of defensive function of *C. difficile* CRISPR-Cas system and active spacer acquisition by this bacterial I-B system has been provided and 5'-CCW defined as protospacer-adjacent motif (PAM) essential for self/non-self discrimination [34,51]. A functional link between the adaptation and interference machineries for *C. difficile* CRISPR-Cas was revealed, since newly acquired spacers are derived from sequences associated with a CCN PAM, which fits the YCN interference consensus. This means that both *C. difficile* CRISPR-Cas adaptation and interference modules share the PAM recognition consensus even if these sequence requirements are stricter during spacer acquisition than those during interference. The majority of new spacers in *C. difficile* 630 Δ erm are acquired from protospacers containing the CCN PAMs, which are the most efficient for interference [51]. Based on this knowledge a new method for genome editing in *C. difficile* using its native CRISPR-Cas system has been recently developed [20]. The analysis of CRISPR spacer content and CRISPR arrays organization in various *C. difficile* strains revealed a large diversity of sequences [34,50]. This information about CRISPR spacer polymorphism and array composition provides interesting perspectives for new high-resolution typing approaches of *C. difficile* strains among potential medical applications to monitor the microevolution and adaptation events in this pathogen [50].

Type I toxin-antitoxin systems

The detection of antisense RNAs encoded in close proximity to CRISPR arrays led to the discovery of genes encoding small proteins (~50 amino acids) that had escaped automatic annotation [46,52*]. A closer look

revealed that these small proteins all have characteristic features of toxins from type I TA systems with a potential transmembrane domain and charged amino acids at the C-terminus (Box 1). Further experiments confirmed their toxic nature and revealed the first antisense RNAs that function as antitoxins in *C. difficile* [52*]. For the majority of these TA modules, in addition to Sigma A-dependent promoter, a Sigma B-dependent promoter was identified upstream of the toxin and antitoxin genes [22*,24,52*] (Figure 2). The intriguing association and co-regulation of these TA modules with CRISPR arrays by general-stress response sigma B and biofilm-related factors such as high intracellular c-di-GMP content [52*] raised the question on the potential functional link between the cell dormancy mechanism and adaptive immunity system. In addition, both the partial and complete *cas* operons (*CD2455* and *CD2982*) encoding the components of type I-B CRISPR-Cas system in *C. difficile* 630 Δ erm strain are regulated by Sigma B factor and are associated with potential Sigma B-dependent promoters [52*] (Figure 2). Five additional type I TA highly conserved within *C. difficile* prophages have been then described that contributed to the stability of these genomic regions [22*]. The localization of type I TA within prophages and pathogenicity islands has also been observed in *Bacillus subtilis* and *S. aureus*, respectively [53,54]. Until now a total of 13 type I TA modules were identified in the genome of *C. difficile* strain 630, 8 within prophages [24] (Table 1) (Figure 2). These TA modules have been included into the T1Tadb developed by Fabien Darfeuille laboratory (<https://d-lab.arna.cnrs.fr/t1tadb>, [55]). A careful inspection of available *C. difficile* phage genomes revealed even more variants of small proteins from 30 to 56 amino acids within type I TA [22*]. Many questions remain to be addressed on these enigmatic TA modules. Notably, the inducible expression of one of these type I toxins led to the development of the improved genetic tool for the *C. difficile* genome editing [22*].

Interplay between TA and CRISPR defense against phage infection

The role in phage inhibition has been suggested among possible functions of chromosomally encoded TA modules [56*]. In *E. coli*, the *hok/Sok* TA system inhibits T4 phage replication, although the underlying mechanism remains unclear [57]. A more recent study showed that a type III *toxIN* TA system protects *E. coli* against multiple phages [58**]. This work showed that the phage activates the toxin by shutting down the host transcription, leading to the fast degradation of the RNA antitoxin. In *C. difficile*, the link between type I TA and adaptive CRISPR-Cas immunity could contribute to the evolution of the chromosome and fitness inside the host. This functional interplay could induce a state of dormancy facilitating CRISPR-Cas-mediated adaptive immunity, in accordance with the recently emerged concept of functional coupling of defense systems in prokaryotes [59*].

The role of this functional coupling between distinct defense strategies might be the induction of dormancy state in infected or stressed cells to allow the activation of adaptive immunity or specific stress responses. Dormancy was suggested to be a strategy of the last resort when the defense strategies fail face of genetic invaders. The induction of CRISPR encoding prophages could also contribute to the spread of CRISPR arrays within the population. In *C. difficile*, type I TA play a role in stabilizing the chromosomal regions carrying CRISPR-Cas systems and thus ensure the retention of CRISPR loci. In particular, by studying four type I TA pairs located within prophage phiCD630-1 carrying CRISPR 3/4 arrays, we showed that these modules are able to confer both plasmid and prophage stability in the absence of selective pressure [22*]. We demonstrated that these TA modules are important for phiCD630-1 maintenance after its excision since the prophage remained in less than 10% of the population in strain deleted for four TA while was still present in 90% of the population in the parental strain after 24 hour of growth. This evolutionary perspective is strengthened by the fact that the physical genomic link between TA pairs and CRISPR arrays can be extended to the majority of sequenced *C. difficile* strains [52*].

In line with these observations, several recent discoveries highlight the connections between defense systems including CRISPR-Cas and TA inside the ‘defense islands’ [2,60]. Cyanobacterial HEPN-MNT TA system encoded in the vicinity of a subtype I-D CRISPR-Cas system has been suggested to contribute to antiviral defense acting as a cellular ATP sensor during phage infection, however the exact mechanism and the crosstalk with CRISPR-Cas remain to be established [61]. In addition to potential functional coupling of immunity and cell dormancy systems [59*], together with our results in *C. difficile*, recent discovery of CreTA system in haloarchaeal species including *Haloarcula hispanica* suggests the direct contribution of TA systems for the maintenance of CRISPR loci within a prokaryotic genome [22*,52*,62**]. In the absence of dedicated maintenance mechanisms, the frequent loss of CRISPR-Cas systems could be hypothesized due to its fitness cost for the bacterial host. While providing the efficient protection against genetic parasites, CRISPR-Cas could lead to the autoimmunity and decreases the new beneficial gene acquisition. In a search for a potential mechanism preventing the loss of CRISPR-Cas system, the recent study identified a new TA module composed of a pair of RNAs that are encoded within the archaeal type I-B CRISPR-Cas locus [62**]. This two-RNA CreTA module consists of the *creT* RNA toxin and CRISPR RNA variant CreA antitoxin, which represses the toxin jointly with CRISPR effector Cascade complex. CreTA system safeguards CRISPR-Cas making the CRISPR effector addictive to the host cell, since in the absence of Cascade the CreT RNA acts to sequester the rare tRNA-Arg leading to the

cell growth arrest. Indeed, CreA becomes antitoxic only in the presence of Cascade, since the partial complementarity between CreA and the *creT* promoter guides Cascade to repress toxin transcription. In CreTA-deleted strain, the genomic region for Cascade components becomes susceptible to disruption by transposon insertion. *In silico* searches identified potential CreTA analogues associated with other haloarchaeal but also some bacterial subtype I-B CRISPR-Cas, as well as with other archaeal subtype I-D and bacterial subtype III-A CRISPR-Cas loci suggesting that this mechanism could be widespread in prokaryotes [62**].

Other ncRNAs that could participate in *C. difficile* interactions with phages

Riboswitches

Riboswitches belong to the class of 5' *cis*-acting RNA regulatory elements that are largely used for the control of gene expression in *C. difficile*. These sequences fold into alternative structures upon binding of effector molecules leading to positive or negative effects on transcription termination, translation, or RNA stability [63]. Among them, cyclic-di-guanosyl-5' monophosphate (c-di-GMP) riboswitches were the first functional ncRNAs reported for *C. difficile* [46,64–66]. C-di-GMP is an important signaling molecule controlling lifestyle switches from free-living motile state to biofilm communities and virulence in bacteria [66,67]. 16 c-di-GMP dependent riboswitches are used as RNA effectors to coordinately control the *C. difficile* infection cycle [44,46,66,68]. Two of these riboswitches, *cdi1_4* and *cdi1_5*, are located inside the phiCD630-1 and phiCD630-2 prophages of *C. difficile* 630 that are associated with antitoxin RNAs from type I TA with potential role in tight regulation of TA toxin production [22*]. In addition to their role in prophage stabilization, these prophage-associated TA systems might link the TA toxin expression to the c-di-GMP signaling in *C. difficile* leading to the inhibition of growth when c-di-GMP levels increase.

Hfq-dependent RNAs

The RNA chaperone protein Hfq has been generally implicated in the regulations mediated by *trans*-encoded sRNAs in Gram-negative bacteria [69,70]. *C. difficile* Hfq protein has unique features in Gram-positive bacteria and likely functions as a key component of post-transcriptional regulatory network [20,41*,43*,44,71]. RNA immunoprecipitation high-throughput sequencing (RIP-seq) analysis identified a large set of Hfq-associated ncRNAs and mRNAs in *C. difficile* [41*,43*]. Among previously identified ncRNAs, in addition to *trans*-encoded RNAs, a number of *cis*-antisense RNAs including the antitoxins from all recently identified type I TA modules [22*,24,52*], numerous riboswitches, and CRISPR RNAs have been enriched in Hfq-associated samples constituting new categories of Hfq ligands [41*]. These findings suggest the new regulatory roles for CRISPR RNAs

beyond the adaptive immunity and potential action *in trans* for *cis*-antisense RNAs including antitoxins beyond toxin mRNA targeting. Recent examples in *Campylobacter jejuni* illustrate the noncanonical regulatory roles of crRNAs and the RNA-guided reprogramming of CRISPR effector Cas9 for gene regulation [72–74]. In *Francisella novicida*, Cas9 represses the expression of immunostimulatory BLP lipoprotein through CRISPR-Cas associated tracrRNA and scaRNA by an unknown mechanism [75]. Our recent identification of crRNAs as the partners of Hfq for mRNA targeting in *C. difficile* [41^{*}] suggests a possibility for the regulatory action independent from Cas machinery reminiscent to the *trans*-encoded sRNA action.

Cross-regulations between *C. difficile* prophage-encoded RNAs and bacterial RNAs might be envisaged and deserve further experimental validations in line with the role of Hfq in RNA-based mechanisms of phage–bacteria interactions in enterobacteria [4^{*},13]. Powerful Grad-seq approach recently revealed extensive RNA–protein interactions in *C. difficile* opening interesting perspectives for future identification of new RNA-binding proteins and their RNA targets [49].

Conclusions

Phage genomes carry a number of genes of unknown function, some of them have been evolved to maintain the cohabitation with host bacteria during lysogenic cycle. *C. difficile* phages represent yet unexplored reservoir of ncRNAs contributing to the complex relationship this pathogen develops with phages. In addition to already identified components of these RNA-based interactions involving CRISPR RNAs and TA RNAs, further studies will uncover new ncRNAs that could participate in host bacteria and phage gene regulation. As described in other bacteria, various prophages found in *C. difficile* genomes could constitute a source of ncRNAs that remain to be studied. In *C. difficile* 630 Δ erm, several additional sRNAs have been revealed by high throughput sequencing encoded inside phiCD630-1, phiCD630-2 prophages and *skin* prophage-like element that deserve further investigations. Understanding all these aspects of bacteria–phage cross-regulations at a post-transcriptional level is of great importance for the development of new phage-based therapeutics to deal with recalcitrant bacterial infections.

Conflict of interest statement

Nothing declared.

Acknowledgements

This work was supported by grants from the Agence Nationale de la Recherche (‘CloSTARn’, ANR-13-JSV3-0005-01 to O.S.), the Institut Universitaire de France (to O.S.), the University Paris-Saclay, the Institute for Integrative Biology of the Cell, the DIM-1HEALTH regional Ile-de-France program (LSP grant no. 173403, V.K. PhD fellowship), the CNRS-RFBR PRC 2019 (grant no.288426 19-54-15003) to O.S.

We thank E. Semenova for help in preparation of this review. We apologize to authors whose work could not be cited in this review because of the space limitations.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Fortier LC, Sekulovic O: **Importance of prophages to evolution and virulence of bacterial pathogens.** *Virulence* 2013, **4**:354-365.
2. Bernheim A, Sorek R: **The pan-immune system of bacteria: antiviral defence as a community resource.** *Nat Rev Microbiol* 2020, **18**:113-119.
3. Malone LM, Birkholz N, Fineran PC: **Conquering CRISPR: how phages overcome bacterial adaptive immunity.** *Curr Opin Biotechnol* 2021, **68**:30-36.
4. Altuvia S, Storz G, Papenfort K: **Cross-regulation between bacteria and phages at a posttranscriptional level.** *Microbiol Spectr* 2018, **6**.
- A recent review summarizing the RNA-based cross-regulations between phages and their host bacteria.
5. Frohlich KS, Papenfort K: **Interplay of regulatory RNAs and mobile genetic elements in enteric pathogens.** *Mol Microbiol* 2016, **101**:701-713.
6. Nussenzweig PM, Marraffini LA: **Molecular mechanisms of CRISPR-Cas immunity in bacteria.** *Annu Rev Genet* 2020, **54**:93-120.
7. Bloch S, Lewandowska N, Wegrzyn G, Nejman-Falenczyk B: **Bacteriophages as sources of small non-coding RNA molecules.** *Plasmid* 2021, **113**:102527.
8. Correia Santos S, Bischler T, Westermann AJ, Vogel J: **MAPS integrates regulation of actin-targeting effector SteC into the virulence control network of *Salmonella* small RNA PinT.** *Cell Rep* 2021, **34**:108722.
9. Hershko-Shalev T, Odenheimer-Bergman A, Elgrably-Weiss M, Ben-Zvi T, Govindarajan S, Seri H, Papenfort K, Vogel J, Altuvia S: **Gifsy-1 prophage IsrK with dual function as small and messenger RNA modulates vital bacterial machineries.** *PLoS Genet* 2016, **12**:e1005975.
10. Owen SV, Canals R, Wenner N, Hammarlof DL, Kroger C, Hinton JCD: **A window into lysogeny: revealing temperate phage biology with transcriptomics.** *Microb Genom* 2020, **6**.
11. Westermann AJ, Forstner KU, Amman F, Barquist L, Chao Y, Schulte LN, Muller L, Reinhardt R, Stadler PF, Vogel J: **Dual RNA-seq unveils noncoding RNA functions in host-pathogen interactions.** *Nature* 2016, **529**:496-501.
12. Sy BM, Lan R, Tree JJ: **Early termination of the Shiga toxin transcript generates a regulatory small RNA.** *Proc Natl Acad Sci U S A* 2020, **117**:25055-25065.
13. Tree JJ, Granneman S, McAteer SP, Tollervey D, Gally DL: **Identification of bacteriophage-encoded anti-sRNAs in pathogenic *Escherichia coli*.** *Mol Cell* 2014, **55**:199-213.
14. Chevallereau A, Blasdel BG, De Smet J, Monot M, Zimmermann M, Kogadeeva M, Sauer U, Jorth P, Whiteley M, Debarbieux L et al.: **Next-generation “-omics” approaches reveal a massive alteration of host RNA metabolism during bacteriophage infection of *Pseudomonas aeruginosa*.** *PLoS Genet* 2016, **12**: e1006134.
15. Dedrick RM, Mavrich TN, Ng WL, Hatfull GF: **Expression and evolutionary patterns of mycobacteriophage D29 and its temperate close relatives.** *BMC Microbiol* 2017, **17**:225.
16. Lin P, Pu Q, Wu Q, Zhou C, Wang B, Schettler J, Wang Z, Qin S, Gao P, Li R et al.: **High-throughput screen reveals sRNAs regulating crRNA biogenesis by targeting CRISPR leader to repress Rho termination.** *Nat Commun* 2019, **10**:3728.

17. Smits WK, Lyras D, Lacy DB, Wilcox MH, Kuijper EJ: ***Clostridium difficile* infection**. *Nat Rev Dis Primers* 2016, **2**:16020.
18. Janoir C: **Virulence factors of *Clostridium difficile* and their role during infection**. *Anaerobe* 2016, **37**:13-24.
19. Heuler J, Fortier L-C, Sun X: ***Clostridioides difficile* phage**
 ●● **biology and application**. *FEMS Microbiol Rev* 2021, **45** fuab012
 A detailed review on current knowledge of *C. difficile* phages and related applications.
20. Maikova A, Kreis V, Boutserin A, Severinov K, Soutourina O: **Using an endogenous CRISPR-Cas system for genome editing in the human pathogen *Clostridium difficile***. *Appl Environ Microbiol* 2019, **85**.
21. Maikova A, Severinov K, Soutourina O: **New insights into functions and possible applications of *Clostridium difficile* CRISPR-Cas system**. *Front Microbiol* 2018, **9**:1740.
22. Peltier J, Hamiot A, Garneau JR, Boudry P, Maikova A,
 ● Hajnsdorf E, Fortier LC, Dupuy B, Soutourina O: **Type I toxin-antitoxin systems contribute to the maintenance of mobile genetic elements in *Clostridioides difficile***. *Commun Biol* 2020, **3**:718
 First characterization of the role of type I TA system in prophage stability and the development of efficient genetic tool using this system as counter-selectable marker.
23. Selle K, Fletcher JR, Tuson H, Schmitt DS, McMillan L, Vridhambal GS, Rivera AJ, Montgomery SA, Fortier LC, Barrangou R *et al.*: **In vivo targeting of *Clostridioides difficile* using phage-delivered CRISPR-Cas3 antimicrobials**. *mBio* 2020, **11**.
24. Soutourina O: **Type I toxin-antitoxin systems in *Clostridia***. *Toxins (Basel)* 2019, **11**.
25. Jackson SA, McKenzie RE, Fagerlund RD, Kieper SN, Fineran PC, Brouns SJ: **CRISPR-Cas: adapting to change**. *Science* 2017, **356**.
26. Makarova KS, Wolf YI, Iranzo J, Shmakov SA, Alkhnbashi OS, Brouns SJJ, Charpentier E, Cheng D, Haft DH, Horvath P *et al.*: **Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants**. *Nat Rev Microbiol* 2020, **18**:67-83.
27. Harms A, Brodersen DE, Mitarai N, Gerdes K: **Toxins, targets, and triggers: an overview of toxin-antitoxin biology**. *Mol Cell* 2018, **70**:768-784.
28. Hayes F: **Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest**. *Science* 2003, **301**:1496-1499.
29. Fraikin N, Goormaghtigh F, Van Melderen L: **Type II toxin-antitoxin systems: evolution and revolutions**. *J Bacteriol* 2020, **202**.
30. Sarpong DD, Murphy ER: **RNA regulated toxin-antitoxin systems in pathogenic bacteria**. *Front Cell Infect Microbiol* 2021, **11**:661026.
31. Fozo EM, Makarova KS, Shabalina SA, Yutin N, Koonin EV, Storz G: **Abundance of type I toxin-antitoxin systems in bacteria: searches for new candidates and discovery of novel families**. *Nucleic Acids Res* 2010, **38**:3743-3759.
32. Sebahia M, Wren BW, Mullany P, Fairweather NF, Minton N, Stabler R, Thomson NR, Roberts AP, Cerdeno-Tarraga AM, Wang H *et al.*: **The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome**. *Nat Genet* 2006, **38**:779-786.
33. Garneau JR, Sekulovic O, Dupuy B, Soutourina O, Monot M, Fortier LC: **High prevalence and genetic diversity of large phiCD211 (phiCDIF1296T)-like prophages in *Clostridioides difficile***. *Appl Environ Microbiol* 2018, **84**.
34. Boudry P, Semenova E, Monot M, Datsenko KA, Lopatina A, Sekulovic O, Ospina-Bedoya M, Fortier LC, Severinov K, Dupuy B *et al.*: **Function of the CRISPR-Cas system of the human pathogen *Clostridium difficile***. *mBio* 2015, **6**:e01112-e01115.
35. Riedel T, Wittmann J, Bunk B, Schober I, Sproer C, Gronow S, Overmann J: **A *Clostridioides difficile* bacteriophage genome encodes functional binary toxin-associated genes**. *J Biotechnol* 2017, **250**:23-28.
36. Govind R, Vedyappan G, Rolfe RD, Dupuy B, Fralick JA: **Bacteriophage-mediated toxin gene regulation in *Clostridium difficile***. *J Virol* 2009, **83**:12037-12045.
37. Hargreaves KR, Flores CO, Lawley TD, Clokie MR: **Abundant and diverse clustered regularly interspaced short palindromic repeat spacers in *Clostridium difficile* strains and prophages target multiple phage types within this pathogen**. *mBio* 2014, **5** e01045-e01013.
38. Bronesky D, Wu Z, Marzi S, Walter P, Geissmann T, Moreau K, Vandenesch F, Caldelari I, Romby P: ***Staphylococcus aureus* RNAIII and its regulon link quorum sensing, stress responses, metabolic adaptation, and regulation of virulence gene expression**. *Annu Rev Microbiol* 2016, **70**:299-316.
39. Sekulovic O, Fortier LC: **Global transcriptional response of *Clostridium difficile* carrying the CD38 prophage**. *Appl Environ Microbiol* 2015, **81**:1364-1374.
40. Li T, Zhang Y, Dong K, Kuo CJ, Li C, Zhu YQ, Qin J, Li QT, Chang YF, Guo X *et al.*: **Isolation and characterization of the novel phage JD032 and global transcriptomic response during JD032 infection of *Clostridioides difficile* ribotype 078**. *mSystems* 2020, **5**.
41. Boudry P, Piattelli E, Drouineau E, Peltier J, Boutserin A, Lejars M,
 ● Hajnsdorf E, Monot M, Dupuy B, Martin-Verstraete I *et al.*: **Identification of RNAs bound by Hfq reveals widespread RNA partners and a sporulation regulator in the human pathogen *Clostridioides difficile***. *RNA Biol* 2021, **18**:1931-1952
 Together with Ref. [43●] this study presents a global identification of RNA partners for RNA chaperone protein Hfq in *C. difficile*, including type I TA, CRISPR, sRNAs, antisense RNAs and riboswitches.
42. Chen Y, Indurthi DC, Jones SW, Papoutsakis ET: **Small RNAs in the genus *Clostridium***. *mBio* 2011, **2** e00340-00310.
43. Fuchs M, Lamm-Schmidt V, Sulzer J, Ponath F, Jenniches L,
 ● Kirk JA, Fagan RP, Barquist L, Vogel J, Faber F: **An RNA-centric global view of *Clostridioides difficile* reveals broad activity of Hfq in a clinically important gram-positive bacterium**. *Proc Natl Acad Sci U S A* 2021, **118**
 See annotation to Ref. [41●].
44. Soutourina O: **RNA-based control mechanisms of *Clostridium difficile***. *Curr Opin Microbiol* 2017, **36**:62-68.
45. Soutourina O, Dubois T, Monot M, Shelyakin PV, Saujet L,
 ● Boudry P, Gelfand MS, Dupuy B, Martin-Verstraete I: **Genome-wide transcription start site mapping and promoter assignments to a sigma factor in the human enteropathogen *Clostridioides difficile***. *Front Microbiol* 2020, **11**:1939
 Global annotation of transcriptional start sites and associated sigma factors in *C. difficile*.
46. Soutourina OA, Monot M, Boudry P, Saujet L, Pichon C, Sismeyro O, Semenova E, Severinov K, Le Bouguenec C, Coppee JY *et al.*: **Genome-wide identification of regulatory RNAs in the human pathogen *Clostridium difficile***. *PLoS Genet* 2013, **9**:e1003493.
47. Brantl S, Bruckner R: **Small regulatory RNAs from low-GC Gram-positive bacteria**. *RNA Biol* 2014, **11**:443-456.
48. Kalvari I, Nawrocki EP, Ontiveros-Palacios N, Argasinska J, Lamkiewicz K, Marz M, Griffiths-Jones S, Toffano-Nioche C, Gautheret D, Weinberg Z *et al.*: **Rfam 14: expanded coverage of metagenomic, viral and microRNA families**. *Nucleic Acids Res* 2021, **49**:D192-D200.
49. Lamm-Schmidt V, Fuchs M, Sulzer J, Gerovac M, Hör J, Dersch P, Vogel J, Faber F: **Grad-seq identifies KhpB as a global RNA-binding protein in *Clostridioides difficile* that regulates toxin production**. *microLife* 2021, **2**.
50. Andersen JM, Shoup M, Robinson C, Britton R, Olsen KE, Barrangou R: **CRISPR diversity and microevolution in *Clostridium difficile***. *Genome Biol Evol* 2016, **8**:2841-2855.
51. Maikova A, Boudry P, Shiriaeva A, Vasileva A, Boutserin A, Medvedeva S, Semenova E, Severinov K, Soutourina O: **Protospacer adjacent motif specificity during *Clostridioides***

- difficile type I-B CRISPR-Cas interference and adaptation.** *mBio* 2021, **12**:e0213621.
52. Maikova A, Peltier J, Boudry P, Hajnsdorf E, Kint N, Monot M, Poquet I, Martin-Verstraete I, Dupuy B, Soutourina O: **Discovery of new type I toxin-antitoxin systems adjacent to CRISPR arrays in *Clostridium difficile*.** *Nucleic Acids Res* 2018, **46**:4733-4751
Discovery of the first functional antisense RNAs in *C. difficile* acting as antitoxins in type I TA systems located near CRISPR arrays.
53. Durand S, Jahn N, Condon C, Brantl S: **Type I toxin-antitoxin systems in *Bacillus subtilis*.** *RNA Biol* 2012, **9**:1491-1497.
54. Sayed N, Jousselin A, Felden B: **A cis-antisense RNA acts in trans in *Staphylococcus aureus* to control translation of a human cytolytic peptide.** *Nat Struct Mol Biol* 2011, **19**:105-112.
55. Tourasse NJ, Darfeuille F: **T1TADB: the database of type I toxin-antitoxin systems.** *RNA* 2021, **27**:1471-1481.
56. Song S, Wood TK: **A primary physiological role of toxin/antitoxin systems is phage inhibition.** *Front Microbiol* 2020, **11**:1895
A review highlighting the role of TA in phage inhibition.
57. Pecota DC, Wood TK: **Exclusion of T4 phage by the *hok/sok* killer locus from plasmid R1.** *J Bacteriol* 1996, **178**:2044-2050.
58. Guegler CK, Laub MT: **Shutoff of host transcription triggers a toxin-antitoxin system to cleave phage RNA and abort infection.** *Mol Cell* 2021, **81**:2361-2373 e2369
This work provides the evidence for the role of type III TA system in phage defense.
59. Koonin EV, Zhang F: **Coupling immunity and programmed cell suicide in prokaryotes: life-or-death choices.** *Bioessays* 2017, **39**:1-9
A review discussing the functional coupling between different defense systems in prokaryotes.
60. Makarova KS, Wolf YI, Snir S, Koonin EV: **Defense islands in bacterial and archaeal genomes and prediction of novel defense systems.** *J Bacteriol* 2011, **193**:6039-6056.
61. Songailiene I, Juozapaitis J, Tamulaitiene G, Ruksenaite A, Sulcius S, Sasnauskas G, Venclovas C, Siksnys V: **HEPN-MNT toxin-antitoxin system: the HEPN ribonuclease is neutralized by OligoAMPylation.** *Mol Cell* 2020, **80**:955-970 e957.
62. Li M, Gong L, Cheng F, Yu H, Zhao D, Wang R, Wang T, Zhang S, Zhou J, Shmakov SA *et al.*: **Toxin-antitoxin RNA pairs safeguard CRISPR-Cas systems.** *Science* 2021, **372**
Discovery of a new two-RNA TA system encoded within CRISPR-Cas locus that contributes to the maintenance of CRISPR-Cas.
63. Bastet L, Turcotte P, Wade JT, Lafontaine DA: **Maestro of regulation: riboswitches orchestrate gene expression at the levels of translation, transcription and mRNA decay.** *RNA Biol* 2018, **15**:679-682.
64. Bordeleau E, Purcell EB, Lafontaine DA, Fortier LC, Tamayo R, Burrus V: **Cyclic di-GMP riboswitch-regulated type IV pili contribute to aggregation of *Clostridium difficile*.** *J Bacteriol* 2015, **197**:819-832.
65. Sudarsan N, Lee ER, Weinberg Z, Moy RH, Kim JN, Link KH, Breaker RR: **Riboswitches in eubacteria sense the second messenger cyclic di-GMP.** *Science* 2008, **321**:411-413.
66. Tamayo R: **Cyclic diguanylate riboswitches control bacterial pathogenesis mechanisms.** *PLoS Pathog* 2019, **15**:e1007529.
67. Hengge R: **High-specificity local and global c-di-GMP signaling.** *Trends Microbiol* 2021, **29**:993-1003.
68. Bordeleau E, Fortier LC, Malouin F, Burrus V: **c-di-GMP turn-over in *Clostridium difficile* is controlled by a plethora of diguanylate cyclases and phosphodiesterases.** *PLoS Genet* 2011, **7**:e1002039.
69. Carrier MC, Lalaouna D, Masse E: **Broadening the definition of bacterial small RNAs: characteristics and mechanisms of action.** *Annu Rev Microbiol* 2018, **72**:141-161.
70. Wagner EGH, Romby P: **Small RNAs in bacteria and archaea: who they are, what they do, and how they do it.** *Adv Genet* 2015, **90**:133-208.
71. Boudry P, Gracia C, Monot M, Caillet J, Saujet L, Hajnsdorf E, Dupuy B, Martin-Verstraete I, Soutourina O: **Pleiotropic role of the RNA chaperone protein Hfq in the human pathogen *Clostridium difficile*.** *J Bacteriol* 2014, **196**:3234-3248.
72. Dugar G, Leenay RT, Eisenbart SK, Bischler T, Aul BU, Beisel CL, Sharma CM: **CRISPR RNA-dependent binding and cleavage of endogenous RNAs by the *Campylobacter jejuni* Cas9.** *Mol Cell* 2018, **69**:893-905 e897.
73. Jiao C, Sharma S, Dugar G, Peeck NL, Bischler T, Wimmer F, Yu Y, Barquist L, Schoen C, Kurzai O *et al.*: **Noncanonical crRNAs derived from host transcripts enable multiplexable RNA detection by Cas9.** *Science* 2021, **372**:941-948.
74. Westra ER, Buckling A, Fineran PC: **CRISPR-Cas systems: beyond adaptive immunity.** *Nat Rev Microbiol* 2014, **12**:317-326.
75. Sampson TR, Saroj SD, Llewellyn AC, Tzeng YL, Weiss DS: **A CRISPR/Cas system mediates bacterial innate immune evasion and virulence.** *Nature* 2013, **497**:254-257.

Publication 2



Using an Endogenous CRISPR-Cas System for Genome Editing in the Human Pathogen *Clostridium difficile*

Anna Maikova,^{a,b,c} Victor Kreis,^c Anaïs Boutserin,^c Konstantin Severinov,^{a,d}  Olga Soutourina^c

^aSkoltech Center of Life Sciences, Skolkovo Institute of Science and Technology, Moscow, Russian Federation

^bSorbonne Paris Cité, Université Paris Diderot, Paris, France

^cInstitute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris-Sud, Université Paris-Saclay, Gif-sur-Yvette, France

^dWaksman Institute for Microbiology, Rutgers, The State University of New Jersey, Piscataway, New Jersey, USA

ABSTRACT The human enteropathogen *Clostridium difficile* constitutes a key public health issue in industrialized countries. Many aspects of *C. difficile* pathophysiology and adaptation inside the host remain poorly understood. We have recently reported that this bacterium possesses an active CRISPR-Cas system of subtype I-B for defense against phages and other mobile genetic elements that could contribute to its success during infection. In this paper, we demonstrate that redirecting this endogenous CRISPR-Cas system toward autoimmunity allows efficient genome editing in *C. difficile*. We provide a detailed description of this newly developed approach and show, as a proof of principle, its efficient application for deletion of a specific gene in reference strain 630 Δ erm and in epidemic *C. difficile* strain R20291. The new method expands the arsenal of the currently limiting set of gene engineering tools available for investigation of *C. difficile* and may serve as the basis for new strategies to control *C. difficile* infections.

IMPORTANCE *Clostridium difficile* represents today a real danger for human and animal health. It is the leading cause of diarrhea associated with health care in adults in industrialized countries. The incidence of these infections continues to increase, and this trend is accentuated by the general aging of the population. Many questions about the mechanisms contributing to *C. difficile*'s success inside the host remain unanswered. The set of genetic tools available for this pathogen is limited, and new developments are badly needed. *C. difficile* has developed efficient defense systems that are directed against foreign DNA and that could contribute to its survival in phage-rich gut communities. We show how one such defense system, named CRISPR-Cas, can be hijacked for *C. difficile* genome editing. Our results also show a great potential for the use of the CRISPR-Cas system for the development of new therapeutic strategies against *C. difficile* infections.

KEYWORDS *Clostridium difficile*, CRISPR, endogenous subtype I-B CRISPR-Cas system, genome editing

The strictly anaerobic spore-forming bacterium *Clostridium difficile* (novel name, *Clostridioides difficile* [1]) is one of the major nosocomial pathogenic clostridia. This enteropathogen causes the majority of cases of antibiotic therapy-associated diarrhea and can lead to pseudomembranous colitis, a potentially lethal disease (2, 3). Over the last few decades, *C. difficile* infections have become one of the most important public health problems due to the emergence of hypervirulent strains (such as the PCR ribotype 027 R20291 strain) (4) and the increased incidence of *C. difficile* antibiotic resistance (5). The disruption of the colonic microflora caused by antibiotic therapy allows *C. difficile* to colonize the intestinal tract after the germination of preexisting or acquired spores (2, 6). Following gut colonization, *C. difficile* produces one or both of

Citation Maikova A, Kreis V, Boutserin A, Severinov K, Soutourina O. 2019. Using an endogenous CRISPR-Cas system for genome editing in the human pathogen *Clostridium difficile*. *Appl Environ Microbiol* 85:e01416-19. <https://doi.org/10.1128/AEM.01416-19>.

Editor Haruyuki Atomi, Kyoto University

Copyright © 2019 American Society for Microbiology. All Rights Reserved.

Address correspondence to Olga Soutourina, olga.soutourina@i2bc.paris-saclay.fr.

Received 24 June 2019

Accepted 6 August 2019

Accepted manuscript posted online 9 August 2019

Published 1 October 2019

the large toxins TcdA and TcdB. These toxins trigger alterations in the intestinal cell cytoskeleton, resulting in cell lysis and inflammation (3, 7). Many aspects of *C. difficile* pathogenesis, including the molecular mechanisms of the infection cycle, remain poorly understood. Therefore, it is important to develop new genome editing approaches for further investigations of this emerging human pathogen.

CRISPR (clustered regularly interspaced short palindromic repeat)-Cas (CRISPR-associated) systems protect bacteria and archaea from phages and other mobile genetic elements (8). These adaptive immunity systems are highly diverse (9) and have been discovered in half of the sequenced bacterial genomes and in almost all archaeal genomes (10). CRISPR-Cas systems comprise CRISPR arrays and *cas* gene operons. CRISPR arrays are arranged into short direct repeats (20 to 40 bp) separated by variable spacers. Some spacers are complementary to protospacers, which are sequences within phage and other mobile genetic element genomes (11). CRISPR arrays are transcribed from promoters localized in leader regions into long pre-CRISPR RNAs (pre-crRNAs). Pre-crRNAs are processed into small protective CRISPR RNAs (crRNAs). In complex with Cas proteins, crRNAs serve as guides to recognize and direct the cleavage of foreign genetic elements by Cas nucleases in a process known as “interference” (12). New spacers are acquired into CRISPR arrays from foreign genomes during the adaptation process (13).

For many CRISPR-Cas systems, an important component of immunity mechanism is a protospacer-adjacent motif (PAM). PAMs are short sequences located on the 3' or 5' end of the protospacer. PAMs are necessary for protospacer recognition, and they are absent in CRISPR arrays; this allows avoidance of autoimmunity (8).

According to a recent classification based on the *cas* genes involved in interference, the CRISPR-Cas systems are divided into two classes and are further subdivided into six types and 33 subtypes (9). Class 1 includes type I, III, and IV CRISPR-Cas systems, which are characterized by multisubunit effector complexes, while class 2 includes type II, V, and VI CRISPR-Cas systems, which carry single-protein effectors. Recent studies showed that *C. difficile* strains possess an active subtype I-B CRISPR-Cas system (14–17). The *C. difficile* CRISPR-Cas system is characterized by an unusually high number of CRISPR arrays (on average, 8.5 CRISPR arrays per genome, with some arrays being localized in prophages) (16) and the presence of two or three *cas* operons belonging to the same subtype (15, 17). In our previous studies, we demonstrated active expression of all CRISPR arrays for the *C. difficile* 630 and R20291 strains, as well as the ability of *C. difficile* 630 to mount robust CRISPR interference (14, 15). We also bioinformatically predicted and experimentally validated *C. difficile* CRISPR-Cas PAMs (15).

During the last few years, substantial efforts have been concentrated on the development of various CRISPR-based biotechnological tools (18). In particular, the type II Cas9- and type V Cpf1 (Cas12a)-based technologies are widely used for genome editing in different organisms (19, 20). Nevertheless, the application of other types of CRISPR-Cas systems has also attracted the attention of the scientific community. Harnessing of endogenous CRISPR-Cas systems for genome editing in bacteria and archaea appears to be a particularly attractive strategy (18, 21). This approach is based on the use of plasmid vectors containing artificial CRISPR miniarrays with spacers targeting a chromosomal gene (21). crRNAs expressed from a plasmid-borne miniarray utilize the endogenous Cas machinery to form an effector complex which recognizes the protospacer of choice, leading to its cleavage. Destruction of chromosomal DNA leads to the killing of wild-type cells (Fig. 1A). An editing plasmid with sequences homologous to sequences flanking the protospacer triggers homologous recombination and allelic exchange with the targeted chromosomal region (Fig. 1B). This results in elimination of the wild-type allele and preservation of chromosomal mutants since they no longer possess the targeted protospacer (Fig. 1B). The endogenous CRISPR-based method is often easier to set up for editing in prokaryotes than the CRISPR-Cas9 and CRISPR-Cpf1 (Cas12a) technologies. Another advantage of this approach is that there is no need to heterologously express potentially toxic Cas proteins inside bacterial or archaeal cells. The genome editing approach based on an endogenous CRISPR-

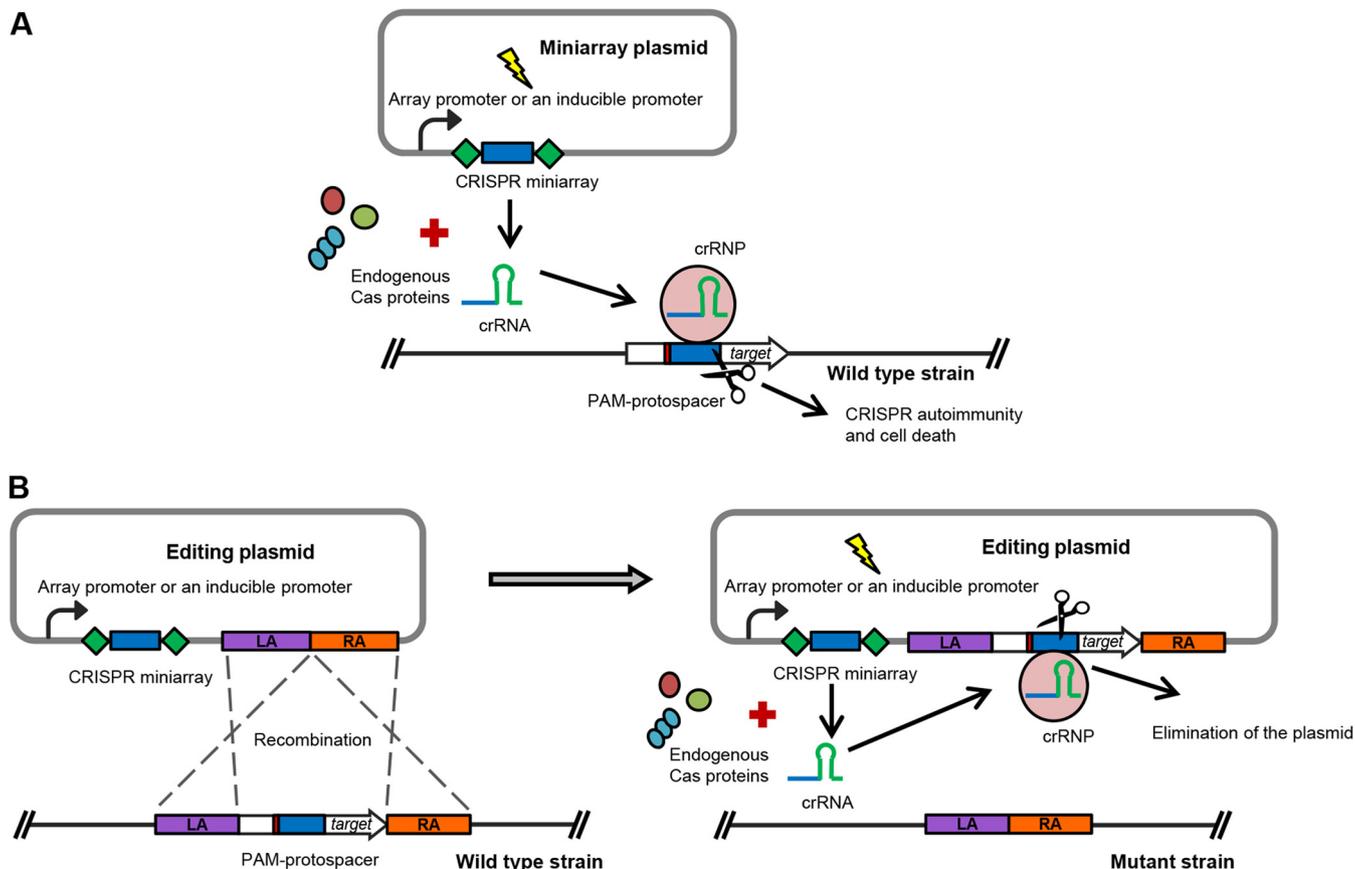


FIG 1 General scheme of using endogenous CRISPR-Cas systems for genome editing in bacteria and archaea. (A) The crRNA is expressed from a vector-borne CRISPR miniarray under the control of native or inducible promoters. The crRNA forms a ribonucleoprotein (crRNP) complex with endogenous Cas proteins, which recognizes and directs the cleavage of the PAM-associated protospacer, localized at the target chromosomal region. This leads to chromosome disruption and cell death. (B) An editing plasmid, additionally carrying homologous arms (the left arm [LA] and the right arm [RA]), allows recombination between the plasmid and the chromosome to occur before CRISPR interference. The crRNP targets the PAM-protospacer on the plasmid, which leads to the elimination of the plasmid and preservation of the chromosomal mutants.

Cas system was successfully applied in several prokaryotic organisms using CRISPR-Cas subtype I-A and III-B or subtype I-B in archaea, *Sulfolobus islandicus* (21) and *Haloarcula hispanica* (22), respectively, and subtype I-B in several clostridial species, *C. pasteurianum* (23), *C. tyrobutyricum* (24), and *C. saccharoperbutylacetonicum* (25).

In *C. difficile*, various genetic tools for genome manipulation have been established. One of the most widely used methods is the ClosTron technology, based on mobile altered type II introns and the utilization of retrotransposable activated markers (RAM) (26, 27). Though this genome editing technique allows targeting of almost any chromosomal region and RAM enable one to easily identify potential mutants, the method has some disadvantages. Most importantly, ClosTron generates insertion mutations that may cause polar effects on downstream genes. An additional limitation comes from difficulties in finding an efficient insertion site within genes of a small size.

Another popular *C. difficile* genome editing approach is the allele-coupled exchange technique, based on a semisuicidal plasmid vector carrying the *Escherichia coli* cytosine deaminase gene (*codA*) or the *C. difficile* orotate phosphoribosyltransferase gene (*pyrE*) as a counterselection marker (28, 29). This method includes a two-step recombination event between the editing plasmid and the genome and the selection of double-crossing-over clones that lost the plasmid on nutrient-poor medium supplemented with 5-fluorocytosine (for *codA*-based plasmids) or fluoroorotic acid (for the *pyrE* allelic exchange system). The counterselection procedure is based on the generation of highly toxic compounds from these substrates. Despite the fact that this approach allows the creation of *C. difficile* mutants carrying point mutations, deletions, and insertions, it can

be difficult to apply in some cases. First, mutations that result in a growth deficiency phenotype or the inactivation of metabolic genes may affect growth on nutrient-poor medium. Second, there are some difficulties with losing the editing plasmids in mutant strains after editing, which could lead to the spontaneous creation of revertant strains.

Recently, a method based on the DNA double-strand breaks in *C. difficile* has been reported (30). This technology uses site-specific cleavage by the *Saccharomyces cerevisiae* yeast homing endonuclease I-SceI, whose recognition site is introduced in the editing plasmid vector. After the integration of the editing vector into the chromosome, another vector containing the I-SceI endonuclease gene under the control of a constitutive promoter is transferred to the single-crossing-over integrants to induce double-strand breaks and genome editing via homologous recombination. The advantage of this method is the possibility to create markerless deletions and the fast loss of the vector. Nevertheless, this method includes time-consuming two-step conjugations and the expression of I-SceI endonuclease, which could induce side effects.

During the past few years, the successful application of CRISPR-Cas9 and CRISPR-Cpf1 (Cas12a) for genome editing in *C. difficile* has been reported (31–34). These approaches have enhanced the possibilities of genetic manipulation in *C. difficile* and have proven to be efficient. However, the Cas9 and Cpf1 technologies require the design of plasmids harboring specific single guide RNAs (sgRNAs), and the editing plasmid is not automatically cured after the editing is complete. The use of an endogenous CRISPR-Cas system can enhance the possibilities of the genetic manipulation of *C. difficile*. The present work describes the utilization of a native *C. difficile* subtype I-B CRISPR-Cas system to generate deletion mutants in the 630 Δ *erm* and R20291 strains.

To evaluate the possibility of using an endogenous *C. difficile* CRISPR-Cas system for the targeting of specific sequences on the bacterial chromosome, we have chosen the *hfq* gene. Hfq is a bacterial RNA-binding protein that plays major roles in RNA metabolism and the global posttranscriptional network, in particular, in Gram-negative bacteria (35). The study of Hfq depletion in *C. difficile* 630 Δ *erm* (36) suggested a pleiotropic role of this protein in *C. difficile* physiology, with the most pronounced effect being on sporulation. The availability of an *hfq* deletion mutant would open new perspectives for further characterization of its role in RNA-based regulation in *C. difficile*. Our previous attempts to inactivate the *hfq* gene using a Clostron gene knockout system were unsuccessful (36). Additionally, we have tried to delete *hfq* using the *codA* allelic exchange approach (28, 29), but also without success (data not shown).

RESULTS

Construction of targeting miniarray plasmids and verification of their functionality. The general strategy for the construction of functional editing plasmids pECrFA_hfq630 and pECrPA_hfqR20291 for use in the 630 Δ *erm* and R20291 strains, respectively, is shown in Fig. 2. We first constructed two CRISPR miniarray plasmids targeting the *hfq* gene (pECrF_hfq and pECrP_hfq). The miniarray was based on the *C. difficile* 630 Δ *erm* CRISPR 16 array, which is highly expressed and capable of interference (15). Two variants of the leader sequence upstream of the miniarray were used (see Fig. S1A and B in the supplemental material): the full leader (a 403-bp sequence upstream of the first direct repeat of the CRISPR 16 array) containing all native promoters, which should allow autonomous expression of the miniarray (pECrF_hfq), and a partial leader (a 154-bp region upstream of the first direct repeat of the CRISPR 16 array), which lacked native promoters but which should allow the inducible expression of the miniarray from a vector-borne anhydrotetracycline (ATc)-inducible promoter (P_{tet}) (pECrP_hfq). The repeat-spacer-repeat motif of the synthetic miniarray was also based on 29-bp repeat sequences of the *C. difficile* 630 Δ *erm* CRISPR 16 array (Fig. 2A and Fig. S1A and B). For successful recognition of protospacers by the *C. difficile* CRISPR-Cas system, a functional PAM-flanking protospacer at the 5' end is necessary (15). Two functional trinucleotide PAMs of the *C. difficile* CRISPR-Cas system, 5' CCA and CCT, have been experimentally validated, and additional alternative motifs, such as CCC,

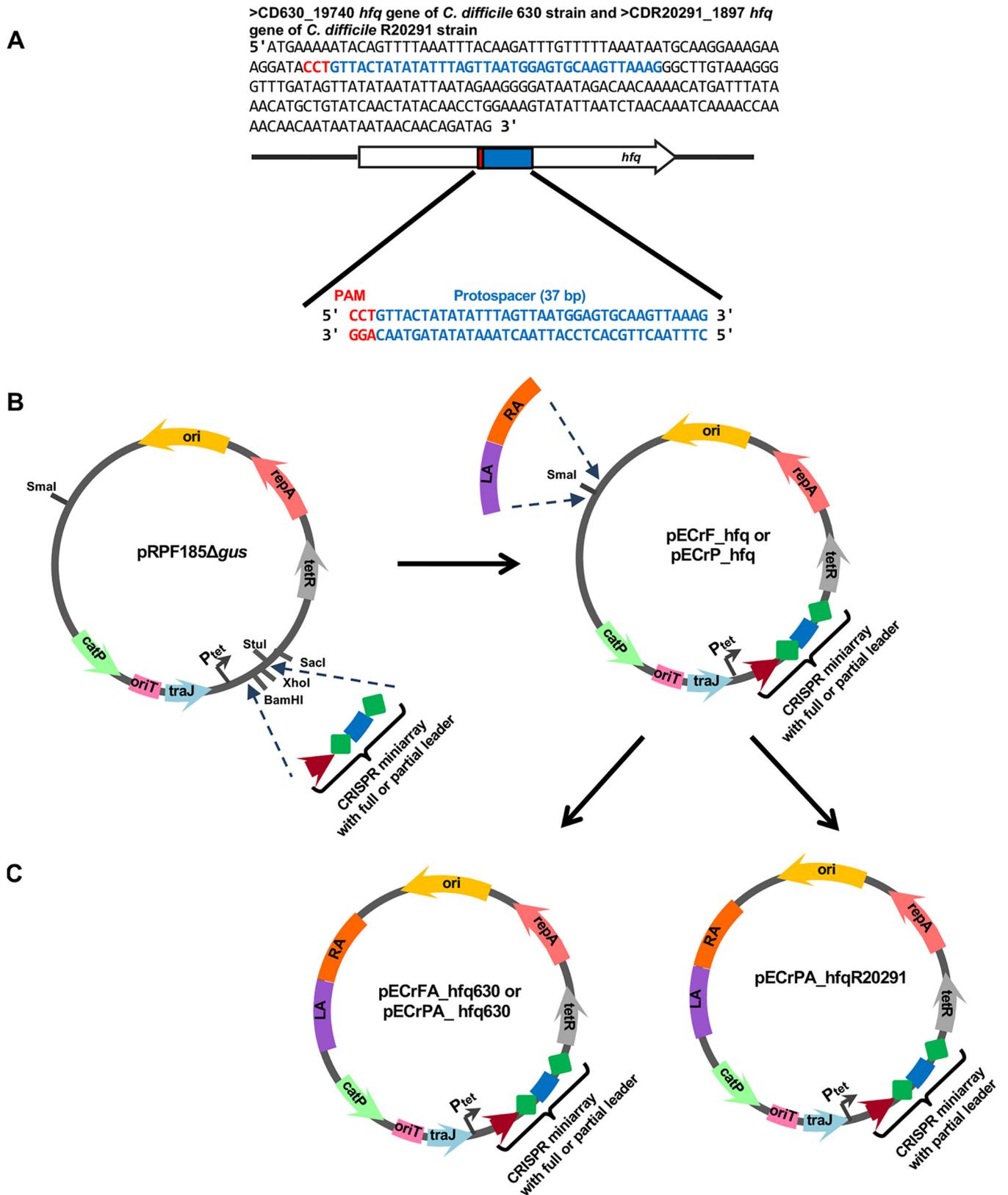


FIG 2 Strategy for the design of the editing plasmids to delete the *hfq* gene in the *C. difficile* 630 Δ erm and R20291 strains. (A) The coding sequences of the *C. difficile* 630 and R20291 *hfq* genes and a 37-bp sequence associated with the 5' CCT PAM, selected as a protospacer for the miniarray. (B) Construction of the pECrF_hfq and pECrP_hfq miniarray plasmids on the basis of the pRPF185 Δ gus vector. The miniarray sequences were cloned into the BamHI and XhoI restriction sites. (C) Construction of the pECrFA_hfq630, pECrPA_hfq630, and pECrPA_hfqR20291 editing plasmids on the basis of pECrF_hfq and pECrP_hfq. The F in the plasmid names represents the full-length leader region for

(Continued on next page)

Downloaded from https://journals.asm.org/journal/aem on 13 December 2021 by 2a01:e0a:5da:d0b0:289b:7ada:f882:8caa.

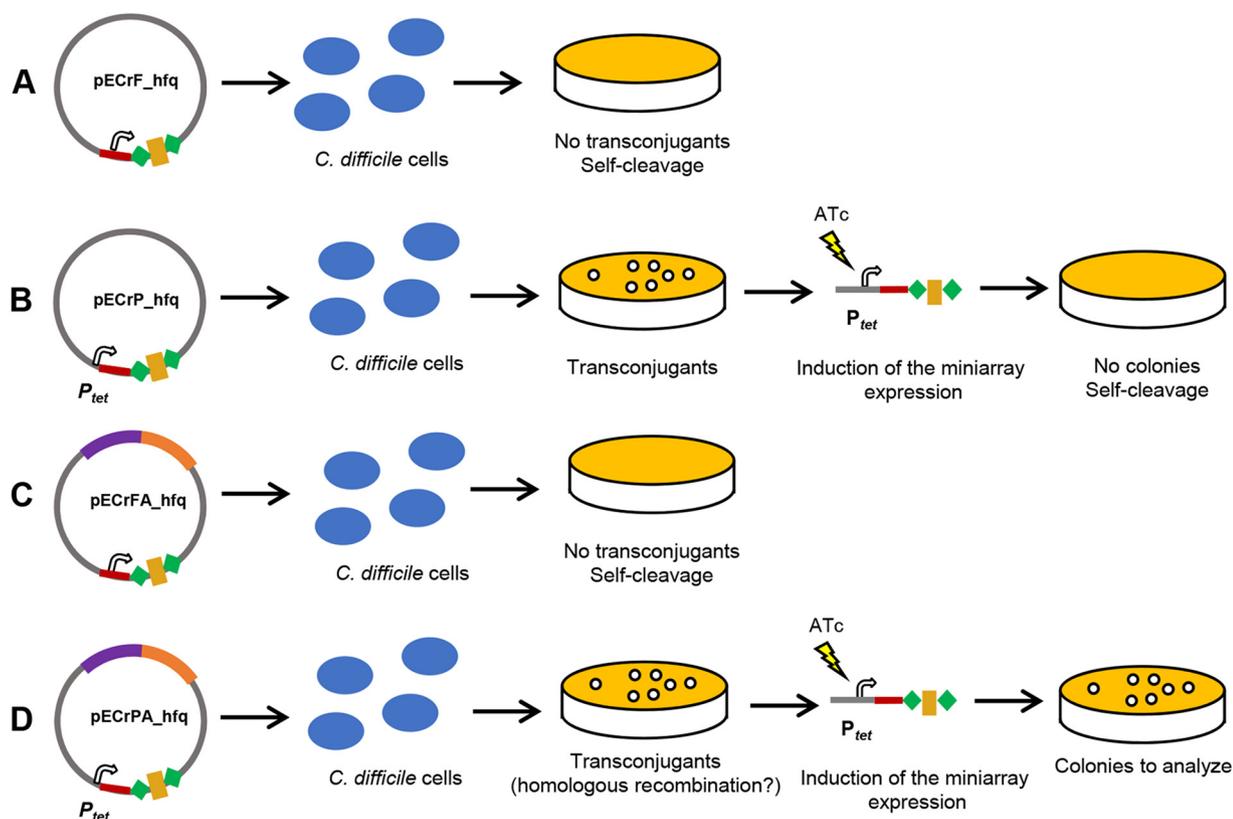


FIG 3 Different effects of the conjugation of the miniarray and editing plasmids into *C. difficile* cells. (A) CRISPR self-cleavage induced by immediate expression of the miniarray from the plasmid pECrF_hfq after conjugation. (B) CRISPR self-cleavage resulted from the ATc-induced expression of the miniarray from the plasmid pECrP_hfq after second plating of the transconjugants. (C) CRISPR self-cleavage induced by the immediate expression of the miniarray from the pECrFA_hfq plasmid after conjugation. (D) Homologous recombination between the chromosome and the pECrPA_hfq plasmid and cleavage of the plasmid resulted from the ATc-induced expression of the miniarray from the plasmid after the second plating of the transconjugants. The effects were tested in the 630 Δ erm strain (B and C), and the effects were tested and gene deletion was performed in both the 630 Δ erm and R20291 strains (A and D).

CCG, and TCA, have been predicted (15). The coding region of the *hfq* gene possesses at least three functional CCW motifs and two alternative TCA motifs. A 37-bp sequence inside the *hfq* gene sequence associated with the 5' CCT PAM was chosen (Fig. 2A; the mean length of the *C. difficile* spacers is 37 bp). The pECrF_hfq and pECrP_hfq plasmids (Fig. 2B) were conjugated to *C. difficile* 630 Δ erm cells using the heat shock method to ensure the highest conjugation efficiency (37). No transconjugants were obtained after conjugation of the pECrF_hfq plasmid in *C. difficile* 630 Δ erm, suggesting CRISPR autoimmunity due to self-targeting (Fig. 3A). The conjugation efficiency of 380 transconjugants/ml was observed after conjugation with pECrP_hfq (approximately 1.9×10^{-6} transconjugants/donor or recipient cell). A control conjugation with the pRPF185 Δ gus vector resulted in 5,480 transconjugants/ml (approximately 27.4×10^{-6} transconjugants/donor or recipient cell). The smaller number of transconjugants in the pECrF_hfq conjugation reaction could be due to P_{tet} promoter leakage leading to autoimmunity caused by self-cleavage in some transconjugants. To check for the efficiency of self-targeting by crRNA expressed from the pECrP_hfq plasmid, eight transconjugant colonies were restreaked on brain heart infusion (BHI) agar plates supplemented with

FIG 2 Legend (Continued)

autonomous expression of the miniarray under the control of native promoters, while the P points out the presence of a partial leader region without native promoters for miniarray expression under the control of an inducible P_{tet} promoter. The presence of homologous arms for recombination within the 630 Δ erm or R20291 strain is indicated by A and the strain name. The pECrFA_hfq630 plasmid carrying the miniarray with the full-length leader region was not efficient for gene deletion in the 630 Δ erm strain; in contrast, pECrPA_hfq630 and pECrPA_hfqR20291 were efficiently used for *hfq* gene deletion in the 630 Δ erm and R20291 strains, respectively.

500 ng/ml ATc to fully induce the expression of the miniarray. No growth was observed on these plates, indicating highly efficient self-targeting by the induced miniarray (Fig. 3B). The same effects were observed after conjugation of the pECrF_hfq plasmid in *C. difficile* R20291 cells, suggesting that the synthetic array based on the *C. difficile* 630 Δ erm CRISPR 16 leader and repeat sequences mimics well native subtype I-B CRISPR arrays in *C. difficile* for at least the 630 and R20291 strains. Therefore, the *C. difficile* endogenous CRISPR-Cas system can recognize and target protospacers on the bacterial chromosome using crRNAs expressed from a plasmid-borne artificial miniarray, and this feature can be utilized for genome editing.

Construction of the genome editing plasmid and deletion of the *hfq* gene of *C. difficile* 630 Δ erm and R20291. We first assessed which miniarray plasmid, pECrF_hfq or pECrP_hfq, is best for *C. difficile* genome manipulation. Approximately 1,200-bp-long regions flanking the *hfq* gene of the 630 Δ erm strain (Fig. S1C) were amplified by PCR and introduced into the SmaI restriction sites of pECrF_hfq or pECrP_hfq using the Gibson assembly reaction (Fig. 2C). No transconjugants were obtained after conjugation of *C. difficile* 630 Δ erm with pECrFA_hfq630, carrying the miniarray with the full-length leader region (Fig. 3C). This may mean that the CRISPR-induced autoimmune degradation of DNA around the targeted protospacer is more efficient than homologous recombination between the chromosome and the homologous region of pECrFA_hfq630. Whatever the reason, the plasmid with the full-length CRISPR array leader sequence is clearly not suitable for genome editing. After conjugation with the pECrPA_hfq630 plasmid carrying the miniarray under the control of the inducible P_{tet} promoter, about 460 transconjugants/ml (approximately 2.3×10^{-6} transconjugants/donor or recipient cell) were obtained. To induce expression of the *hfq*-targeting miniarray, 10 transconjugants were restreaked on BHI agar supplemented with 500 ng/ml ATc. We observed the growth of each transconjugant tested, suggesting that homologous recombination between the chromosome and plasmid had occurred (Fig. 3D) or that CRISPR interference was not efficient. One clone from each plate was then restreaked on BHI plates with or without thiamphenicol (Tm) to check for plasmid loss. Three out of 10 clones lost the plasmid. When analyzed by PCR, these clones turned out to be Δ *hfq* mutants (Fig. 4A). The experiment was independently repeated at least three times. In all cases, when testing 6 to 10 clones, the mutant strains could be reproducibly obtained with an overall efficiency varying from 30% to 100%. Thus, a plasmid containing an inducibly transcribed CRISPR miniarray and arms for homologous recombination at the targeted protospacer allows efficient genome editing in *C. difficile*.

The coding region of the *hfq* gene of the *C. difficile* R20291 strain is identical to that of the 630 Δ erm strain, but the flanking sequences are different. Therefore, to delete the R20291 *hfq* gene, we constructed the pECrPA_hfqR20291 plasmid on the basis of the pECrP_hfq miniarray plasmid with homologous arms of R20291 *hfq* flanking sequences (Fig. 2C and Fig. S1D). Nine out of 10 selected transconjugants had lost the plasmid, and PCR analysis showed that seven out of nine clones without the plasmid were Δ *hfq* mutants (Fig. 4A).

Validation and complementation of *hfq* deletion strains. To validate the *hfq* deletion, we assessed *hfq* mRNA expression in the wild-type and Δ *hfq* mutant strains carrying an empty pRPF185 Δ gus vector (the wt/p and Δ *hfq*/p strains, respectively) as well as in complemented *C. difficile* Δ *hfq* strain Δ *hfq*/p-*hfq* expressing plasmid-borne *hfq*. Quantitative reverse transcription-PCR (qRT-PCR) analysis confirmed the absence of *hfq* expression in the *C. difficile* 630 Δ erm Δ *hfq* and R20291 Δ *hfq* strains and the presence of the transcript in the wild-type strains (Fig. 4B). A high 400- to 500-fold increase in *hfq* mRNA abundance compared to that in the wild type was detected in complemented strains due to strong P_{tet} induction in the presence of ATc (Fig. 4B). Western blotting with polyclonal anti-Hfq antibodies confirmed the lack of the Hfq protein in the Δ *hfq*/p strains (Fig. 4C; InstantBlue dye-stained protein gels, used as loading controls, are shown in Fig. S2).

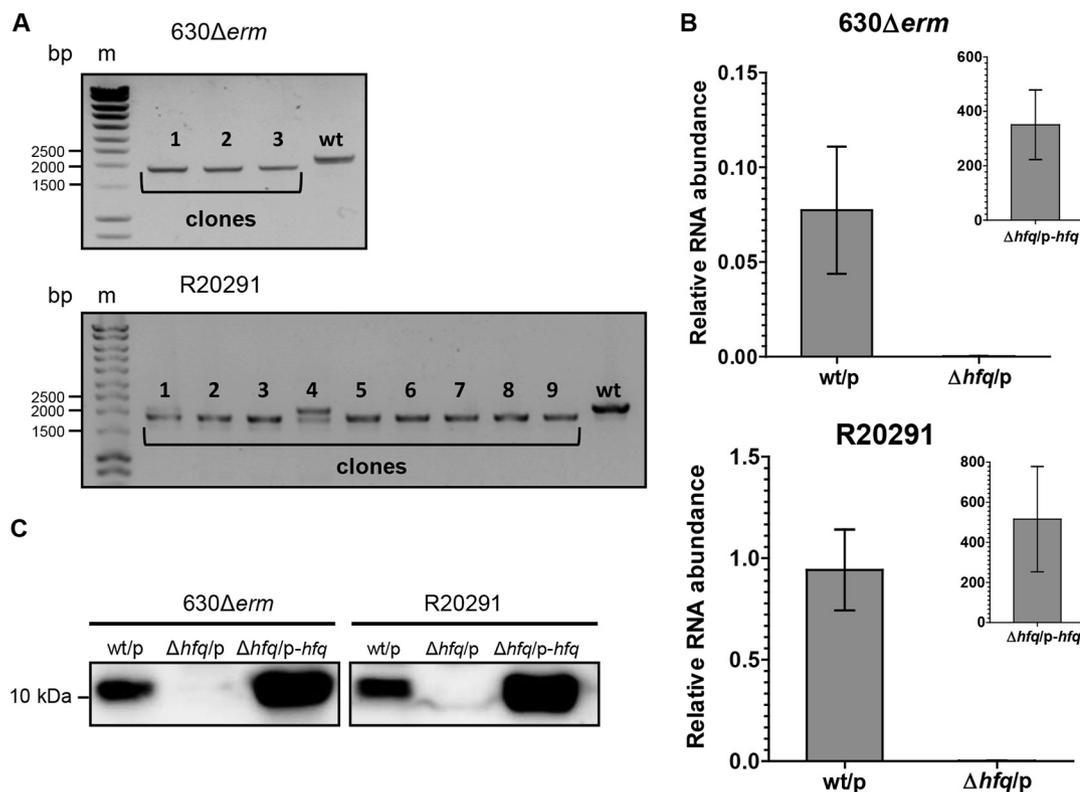


FIG 4 Validation of *hfq* deletion mutants. (A) PCR analysis of the *C. difficile* clones which lost the plasmid after genome editing. The 2,151-bp PCR bands correspond to the wild-type genotype; the 1,893-bp PCR bands correspond to the mutant genotype. For the R20291 strain, both the wild-type and mutant copies were detected with clone 4 (lane 4); this clone was discarded from further analysis. Lanes m, molecular mass markers. (B) qRT-PCR analysis of the wild-type and *Δhfq* mutant strains carrying an empty pRPF185Δ*gus* (the wt/p and *Δhfq*/p strains, respectively) and the complemented *Δhfq* *C. difficile* strain (the *Δhfq*/p-*hfq* strain). mRNA levels are relative to those of 16S rRNA. (C) Western blot analysis of the wt/p, *Δhfq*-p, and *Δhfq*/p-*hfq* *C. difficile* strains. As loading controls, InstantBlue dye-stained protein gels were used (see Fig. S2 in the supplemental material).

Sporulation assay of *C. difficile* 630Δ*erm* *Δhfq* mutants. Sporulation represents one of the crucial features of *C. difficile* as a successful pathogen. In our previous work, we revealed that the Hfq protein likely controls the sporulation rates in *C. difficile* 630Δ*erm*-derived strains (36). The Hfq-depleted strain demonstrated higher levels of sporulation than the control strain. To analyze the effect of the *hfq* gene deletion on this phenotype, we compared the sporulation rates in the 630Δ*erm* wt/p, *Δhfq*/p, and *Δhfq*/p-*hfq* strains. After 24 h and 48 h in BHIS medium supplemented with Tm and ATc, the mutant strain (the *Δhfq*/p strain) demonstrated a higher level of sporulation than the wild-type strain (the wt/p strain) (Fig. 5). In addition, the complemented strain (the *Δhfq*/p-*hfq* strain) showed a reversion of sporulation efficiency to a level close to that seen in the wild type (Fig. 5). Thus, these results are consistent with previously obtained data and confirm the potential involvement of the Hfq protein in the control of sporulation in *C. difficile* (36).

DISCUSSION

Over the last decade, the rapid development of various biotechnological tools based on prokaryotic adaptive immune CRISPR-Cas systems has occurred (18). In addition to the most popular CRISPR tools, based on class 2 Cas9 and Cpf1 (Cas12a) proteins (19, 20), other CRISPR-Cas systems are also being actively explored for genetic manipulation purposes. One of the most promising applications is the use of endogenous CRISPR-Cas systems for genome editing and engineering of bacteria and archaea (18, 21).

In the present work, we utilized the endogenous CRISPR-Cas system for genome editing of enteropathogenic *C. difficile*. Although other techniques for genome manipulation in this bacterium are available (26–34), they present some limitations in their

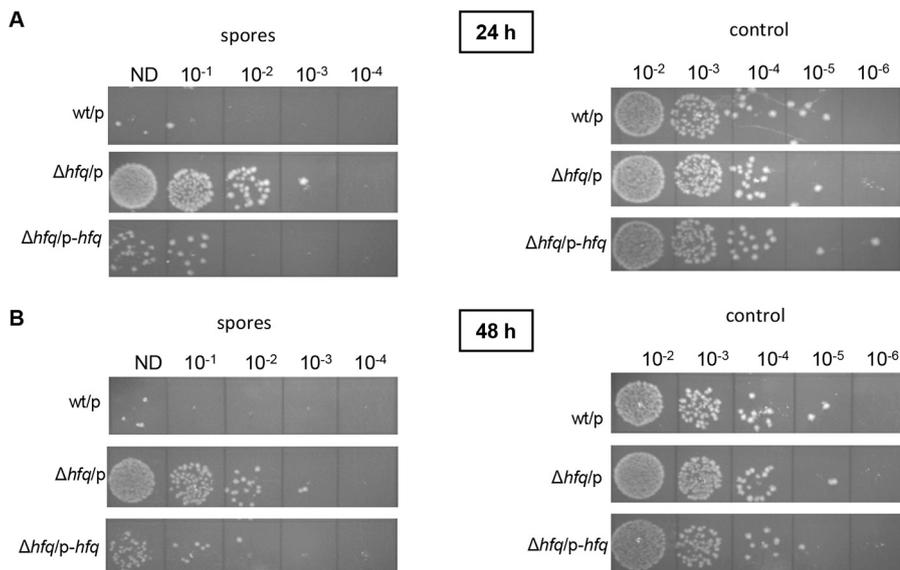


FIG 5 Sporulation levels in the *C. difficile* 630Δerm wt/p, Δhfq/p, and Δhfq/p-hfq strains (as numbers of spores) and the total amount of bacteria (as numbers of CFU) (control) after 24 h (A) and 48 h (B) of growth in BHIS supplemented with Tm and ATc. The serial dilutions of the cultures spotted on BHI plates supplemented with taurocholate are indicated (ND, not diluted). Spore samples were heated to kill all cells other than spores, while the control samples were not heated to estimate the total amount of bacteria.

applications. Harnessing the native subtype I-B CRISPR-Cas system for genome editing in *C. difficile* allowed us to create deletion mutants of the *hfq* gene, encoding the RNA chaperone Hfq. Attempts to inactivate this gene using other approaches, including the CloStron technology (36) and *codA* allelic exchange, were not successful (data not shown). Though a strain depleted of Hfq by expression of antisense RNA is available, the construction of the *hfq* deletion mutant opens up interesting possibilities for future studies of the regulatory role of Hfq and its RNA network in *C. difficile*.

The general work flow for the application of native CRISPR-Cas genome editing method in *C. difficile* is presented in Fig. 6. To repurpose the endogenous CRISPR-Cas system for deletion of the *hfq* gene, we designed plasmid vectors carrying targeting miniarray and editing plasmids carrying, in addition, homologous arms for recombination (Fig. 2B and C). The *C. difficile* 630Δerm CRISPR 16 array was chosen as a basis for synthetic miniarray construction, since it is functional for interference (15). The repeat-spacer-repeat motif for the artificial miniarray was composed of 29-bp repeat sequences and a 37-bp spacer sequence associated with a functional 5' CCT PAM inside the *hfq* gene coding region. To facilitate the genome editing procedure, we used the pECrPA_hfq630 plasmid containing the miniarray under the control of the inducible P_{tet} promoter. This strategy allowed us to successfully generate *hfq* deletion mutants in both *C. difficile* 630Δerm and epidemic R20291 strains. The CRISPR repeats in the 630Δerm and R20291 strains have similar consensus sequences (15). Moreover, both strains possess homologous complete and partial subtype I-B *cas* operons also present in the majority of sequenced *C. difficile* strains (15). The Cas machineries of the R20291 strain (and, by extension, those of other *C. difficile* isolates) can successfully recognize and utilize crRNAs expressed from a 630Δerm-based miniarray. Thus, the artificial miniarray designed from the *C. difficile* 630Δerm CRISPR 16 leader and repeat sequences is suitable for targeting specific chromosomal protospacer sequences and can be used for genome editing in at least two *C. difficile* strains. The general conservation of subtype I-B *cas* operons in *C. difficile* makes it likely that the same targeting arrays will be suitable for the majority of *C. difficile* strains, though this conjecture remains to be experimentally verified.

Repurposing of native CRISPR-Cas systems for genome editing in *C. difficile* has considerable advantages over other techniques applied to this bacterium. First of all,

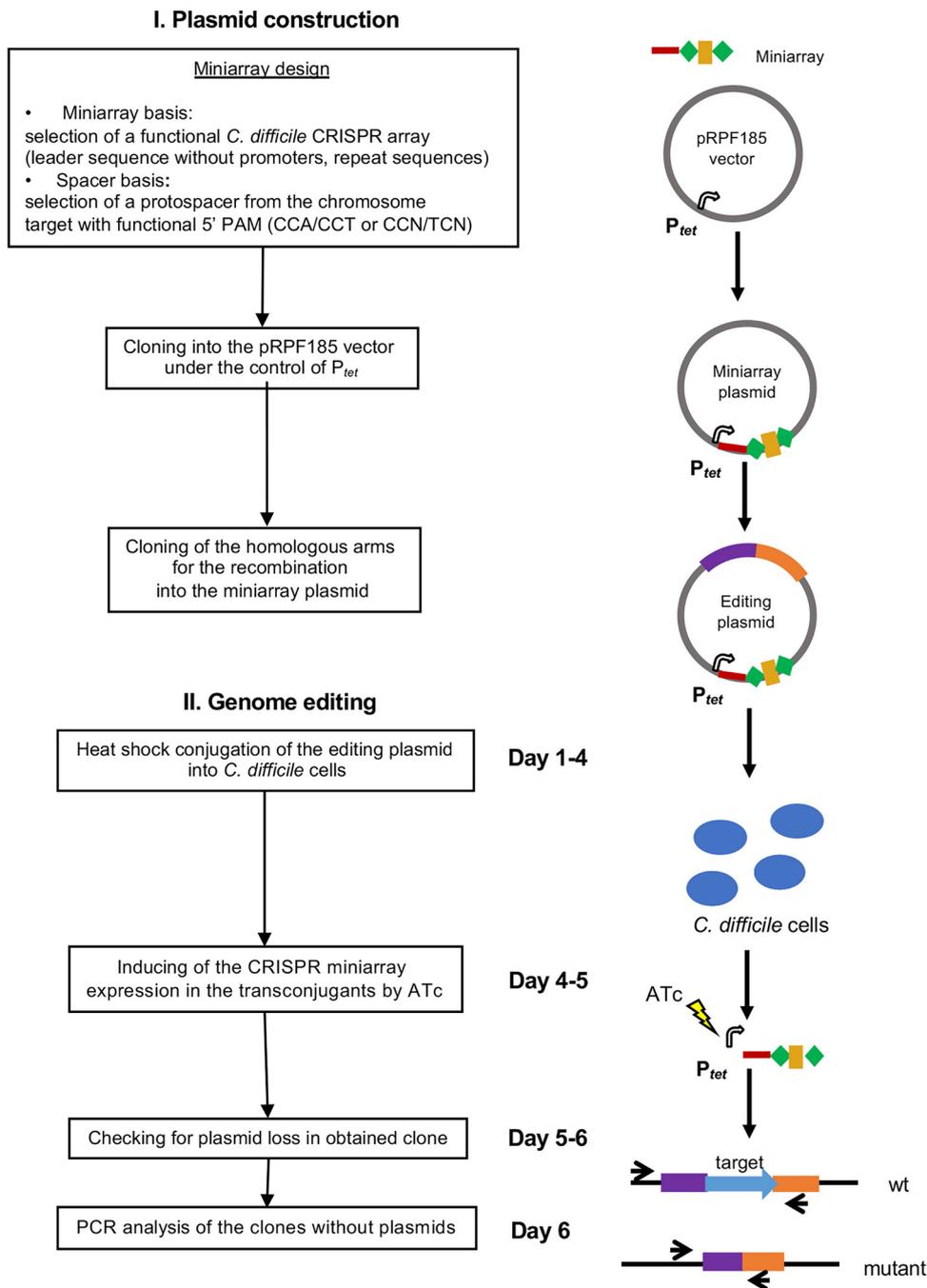


FIG 6 General work flow for application of endogenous CRISPR-Cas-based genome editing method in *C. difficile*.

this method does not require the expression of heterologous proteins inside *C. difficile* cells, which may have toxic or other unpredictable effects. A miniarray localized on an editing plasmid mimics the natural *C. difficile* CRISPR array and should not have an undesirable impact during genome manipulation. Second, this approach includes only one conjugation round and fewer plating steps, giving significant time savings (Fig. 6). For example, the *codA* allelic exchange method requires at least three more colony plating steps than the method with the miniarray editing plasmid, increasing the time needed to complete the editing experiment by at least 3 days. Finally, the miniarray editing plasmid is readily lost after the editing process, preventing the spontaneous emergence of revertants.

Among the possible challenges for the application of the method could be the choice of the best protospacer on the target genome region. The presence of a functional PAM upstream of the protospacer is imperative for successful targeting. For this reason, the choice of the genome sequence for editing should be guided by the availability of PAMs. For the moment, only two PAMs (CCA and CCT) have been experimentally confirmed for *C. difficile* CRISPR-Cas target recognition (15). At the same time, general *in silico* analysis of CRISPR spacer homology to phage protospacers revealed a rather unconstrained PAM consensus CCN/TCN for the *C. difficile* CRISPR-Cas system (15). These data increase the possibilities of target sequence selection. In addition, type I CRISPR-Cas systems can recognize protospacers on both strands of the target DNA, which expands the opportunities to find functional PAMs in the target region (21).

The applications of endogenous CRISPR-Cas system for genome editing in *C. difficile* could be potentially larger than those for the generation of deletion mutants. This technique could be readily applied for introducing other types of mutations, i.e., point mutations and insertions (21). For a point mutation, the homologous arms on the editing plasmid could be designed to introduce changes in the functional PAMs at the editing region to a nonfunctional motif. Alternatively, substitutions could be introduced into a seed region, the first 8 nucleotides of the protospacer, crucial for CRISPR targeting (38). As a priority choice, a point mutation design could be achieved by introducing changes at the first or second position of PAMs. Combining the changes within PAMs and the seed region could even increase the efficiency of editing, as reported for other endogenous CRISPR editing tools (21, 25). We have previously shown that a nonfunctional PAM and mutation in the first position of protospacer within the seed region abolished or considerably impaired CRISPR interference (15). Genome insertions designed to make a break in the integrity of the chosen protospacer or/and PAM of the targeted genome sequence (21) or to insert a mutation to knock out the PAM (25) could be introduced by the homologous arms.

The role of essential genes cannot be easily investigated since no deletion mutant can be generated. Therefore, the CRISPRi method (utilizing CRISPR interference), which allows repression of the expression of target genes, has recently been developed (39). This technology is primarily based on CRISPR-Cas9 systems with a mutated catalytic site of the Cas9 protein (catalytically dead Cas9 [dCas9]) (40). The dCas9-based method has already been used in *C. difficile* (41). In addition, it was shown that an *E. coli* native subtype I-E CRISPR-Cas system lacking *cas3* could be repurposed for programmable transcriptional repression (42). Furthermore, a recent study showed that the subtype I-B CRISPR-Cas system of *Haloferax volcanii* lacking the *cas3* and *cas6* genes could be used for gene repression in this archaeon (43). Altogether, these data suggest that the *C. difficile* native CRISPR-Cas system may be used for this goal, too, in a particular context. However, about 90% of the sequenced *C. difficile* strains possess two subtype I-B *cas* operons, each carrying the *cas3* nuclease gene. An additional partial *cas* operon with the *cas3* gene is present in the majority of the multilocus sequence type 3 group of *C. difficile* strains, including the PCR ribotype 027 strains (15). Thus, depending on the strain, the creation of a double- or triple-*cas3*-mutant background would be necessary to consider application of this CRISPRi method.

CRISPR self-targeting could lead to bacterial cell death. This feature of CRISPR-Cas systems can be applied for the development of new antimicrobial agents (44). Among the suggested strategies reside the use of phage particles and phagemids as vectors to deliver all the necessary autotargeting CRISPR-Cas components inside the cell of the targeted pathogen (44). In the present study, we showed the active killing of *C. difficile* cells by CRISPR self-targeting via expression of the miniarray from a plasmid vector. Therefore, in perspective, this approach could be promising for the future development of alternative strategies for the treatment of *C. difficile* infections.

In conclusion, the repurposing of the endogenous CRISPR-Cas system for genome editing in *C. difficile* extends the range of biotechnological techniques available for this enteropathogenic bacterium and could be valuable in further studies.

TABLE 1 Bacterial strains and plasmids used in this study

Strain or plasmid	Genotype or description	Source or reference
Strains		
<i>E. coli</i>		
NEB-10 beta	$\Delta(\text{ara-leu})7697 \text{ araD139 fhuA } \Delta\text{lacX74 galk16 galE15 e14 mutant } \phi 80\text{dlacZ}\Delta\text{M15 recA1 relA1 endA1 nupG rpsL (Str}^r\text{) rph spoT1 } \Delta(\text{mrr hsdRMS-mcrBC})$	New England Biolabs
HB101(RP4)	$\text{supE44 aa14 galk2 lacY1 } \Delta(\text{gpt-proA})62 \text{ rpsL20 (Str}^r\text{) xyl-5 mtl-1 recA13 } \Delta(\text{mcrC-mrr}) \text{ hsdS}_B(\text{r}_B^- \text{m}_B^-)$ RP4 (Tra ⁺ IncP Ap ^r Km ^r Tc ^r)	Laboratory stock
<i>C. difficile</i>		
630 Δerm	Sequenced reference strain, ΔermB	Laboratory stock (52)
R20291	PCR ribotype 027 epidemic strain	Laboratory stock
wt/p	630 Δerm or R20291 carrying the pRPF Δgus plasmid	This work
$\Delta\text{hfq/p}$	630 $\Delta\text{erm } \Delta\text{hfq}$ or R20291 Δhfq carrying the pRPF Δgus plasmid	This work
$\Delta\text{hfq/p-hfq}$	630 $\Delta\text{erm } \Delta\text{hfq}$ or R20291 Δhfq carrying the p-hfq plasmid	This work
Plasmid		
pRPF185 Δgus	pRPF185 Δgus vector derivative	14, 47
pECrF_hfq	pRPF185 Δgus carrying the hfq gene targeting the CRISPR miniarray with the full leader sequence	This work
pECrP_hfq	pRPF185 Δgus carrying the hfq gene targeting the CRISPR miniarray with the partial leader sequence under the control of the P _{tet} promoter	This work
pECrFA_hfq630	pECrF_hfq carrying arms for recombination in the 630 Δerm strain	This work
pECrPA_hfq630	pECrP_hfq carrying arms for recombination in the 630 Δerm strain	This work
pECrPA_hfqR20291	pECrP_hfq carrying arms for recombination in the R20291 strain	This work
p-hfq	pRPF185 Δgus carrying the hfq gene under the control of the P _{tet} promoter	This work

MATERIALS AND METHODS

Bacterial strains, plasmids, and growth conditions. All the plasmids and bacterial strains used in this study are listed in Table 1. *C. difficile* strains were grown in brain heart infusion (BHI; Difco) or tryptone, yeast extract (TY) (45) medium at 37°C under anaerobic conditions (5% H₂, 5% CO₂, 90% N₂) in an anaerobic chamber (Jacomex). BHI medium supplemented with yeast extract (5 mg/ml) and L-cysteine (0.1%) (BHIS) was used in the sporulation experiments. When needed, thiamphenicol (Tm) at a final concentration of 15 μg/ml was added to the *C. difficile* cultures. The *E. coli* strains (Table 1) were grown in LB medium (46) supplemented with ampicillin (100 μg/ml) and chloramphenicol (15 μg/ml) when it was suitable. The nonantibiotic analog anhydrotetracycline (ATc) was used for induction of the P_{tet} promoter of pRPF185 vector derivatives in *C. difficile* (47).

Plasmid construction and conjugation into *C. difficile*. All the oligonucleotides used in this work are listed in Table 2. To create artificial CRISPR miniarrays targeting the *C. difficile* hfq gene, the full leader sequence (positions –403 to –1 relative to the first nucleotide of the first repeat in the array) and a partial leader sequence (positions –154 to –1 relative to the first nucleotide of the first repeat in the array) of the *C. difficile* 630 Δerm CRISPR 16 array were amplified by PCR of genomic DNA (see Fig. S1A and B in the supplemental material). The artificial repeat-spacer-repeat motif was amplified by PCR from synthetic oligonucleotides to generate the double-stranded fragment. The full or partial leader sequence and the repeat-spacer-repeat motif were assembled and cloned into the BamHI and XhoI sites of the pRPF185 Δgus plasmid vector (14) using the Gibson assembly reaction (48), giving the pECrF_hfq and pECrP_hfq miniarray plasmids (Fig. 2B).

To construct editing plasmids, approximately 1,200-bp-long regions flanking the hfq gene of the 630 Δerm and R20291 strains (Fig. S1C and D) were amplified by PCR and introduced into the SmaI restriction site of pECrF_hfq or pECrP_hfq using the Gibson assembly reaction, resulting in the pECrFA_hfq630, pECrPA_hfq630, and pECrPA_hfqR20291 plasmids (Fig. 2C).

To construct a plasmid for complementation of the hfq deletion, the hfq gene sequence, including the ribosome-binding site (positions –50 to +397 relative to the translational start site), was amplified by PCR and cloned into the StuI and BamHI sites of pRPF185 Δgus under the control of the ATc-inducible P_{tet} promoter, giving the p-hfq plasmid.

DNA sequencing was performed to verify the plasmid constructs. pRPF185 Δgus is a shuttle vector that replicates both in *E. coli* (ColE1 origin) and in *C. difficile*. All resulting plasmids were transformed into the *E. coli* HB101(RP4) strain and further transferred into *C. difficile* cells by conjugation. The heat shock method with incubation for 15 min at 50°C was used to get the highest conjugation efficiency (37). *C. difficile* transconjugants were selected on BHI agar containing Tm (15 μg/ml), D-cycloserine (25 μg/ml), and ceftiofloxacin (8 μg/ml).

Deletion of the hfq gene and validation of Δhfq mutants. To induce the expression of the CRISPR miniarrays under the control of the P_{tet} promoter, *C. difficile* transconjugants containing the pECrP_hfq, pECrFA_hfq630, or pECrPA_hfqR20291 plasmid were subsequently restreaked onto BHI agar supplemented with ATc (500 ng/ml). The resulting *C. difficile* colonies were then restreaked in parallel onto BHI agar supplemented or not with Tm (15 μg/ml) to check for plasmid loss. Subsequently, selected clones without plasmids were analyzed by PCR to detect the chromosomal deletion of the hfq gene. The resulting PCR fragments were sequenced to confirm the gene deletion.

TABLE 2 Oligonucleotides used in this study

Primer purpose and name	Sequence (5'–3') ^a	Description ^b
Construction of CRISPR miniarray plasmid		
AM81	TAACAGATCTGAGCTCCAGGCCCTT CAATTATATGATAGGTTTTTTATTAAGCATACTAGCTGGTGTATATC	Full leader CR16 F
AM82	GTTAATCTAAAAC CCCCAAAATAAACTTAGTATTTCCAATATCTACACATACAC	Leader CR16-R
AM83	TAACAGATCTGAGCTCCAGGCCCTT CTGAGCAATATTTGCGATAAATTGAAGTTTAAACAATTG	Partial leader CR16-F
AM91	GTTTTAGATTAACCTATATGGAATGTAATGTTACTATATATTTAGTTAATGGAGTGCAAGTTAAAGGTT TTAGATTAACCTATATGGAATGTAAT	<i>hfq</i> repeat-spacer-repeat motif
AM92	AGTTTTATTTGGG GTTTTAGATTAACCTATATGGAATGTAATGTTACTATATATTTAGTTAATGGAGTG	Repeat-spacer-repeat F
AM93	TTTAAAGTTTTATAAAACCTTATAG ATTACATCCATATAGTTAATCTAAAACCTTTAACTGCAC	Repeat-spacer-repeat R
Construction of editing plasmids		
AM158	GAACACTTGCCGAAAAAGAAAACTGCCGGGTACGTACCCG GATATTGAAATAAAAAGTTTATTG	Left arm 630 and R20291 F
AM159	TCTTAAATTAATAATTAT TAGATTTGTACCCCTCCCAAG	Left arm 630 and R20291 R
AM160	CTTGGGAGGGTACAAATCTA ATAATTAATTTAATTAAGATGATTGAG	Right arm 630 and R20291 F
AM161	GAGCGAGGAAGCGGAAGAGCGCTCGCGGGGATCGATCCCG GAACAGGTTTTACATAAGAATC	Right arm 630 and R20291 R
<i>Δhfq</i> mutant detection		
AM106	ACTAAAAGGGTCATAAGAGC	<i>Δhfq</i> F
AM169	TATAAGGAGGCTTATTGGAGC	<i>Δhfq</i> R
Construction of plasmids for complementation		
HFQ1	<u>GAAGGCCTGGTAGGAATTTTTAGAAGT</u>	5' <i>hfq</i> <i>Stul</i>
HFQ2	<u>GGGGATCCCATTAAGCATTTTATCACCTGTC</u>	3' <i>hfq</i> <i>Bam</i> HI
qRT-PCR		
QRTBD37	GGGAGACTTGAGTGCAGGAG	16S RNA F
QRTBD38	GTGCCTCAGCGTCAGTTACA	16S RNA R
IMV447	AGGGCTGTAAAGGGGTTTG	qRT-PCR <i>hfq</i> F
IMV448	TTGTTGTTTTGGTTTTGATTGTT	qRT-PCR <i>hfq</i> R

^aOverlapping regions are indicated in boldface, and underlined sequences represent those of the restriction endonucleases.

^bCR16, CRISPR 16 array; F, forward; R, reverse.

RNA extraction and qRT-PCR. For total RNA extraction, *C. difficile* 630 Δ erm- and R20291-derived pRPF185 Δ gus- and p-*hfq*-carrying strains were grown for 6 h or 8 h in TY medium supplemented with Tm (7.5 μ g/ml) and Atc (250 ng/ml). Total RNA isolation was performed as previously described (49). cDNA synthesis by reverse transcription and quantitative reverse transcription PCR (qRT-PCR) was performed as previously described (50) using a Bio-Rad CFX Connect real-time system. The expression level of the *hfq* gene relative to that of the 16S RNA gene was calculated (51).

Protein extract preparation and Western blotting. To extract total proteins, *C. difficile* 630 Δ erm- and R20291-derived pRPF185 Δ gus- and p-*hfq*-carrying strains were grown for 6 h or 16 h in TY medium supplemented with Tm (7.5 μ g/ml) and Atc (250 ng/ml). Cell lysis and protein extraction were performed as previously described (36).

For each sample, 30 μ g of protein extract was loaded onto two 15% SDS polyacrylamide gels in parallel. After the electrophoresis, proteins from the 1st gel were transferred to a polyvinylidene fluoride membrane. Membrane hybridization with primary and secondary antibodies was then performed as described before (36). The bioluminescent signal from the secondary antibodies was detected using the SuperSignal West Pico chemiluminescent substrate (Thermo Scientific) and a Fusion FX (Vilber Lourmat) digital camera. The 2nd gel was stained with InstantBlue dye (Expedeon) and used as a loading control (Fig. S2).

Sporulation assay. *C. difficile* strains harboring the pRPF185 Δ gus and p-*hfq* plasmids were grown overnight in TY medium containing Tm (15 μ g/ml). Overnight cultures were used to inoculate the strain at an optical density at 600 nm (OD₆₀₀) of 0.1 in fresh TY medium supplemented with taurocholate (0.1%), D-fructose (0.5%), Tm (7.5 μ g/ml), and Atc (10 ng/ml) to get only vegetative cells. When the cultures had reached an OD₆₀₀ of 1.0 to 1.5, they were diluted to an OD₆₀₀ of 0.01 in BHIS medium containing Tm (7.5 μ g/ml) and Atc (10 ng/ml) and grown at 37°C. After 24 h and 48 h of growth, 1 ml of each culture was divided into two samples. To determine the total amount of bacteria (in number of CFU), the first sample was serially diluted and spotted (10 μ l per spot) onto BHI agar containing 0.1% taurocholate. The second sample was incubated at 65°C for 30 min to eliminate vegetative cells. Subsequently, the sample was serially diluted and spotted (10 μ l per spot) onto BHI agar containing 0.1% taurocholate to estimate the number of spores.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/AEM.01416-19>.

SUPPLEMENTAL FILE 1, PDF file, 0.3 MB.

ACKNOWLEDGMENTS

This work was supported by grants from the Agence Nationale de la Recherche (CloSTARn; grant ANR-13-JSV3-0005-01 to O.S.), the Institut Universitaire de France (to O.S.), the University Paris-Sud, the Institute for Integrative Biology of the Cell, the DIM-1HEALTH regional Ile de France program (LSP grant no. 164466), the CNRS-RFBR PRC 2019 (grant no. 288426) to O.S. and K.S., a Vernadski fellowship to A.M., and a Skoltech Biomedical Initiative grant (grant no. SBI RF-000000136) to K.S.

We are grateful to P. Boudry, E. Semenova, and J. Peltier for helpful discussions during the preparation of the manuscript.

REFERENCES

- Oren A, Rupnik M. 2018. *Clostridium difficile* and *Clostridioides difficile*: two validly published and correct names. *Anaerobe* 52:125–126. <https://doi.org/10.1016/j.anaerobe.2018.07.005>.
- Rupnik M, Wilcox MH, Gerding DN. 2009. *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. *Nat Rev Microbiol* 7:526–536. <https://doi.org/10.1038/nrmicro2164>.
- Abt MC, McKenney PT, Pamer EG. 2016. *Clostridium difficile* colitis: pathogenesis and host defence. *Nat Rev Microbiol* 14:609–620. <https://doi.org/10.1038/nrmicro.2016.108>.
- Warny M, Pepin J, Fang A, Killgore G, Thompson A, Brazier J, Frost E, McDonald LC. 2005. Toxin production by an emerging strain of *Clostridium difficile* associated with outbreaks of severe disease in North America and Europe. *Lancet* 366:1079–1084. [https://doi.org/10.1016/S0140-6736\(05\)67420-X](https://doi.org/10.1016/S0140-6736(05)67420-X).
- Banawas SS. 2018. *Clostridium difficile* infections: a global overview of drug sensitivity and resistance mechanisms. *Biomed Res Int* 2018: 8414257. <https://doi.org/10.1155/2018/8414257>.
- Seekatz AM, Young VB. 2014. *Clostridium difficile* and the microbiota. *J Clin Invest* 124:4182–4189. <https://doi.org/10.1172/JCI72336>.
- Just I, Selzer J, Wilm M, von Eichel-Streiber C, Mann M, Aktories K. 1995. Glucosylation of Rho proteins by *Clostridium difficile* toxin B. *Nature* 375:500–503. <https://doi.org/10.1038/375500a0>.
- Sorek R, Lawrence CM, Wiedenheft B. 2013. CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annu Rev Biochem* 82: 237–266. <https://doi.org/10.1146/annurev-biochem-072911-172315>.
- Koonin EV, Makarova KS, Zhang F. 2017. Diversity, classification and evolution of CRISPR-Cas systems. *Curr Opin Microbiol* 37:67–78. <https://doi.org/10.1016/j.mib.2017.05.008>.
- Grissa I, Vergnaud G, Pourcel C. 2007. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* 8:172. <https://doi.org/10.1186/1471-2105-8-172>.
- Shmakov SA, Sitnik V, Makarova KS, Wolf YI, Severinov KV, Koonin EV. 2017. The CRISPR spacer space is dominated by sequences from species-specific mobilomes. *mBio* 8:e01397-17. <https://doi.org/10.1128/mBio.01397-17>.
- Garneau JE, Dupuis M-V, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadán AH, Moineau S. 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468:67–71. <https://doi.org/10.1038/nature09523>.
- Marraffini LA. 2015. CRISPR-Cas immunity in prokaryotes. *Nature* 526: 55–61. <https://doi.org/10.1038/nature15386>.
- Soutourina O, Monot M, Boudry P, Saujet L, Pichon C, Sismeiro O, Semenova E, Severinov K, Le Bouguenec C, Coppée JY, Dupuy B, Martin-Verstraete I. 2013. Genome-wide identification of regulatory RNAs in the human pathogen *Clostridium difficile*. *PLoS Genet* 9:e1003493. <https://doi.org/10.1371/journal.pgen.1003493>.
- Boudry P, Semenova E, Monot M, Datsenko KA, Lopatina A, Sekulovic O, Ospina-Bedoya M, Fortier LC, Severinov K, Dupuy B, Soutourina O. 2015. Function of the CRISPR-Cas system of the human pathogen *Clostridium difficile*. *mBio* 6:e01112-15. <https://doi.org/10.1128/mBio.01112-15>.
- Andersen JM, Shoup M, Robinson C, Britton R, Olsen KEP, Barrangou R. 2016. CRISPR diversity and microevolution in *Clostridium difficile*. *Genome Biol Evol* 8:2841–2855. <https://doi.org/10.1093/gbe/evw203>.
- Hargreaves KR, Flores CO, Lawley TD, Clokie RJ, Trevor D, Hargreaves KR, Flores CO, Lawley TD, Clokie RJ. 2014. Abundant and diverse clustered regularly interspaced short palindromic repeat spacers in *Clostridium difficile* strains and prophages target multiple phage types within this pathogen. *mBio* 5:e01045-13. <https://doi.org/10.1128/mBio.01045-13>.
- Barrangou R, Horvath P. 2017. A decade of discovery: CRISPR functions and applications. *Nat Microbiol* 2:17092. <https://doi.org/10.1038/nmicrobiol.2017.92>.
- Hsu PD, Lander ES, Zhang F. 2014. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157:1262–1278. <https://doi.org/10.1016/j.cell.2014.05.010>.
- Safari F, Zare K, Negahdaripour M, Barekati-Mowahed M, Ghasemi Y. 2019. CRISPR Cpf1 proteins: structure, function and implications for genome editing. *Cell Biosci* 9:36. <https://doi.org/10.1186/s13578-019-0298-7>.
- Li Y, Pan S, Zhang Y, Ren M, Feng M, Peng N, Chen L, Liang YX, She Q. 2016. Harnessing type I and type III CRISPR-Cas systems for genome editing. *Nucleic Acids Res* 44:e34. <https://doi.org/10.1093/nar/gkv1044>.
- Cheng F, Gong L, Zhao D, Yang H, Zhou J, Li M, Xiang H. 2017. Harnessing the native type I-B CRISPR-Cas for genome editing in a polyploid archaeon. *J Genet Genomics* 44:541–548. <https://doi.org/10.1016/j.jgg.2017.09.010>.
- Pyne ME, Bruder MR, Moo-Young M, Chung DA, Chou CP. 2016. Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in *Clostridium*. *Sci Rep* 6:25666. <https://doi.org/10.1038/srep25666>.
- Zhang J, Zong W, Hong W, Zhang Z-T, Wang Y. 2018. Exploiting endogenous CRISPR-Cas system for multiplex genome editing in *Clostridium tyrobutyricum* and engineer the strain for high-level butanol production. *Metab Eng* 47:49–59. <https://doi.org/10.1016/j.ymben.2018.03.007>.
- Atmadjaja AN, Holby V, Harding AJ, Krabben P, Smith HK, Jenkinson ER. 2019. CRISPR-Cas, a highly effective tool for genome editing in *Clostridium saccharoperbutylacetonicum* N1-4(HMT). *FEMS Microbiol Lett* 366: fnz059. <https://doi.org/10.1093/femsle/fnz059>.
- Heap JT, Kuehne SA, Ehsaan M, Cartman ST, Cooksley CM, Scott JC, Minton NP. 2010. The ClosTron: mutagenesis in *Clostridium* refined and streamlined. *J Microbiol Methods* 80:49–55. <https://doi.org/10.1016/j.mimet.2009.10.018>.
- Kuehne SA, Heap JT, Cooksley CM, Cartman ST, Minton NP. 2011. ClosTron-mediated engineering of *Clostridium*. *Methods Mol Biol* 765: 389–407. https://doi.org/10.1007/978-1-61779-197-0_23.
- Cartman ST, Kelly ML, Heeg D, Heap JT, Minton NP. 2012. Precise manipulation of the *Clostridium difficile* chromosome reveals a lack of association between the *tdcC* genotype and toxin production. *Appl Environ Microbiol* 78:4683–4690. <https://doi.org/10.1128/AEM.00249-12>.
- Ng YK, Ehsaan M, Philip S, Coltery MM, Janoir C, Collignon A, Cartman ST, Minton NP. 2013. Expanding the repertoire of gene tools for precise manipulation of the *Clostridium difficile* genome: allelic exchange using *pyrE* alleles. *PLoS One* 8:e56051. <https://doi.org/10.1371/journal.pone.0056051>.
- Theophilou ES, Vohra P, Gallagher MP, Poxton IR, Blakely GW. 2018. Generation of markerless deletions in the nosocomial pathogen *Clostridium difficile* by induction of DNA double-strand breaks. *Appl Environ Microbiol* 85:e02055-18. <https://doi.org/10.1128/AEM.02055-18>.
- McAllister KN, Bouillaut L, Kahn JN, Self WT, Sorg JA. 2017. Using CRISPR-Cas9-mediated genome editing to generate *C. difficile* mutants defective in selenoproteins synthesis. *Sci Rep* 7:14672. <https://doi.org/10.1038/s41598-017-15236-5>.
- Ingle P, Groothuis D, Rowe P, Huang H, Cockayne A, Kuehne SA, Jiang W, Gu Y, Humphreys CM, Minton NP. 2019. Generation of a fully erythromycin-sensitive strain of *Clostridioides difficile* using a novel CRISPR-Cas9 genome editing system. *Sci Rep* 9:8123. <https://doi.org/10.1038/s41598-019-44458-y>.

33. Inés I, Cañ C, Groothuis D, Zygouropoulou M, Rodrigues R, Minton NP, 2019. RiboCas: a universal CRISPR-based editing tool for *Clostridium*. *ACS Synth Biol* 8:1379–1390. <https://doi.org/10.1021/acssynbio.9b00075>.
34. Hong W, Zhang J, Cui G, Wang L, Wang Y. 2018. Multiplexed CRISPR-Cpf1-mediated genome editing in *Clostridium difficile* toward the understanding of pathogenesis of *C. difficile* infection. *ACS Synth Biol* 7:1588–1600. <https://doi.org/10.1021/acssynbio.8b00087>.
35. Sobrero P, Valverde C. 2012. The bacterial protein Hfq: much more than a mere RNA-binding factor. *Crit Rev Microbiol* 38:276–299. <https://doi.org/10.3109/1040841X.2012.664540>.
36. Boudry P, Gracia C, Monot M, Caillet J, Saujet L, Hajnsdorf E, Dupuy B, Martin-Verstraete I, Soutourina O. 2014. Pleiotropic role of the RNA chaperone protein Hfq in the human pathogen *Clostridium difficile*. *J Bacteriol* 196:3234–3248. <https://doi.org/10.1128/JB.01923-14>.
37. Kirk JA, Fagan RP. 2016. Heat shock increases conjugation efficiency in *Clostridium difficile*. *Anaerobe* 42:1–5. <https://doi.org/10.1016/j.anaerobe.2016.06.009>.
38. Semenova E, Jore MM, Datsenko KA, Semenova A, Westra ER, Wanner B, van der Oost J, Brouns SJJ, Severinov K. 2011. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc Natl Acad Sci U S A* 108:10098–10103. <https://doi.org/10.1073/pnas.1104144108>.
39. Peters JM, Colavin A, Shi H, Czarny TL, Larson MH, Wong S, Hawkins JS, Lu CHS, Koo BM, Marta E, Shiver AL, Whitehead EH, Weissman JS, Brown ED, Qi LS, Huang KC, Gross C. 2016. A comprehensive, CRISPR-based functional analysis of essential genes in bacteria. *Cell* 165:1493–1506. <https://doi.org/10.1016/j.cell.2016.05.003>.
40. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA. 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152:1173–1183. <https://doi.org/10.1016/j.cell.2013.02.022>.
41. Müh U, Pannullo AG, Weiss DS, Ellermeier CD. 2019. A xylose-inducible expression system and a CRISPRi-plasmid for targeted knock-down of gene expression in *Clostridioides difficile*. *J Bacteriol* 201:e00711-18. <https://doi.org/10.1128/JB.00711-18>.
42. Luo ML, Mullis AS, Leenay RT, Beisel CL. 2015. Repurposing endogenous type I CRISPR-Cas systems for programmable gene repression. *Nucleic Acids Res* 43:674–681. <https://doi.org/10.1093/nar/gku971>.
43. Stachler A-E, Marchfelder A. 2016. Gene repression in *Haloarchaea* using the CRISPR (clustered regularly interspaced short palindromic repeats)-Cas I-B system. *J Biol Chem* 291:15226–15242. <https://doi.org/10.1074/jbc.M116.724062>.
44. Bikard D, Barrangou R. 2017. Using CRISPR-Cas systems as antimicrobials. *Curr Opin Microbiol* 37:155–160. <https://doi.org/10.1016/j.mib.2017.08.005>.
45. Dupuy B, Sonenshein AL. 1998. Regulated transcription of *Clostridium difficile* toxin genes. *Mol Microbiol* 27:107–120. <https://doi.org/10.1046/j.1365-2958.1998.00663.x>.
46. Bertani G. 1951. Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J Bacteriol* 62:293–300.
47. Fagan RP, Fairweather NF. 2011. *Clostridium difficile* has two parallel and essential Sec secretion systems. *J Biol Chem* 286:27483–27493. <https://doi.org/10.1074/jbc.M111.263889>.
48. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6:343–345. <https://doi.org/10.1038/nmeth.1318>.
49. André G, Even S, Putzer H, Burguière P, Croux C, Danchin A, Martin-Verstraete I, Soutourina O. 2008. S-box and T-box riboswitches and antisense RNA control a sulfur metabolic operon of *Clostridium acetobutylicum*. *Nucleic Acids Res* 36:5955–5969. <https://doi.org/10.1093/nar/gkn601>.
50. Saujet L, Monot M, Dupuy B, Soutourina O, Martin-Verstraete I. 2011. The key sigma factor of transition phase, SigH, controls sporulation, metabolism, and virulence factor expression in *Clostridium difficile*. *J Bacteriol* 193:3186–3196. <https://doi.org/10.1128/JB.00272-11>.
51. Metcalf D, Sharif S, Weese JS. 2010. Evaluation of candidate reference genes in *Clostridium difficile* for gene expression normalization. *Anaerobe* 16:439–443. <https://doi.org/10.1016/j.anaerobe.2010.06.007>.
52. Hussain HA, Roberts AP, Mullany P. 2005. Generation of an erythromycin-sensitive derivative of *Clostridium difficile* strain 630 (630 Δ erm) and demonstration that the conjugative transposon Tn916 Δ E enters the genome of this strain at multiple sites. *J Med Microbiol* 54:137–141. <https://doi.org/10.1099/jmm.0.45790-0>.

1. Lawson, P. A., Citron, D. M., Tyrrell, K. L. & Finegold, S. M. Reclassification of *Clostridium difficile* as *Clostridioides difficile* (Hall and O'Toole 1935) Prévot 1938. *Anaerobe* **40**, 95–99 (2016).
2. Smits, W. K., Lyras, D., Lacy, D. B., Wilcox, M. H. & Kuijper, E. J. *Clostridium difficile* infection. *Nat. Rev. Dis. Primer* **2**, 16020 (2016).
3. Hall, I. C. INTESTINAL FLORA IN NEW-BORN INFANTS: WITH A DESCRIPTION OF A NEW PATHOGENIC ANAEROBE, BACILLUS DIFFICILIS. *Am. J. Dis. Child.* **49**, 390 (1935).
4. Cohen, L. E., McNeill, C. J. & Wells, R. F. Clindamycin-Associated Colitis. *JAMA* **223**, 1379–1380 (1973).
5. Tedesco, F. J. Clindamycin-Associated Colitis: A Prospective Study. *Ann. Intern. Med.* **81**, 429 (1974).
6. Bartlett, J. G. *Clostridium difficile*: History of Its Role as an Enteric Pathogen and the Current State of Knowledge About the Organism. *Clin. Infect. Dis.* **18**, S265–S272 (1994).
7. Bartlett, J. G., Onderdonk, A. B., Cisneros, R. L. & Kasper, D. L. Clindamycin-Associated Colitis Due to a Toxin-Producing Species of *Clostridium* in Hamsters. *J. Infect. Dis.* **136**, 701–705 (1977).
8. Bartlett, J. G., Chang, T. W., Gurwith, M., Gorbach, S. L. & Onderdonk, A. B. Antibiotic-associated pseudomembranous colitis due to toxin-producing clostridia. *N. Engl. J. Med.* **298**, 531–534 (1978).
9. Bartlett, J. G., Chang, T. W., Moon, N. & Onderdonk, A. B. Antibiotic-induced lethal enterocolitis in hamsters: studies with eleven agents and evidence to support the pathogenic role of toxin-producing *Clostridia*. *Am. J. Vet. Res.* **39**, 1525–1530 (1978).
10. Rupnik, M., Wilcox, M. H. & Gerding, D. N. *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. *Nat. Rev. Microbiol.* **7**, 526–536 (2009).
11. Karen C., C. & John G., B. Biology of *Clostridium difficile*: Implications for Epidemiology and Diagnosis. *Annu. Rev. Microbiol.* **65**, 501–521 (2011).
12. Czepiel, J. *et al.* *Clostridium difficile* infection: review. *Eur. J. Clin. Microbiol. Infect. Dis.* **38**, 1211–1221 (2019).
13. Kazanowski, M., Smolarek, S., Kinnarney, F. & Grzebieniak, Z. *Clostridium difficile*: epidemiology, diagnostic and therapeutic possibilities—a systematic review. *Tech. Coloproctology* **18**, 223–232 (2014).
14. De Roo, A. C. & Regenbogen, S. E. *Clostridium difficile* Infection: An Epidemiology Update. *Clin. Colon Rectal Surg.* **33**, 049–057 (2020).
15. Rea, M. C. *et al.* *Clostridium difficile* Carriage in Elderly Subjects and Associated Changes in the Intestinal Microbiota. *J. Clin. Microbiol.* **50**, 867–875 (2012).
16. Kyne, L., Warny, M., Qamar, A. & Kelly, C. P. Asymptomatic Carriage of *Clostridium difficile* and Serum Levels of IgG Antibody against Toxin A. *N. Engl. J. Med.* **342**, 390–397 (2000).
17. DePestel, D. D. & Aronoff, D. M. Epidemiology of *Clostridium difficile* Infection. *J. Pharm. Pract.* **26**, 464–475 (2013).
18. Gerding, D. N. & Lessa, F. C. The Epidemiology of *Clostridium difficile* Infection Inside and Outside Health Care Institutions. *Infect. Dis. Clin. North Am.* **29**, 37–50 (2015).
19. Khanna, S. *et al.* The Epidemiology of Community-Acquired *Clostridium difficile* Infection: A Population-Based Study. *Am. J. Gastroenterol.* **107**, 89–95 (2012).

20. Goorhuis, A. *et al.* Emergence of *Clostridium difficile* Infection Due to a New Hypervirulent Strain, Polymerase Chain Reaction Ribotype 078. *Clin. Infect. Dis.* **47**, 1162–1170 (2008).
21. Killgore, G. *et al.* Comparison of Seven Techniques for Typing International Epidemic Strains of *Clostridium difficile*: Restriction Endonuclease Analysis, Pulsed-Field Gel Electrophoresis, PCR-Ribotyping, Multilocus Sequence Typing, Multilocus Variable-Number Tandem-Repeat Analysis, Amplified Fragment Length Polymorphism, and Surface Layer Protein A Gene Sequence Typing. *J. Clin. Microbiol.* **46**, 431–437 (2008).
22. Brazier, J. S. Typing of *Clostridium difficile*. *Clin. Microbiol. Infect.* **7**, 428–431 (2001).
23. Freeman, J. *et al.* Pan-European longitudinal surveillance of antibiotic resistance among prevalent *Clostridium difficile* ribotypes. *Clin. Microbiol. Infect.* **21**, 248.e9–248.e16 (2015).
24. *Updates on Clostridium difficile in Europe: Advances in Microbiology, Infectious Diseases and Public Health Volume 8.* vol. 1050 (Springer International Publishing, 2018).
25. Eckert, C. *et al.* Clinical and microbiological features of *Clostridium difficile* infections in France: The ICD-RAISIN 2009 national survey. *Médecine Mal. Infect.* **43**, 67–74 (2013).
26. Freeman, J. *et al.* The Changing Epidemiology of *Clostridium difficile* Infections. *Clin. Microbiol. Rev.* **23**, 529–549 (2010).
27. Kelly, C. P. & LaMont, J. T. *Clostridium difficile* — More Difficult Than Ever. *N. Engl. J. Med.* **359**, 1932–1940 (2008).
28. Kuijper, E. J., Coignard, B. & Tüll, P. Emergence of *Clostridium difficile*-associated disease in North America and Europe. *Clin. Microbiol. Infect.* **12**, 2–18 (2006).
29. McDonald, L. C. *et al.* An Epidemic, Toxin Gene-Variant Strain of *Clostridium difficile*. *N. Engl. J. Med.* **353**, 2433–2441 (2005).
30. Loo, V. G. *et al.* A Predominantly Clonal Multi-Institutional Outbreak of *Clostridium difficile*-Associated Diarrhea with High Morbidity and Mortality. *N. Engl. J. Med.* **353**, 2442–2449 (2005).
31. Dallal, R. M. *et al.* Fulminant *Clostridium difficile*: An Underappreciated and Increasing Cause of Death and Complications: *Ann. Surg.* **235**, 363–372 (2002).
32. Oake, N. *et al.* The Effect of Hospital-Acquired *Clostridium difficile* Infection on In-Hospital Mortality. *Arch. Intern. Med.* **170**, (2010).
33. Muto, C. A. *et al.* A Large Outbreak of *Clostridium difficile*-Associated Disease with an Unexpected Proportion of Deaths and Colectomies at a Teaching Hospital Following Increased Fluoroquinolone Use. *Infect. Control Hosp. Epidemiol.* **26**, 273–280 (2005).
34. McDonald, L. C., Owings, M. & Jernigan, D. B. *Clostridium difficile* Infection in Patients Discharged from US Short-stay Hospitals, 1996–2003. *Emerg. Infect. Dis.* **12**, 7 (2006).
35. Pepin, J. *Clostridium difficile*-associated diarrhea in a region of Quebec from 1991 to 2003: a changing pattern of disease severity. *Can. Med. Assoc. J.* **171**, 466–472 (2004).
36. Bauer, M. P. *et al.* *Clostridium difficile* infection in Europe: a hospital-based survey. *The Lancet* **377**, 63–73 (2011).
37. Dubberke, E. R. & Olsen, M. A. Burden of *Clostridium difficile* on the Healthcare System. *Clin. Infect. Dis.* **55**, S88–S92 (2012).
38. Bartlett, J. G. Narrative Review: The New Epidemic of *Clostridium difficile*-Associated Enteric Disease. *Ann. Intern. Med.* **145**, 758 (2006).
39. Valiente, E., Cairns, M. D. & Wren, B. W. The *Clostridium difficile* PCR ribotype 027 lineage: a

- pathogen on the move. *Clin. Microbiol. Infect.* **20**, 396–404 (2014).
40. Pepin, J. Mortality attributable to nosocomial *Clostridium difficile*-associated disease during an epidemic caused by a hypervirulent strain in Quebec. *Can. Med. Assoc. J.* **173**, 1037–1042 (2005).
 41. Popoff, M. R., Rubin, E. J., Gill, D. M. & Boquet, P. Actin-specific ADP-ribosyltransferase produced by a *Clostridium difficile* strain. *Infect. Immun.* **56**, 2299–2306 (1988).
 42. Kuijper, E. J. *et al.* Update of *Clostridium difficile* infection due to PCR ribotype 027 in Europe, 2008. *Euro Surveill. Bull. Eur. Sur Mal. Transm. Eur. Commun. Dis. Bull.* **13**, 18942 (2008).
 43. Kim, H. *et al.* Emergence of *Clostridium difficile* Ribotype 027 in Korea. *Ann. Lab. Med.* **31**, 191–196 (2011).
 44. Cheng, V. C. C. *et al.* *Clostridium difficile* ribotype 027 arrives in Hong Kong. *Int. J. Antimicrob. Agents* **34**, 492–493 (2009).
 45. Kato, H., Ito, Y., van den Berg, R. J., Kuijper, E. J. & Arakawa, Y. First isolation of *Clostridium difficile* 027 in Japan. *Euro Surveill. Bull. Eur. Sur Mal. Transm. Eur. Commun. Dis. Bull.* **12**, E070111.3 (2007).
 46. Riley, T. V., Thean, S., Hool, G. & Golledge, C. L. First Australian isolation of epidemic *Clostridium difficile* PCR ribotype 027. *Med. J. Aust.* **190**, 706–708 (2009).
 47. Clements, A. C., Magalhães, R. J. S., Tatem, A. J., Paterson, D. L. & Riley, T. V. *Clostridium difficile* PCR ribotype 027: assessing the risks of further worldwide spread. *Lancet Infect. Dis.* **10**, 395–404 (2010).
 48. Tachon, M. *et al.* First cluster of *C. difficile* toxinotype III, PCR-ribotype 027 associated disease in France: preliminary report. *Euro Surveill. Bull. Eur. Sur Mal. Transm. Eur. Commun. Dis. Bull.* **11**, E060504.1 (2006).
 49. Bauer, K. A. *et al.* Impact of the NAP-1 strain on disease severity, mortality, and recurrence of healthcare-associated *Clostridium difficile* infection. *Anaerobe* **48**, 1–6 (2017).
 50. Mani, N. S., Lynch, J. B., Fang, F. C. & Chan, J. D. Risk Factors for BI/NAP1/027 *Clostridioides difficile* Infections and Clinical Outcomes Compared With Non-NAP1 Strains. *Open Forum Infect. Dis.* **6**, ofz433 (2019).
 51. Stabler, R. A. *et al.* Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the evolution of a hypervirulent bacterium. *Genome Biol.* **10**, R102 (2009).
 52. Stabler, R. A. *et al.* In-depth genetic analysis of *Clostridium difficile* PCR-ribotype 027 strains reveals high genome fluidity including point mutations and inversions. *Gut Microbes* **1**, 269–276 (2010).
 53. Dupuy, B., Govind, R., Antunes, A. & Matamouros, S. *Clostridium difficile* toxin synthesis is negatively regulated by TcdC. *J. Med. Microbiol.* **57**, 685–689 (2008).
 54. Spigaglia, P. & Mastrantonio, P. Molecular Analysis of the Pathogenicity Locus and Polymorphism in the Putative Negative Regulator of Toxin Production (TcdC) among *Clostridium difficile* Clinical Isolates. *J. Clin. Microbiol.* **40**, 3470–3475 (2002).
 55. Curry, S. R. *et al.* *tcdC* Genotypes Associated with Severe TcdC Truncation in an Epidemic Clone and Other Strains of *Clostridium difficile*. *J. Clin. Microbiol.* **45**, 215–221 (2007).
 56. Murray, R., Boyd, D., Levett, P. N., Mulvey, M. R. & Alfa, M. J. Truncation in the *tcdC* 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 (2009).
57. Åkerlund, T., Svenungsson, B., Lagergren, Å. & Burman, L. G. Correlation of Disease Severity with Fecal Toxin Levels in Patients with *Clostridium difficile* -Associated Diarrhea and Distribution of PCR Ribotypes and Toxin Yields In Vitro of Corresponding Isolates. *J. Clin. Microbiol.* **44**, 353–358 (2006).
 58. Merrigan, M. *et al.* Human Hypervirulent *Clostridium difficile* Strains Exhibit Increased Sporulation as Well as Robust Toxin Production. *J. Bacteriol.* **192**, 4904–4911 (2010).
 59. Burns, D. A., Heeg, D., Cartman, S. T. & Minton, N. P. Reconsidering the Sporulation Characteristics of Hypervirulent *Clostridium difficile* BI/NAP1/027. *PLoS ONE* **6**, e24894 (2011).
 60. Bartlett, J. G. & Gerding, D. N. Clinical Recognition and Diagnosis of *Clostridium difficile* Infection. *Clin. Infect. Dis.* **46**, S12–S18 (2008).
 61. Planche, T. & Wilcox, M. H. Diagnostic Pitfalls in *Clostridium difficile* Infection. *Infect. Dis. Clin. North Am.* **29**, 63–82 (2015).
 62. Planche, T. & Wilcox, M. Reference assays for *Clostridium difficile* infection: one or two gold standards? *J. Clin. Pathol.* **64**, 1–5 (2011).
 63. Carey-Ann, B. D. & Carroll, K. C. Diagnosis of *Clostridium difficile* Infection: an Ongoing Conundrum for Clinicians and for Clinical Laboratories. *Clin. Microbiol. Rev.* **26**, 604–630 (2013).
 64. Taylor, N. S., Thorne, G. M. & Bartlett, J. G. Comparison of two toxins produced by *Clostridium difficile*. *Infect. Immun.* **34**, 1036–1043 (1981).
 65. Delmée, M. Laboratory diagnosis of *Clostridium difficile* disease. *Clin. Microbiol. Infect.* **7**, 411–416 (2001).
 66. Monot, M. *et al.* *Clostridium difficile*: New Insights into the Evolution of the Pathogenicity Locus. *Sci. Rep.* **5**, 15023 (2015).
 67. Curry, S. *Clostridium difficile*. *Clin. Lab. Med.* **30**, 329–342 (2010).
 68. Crobach, M. J. T. *et al.* European Society of Clinical Microbiology and Infectious Diseases: update of the diagnostic guidance document for *Clostridium difficile* infection. *Clin. Microbiol. Infect.* **22**, S63–S81 (2016).
 69. Crobach, M. J. T., Dekkers, O. M., Wilcox, M. H. & Kuijper, E. J. European Society of Clinical Microbiology and Infectious Diseases (ESCMID): Data review and recommendations for diagnosing *Clostridium difficile*-infection (CDI). *Clin. Microbiol. Infect.* **15**, 1053–1066 (2009).
 70. She, R. C., Durrant, R. J. & Petti, C. A. Evaluation of Enzyme Immunoassays to Detect *Clostridium difficile* Toxin From Anaerobic Stool Culture. *Am. J. Clin. Pathol.* **131**, 81–84 (2009).
 71. Eastwood, K., Else, P., Charlett, A. & Wilcox, M. Comparison of Nine Commercially Available *Clostridium difficile* Toxin Detection Assays, a Real-Time PCR Assay for *C. difficile tcdB*, and a Glutamate Dehydrogenase Detection Assay to Cytotoxin Testing and Cytotoxigenic Culture Methods. *J. Clin. Microbiol.* **47**, 3211–3217 (2009).
 72. Bartlett, J. G. *Clostridium difficile*: progress and challenges. *Ann. N. Y. Acad. Sci.* **1213**, 62–69 (2010).
 73. Carroll, K. C. Tests for the diagnosis of *Clostridium difficile* infection: The next generation. *Anaerobe* **17**, 170–174 (2011).
 74. Ticehurst, J. R. *et al.* Effective Detection of Toxigenic *Clostridium difficile* by a Two-Step

- Algorithm Including Tests for Antigen and Cytotoxin. *J. Clin. Microbiol.* **44**, 1145–1149 (2006).
75. Boyanton, B. L. *et al.* Loop-Mediated Isothermal Amplification Compared to Real-Time PCR and Enzyme Immunoassay for Toxigenic *Clostridium difficile* Detection. *J. Clin. Microbiol.* **50**, 640–645 (2012).
 76. Tenover, F. C. *et al.* Impact of Strain Type on Detection of Toxigenic *Clostridium difficile*: Comparison of Molecular Diagnostic and Enzyme Immunoassay Approaches. *J. Clin. Microbiol.* **48**, 3719–3724 (2010).
 77. Crobach, M. J. T., Baktash, A., Duszenko, N. & Kuijper, E. J. Diagnostic Guidance for *C. difficile* Infections. in *Updates on Clostridium difficile in Europe: Advances in Microbiology, Infectious Diseases and Public Health Volume 8* (eds. Mastrantonio, P. & Rupnik, M.) 27–44 (Springer International Publishing, 2018). doi:10.1007/978-3-319-72799-8_3.
 78. Lee, H. S., Plechot, K., Gohil, S. & Le, J. *Clostridium difficile*: Diagnosis and the Consequence of Over Diagnosis. *Infect. Dis. Ther.* **10**, 687–697 (2021).
 79. McDonald, L. C. *et al.* Clinical Practice Guidelines for *Clostridium difficile* Infection in Adults and Children: 2017 Update by the Infectious Diseases Society of America (IDSA) and Society for Healthcare Epidemiology of America (SHEA). *Clin. Infect. Dis.* **66**, e1–e48 (2018).
 80. Surawicz, C. M. *et al.* Guidelines for Diagnosis, Treatment, and Prevention of *Clostridium difficile* Infections. *Am. J. Gastroenterol.* **108**, 478–498 (2013).
 81. Dial, S., Delaney, J. A. C., Barkun, A. N. & Suissa, S. Use of Gastric Acid-Suppressive Agents and the Risk of Community-Acquired *Clostridium difficile*-Associated Disease. *JAMA* **294**, 2989–2995 (2005).
 82. Aslam, S., Hamill, R. J. & Musher, D. M. Treatment of *Clostridium difficile*-associated disease: old therapies and new strategies. *Lancet Infect. Dis.* **5**, 549–557 (2005).
 83. Hu, M. Y. *et al.* Prospective Derivation and Validation of a Clinical Prediction Rule for Recurrent *Clostridium difficile* Infection. *Gastroenterology* **136**, 1206–1214 (2009).
 84. Dinh, A., Bouchand, F. & Le Monnier, A. Actualités épidémiologiques et thérapeutiques des infections à *Clostridium difficile*. *Rev. Médecine Interne* **36**, 596–602 (2015).
 85. Debast, S. B., Bauer, M. P. & Kuijper, E. J. European Society of Clinical Microbiology and Infectious Diseases: Update of the Treatment Guidance Document for *Clostridium difficile* Infection. *Clin. Microbiol. Infect.* **20**, 1–26 (2014).
 86. Vonberg, R.-P. *et al.* Infection control measures to limit the spread of *Clostridium difficile*. *Clin. Microbiol. Infect.* **14**, 2–20 (2008).
 87. Drekonja, D. M. *et al.* Comparative Effectiveness of *Clostridium difficile* Treatments: A Systematic Review. *Ann. Intern. Med.* **155**, 839 (2011).
 88. Bagdasarian, N., Rao, K. & Malani, P. N. Diagnosis and Treatment of *Clostridium difficile* in Adults: A Systematic Review. *JAMA* **313**, 398 (2015).
 89. Zhang, K. *et al.* The non-toxigenic *Clostridium difficile* CD37 protects mice against infection with a BI/NAP1/027 type of *C. difficile* strain. *Anaerobe* **36**, 49–52 (2015).
 90. Gerding, D. N. *et al.* Administration of Spores of Nontoxigenic *Clostridium difficile* Strain M3 for Prevention of Recurrent *C. difficile* Infection: A Randomized Clinical Trial. *JAMA* **313**, 1719 (2015).
 91. Gujja, D. & Friedenber, F. K. Predictors of serious complications due to *Clostridium difficile*

- infection. *Aliment. Pharmacol. Ther.* **29**, 635–642 (2009).
92. Fujitani, S., George, W. L. & Murthy, A. R. Comparison of Clinical Severity Score Indices for *Clostridium difficile* Infection. *Infect. Control Hosp. Epidemiol.* **32**, 220–228 (2011).
 93. Zar, F. A., Bakkanagari, S. R., Moorthi, K. M. L. S. T. & Davis, M. B. A Comparison of Vancomycin and Metronidazole for the Treatment of *Clostridium difficile*-Associated Diarrhea, Stratified by Disease Severity. *Clin. Infect. Dis.* **45**, 302–307 (2007).
 94. Johnson, S. *et al.* Vancomycin, Metronidazole, or Tolevamer for *Clostridium difficile* Infection: Results From Two Multinational, Randomized, Controlled Trials. *Clin. Infect. Dis.* **59**, 345–354 (2014).
 95. Nelson, R. L., Suda, K. J. & Evans, C. T. Antibiotic treatment for *Clostridium difficile* -associated diarrhoea in adults. *Cochrane Database Syst. Rev.* **2017**, (2017).
 96. Cornely, O. A., Miller, M. A., Louie, T. J., Crook, D. W. & Gorbach, S. L. Treatment of First Recurrence of *Clostridium difficile* Infection: Fidaxomicin Versus Vancomycin. *Clin. Infect. Dis.* **55**, S154–S161 (2012).
 97. Louie, T. J. *et al.* Fidaxomicin versus Vancomycin for *Clostridium difficile* Infection. *N. Engl. J. Med.* **364**, 422–431 (2011).
 98. Nerandzic, M. M., Mullane, K., Miller, M. A., Babakhani, F. & Donskey, C. J. Reduced Acquisition and Overgrowth of Vancomycin-Resistant Enterococci and *Candida* Species in Patients Treated With Fidaxomicin Versus Vancomycin for *Clostridium difficile* Infection. *Clin. Infect. Dis.* **55**, S121–S126 (2012).
 99. Johnson, S. Recurrent *Clostridium difficile* infection: A review of risk factors, treatments, and outcomes. *J. Infect.* **58**, 403–410 (2009).
 100. Gough, E., Shaikh, H. & Manges, A. R. Systematic Review of Intestinal Microbiota Transplantation (Fecal Bacteriotherapy) for Recurrent *Clostridium difficile* Infection. *Clin. Infect. Dis.* **53**, 994–1002 (2011).
 101. Smits, W. K. Hype or hypervirulence: A reflection on problematic *C. difficile* strains. *Virulence* **4**, 592–596 (2013).
 102. Paredes-Sabja, D., Shen, A. & Sorg, J. A. *Clostridium difficile* spore biology: sporulation, germination, and spore structural proteins. *Trends Microbiol.* **22**, 406–416 (2014).
 103. Edwards, A. N. *et al.* Chemical and Stress Resistances of *Clostridium difficile* Spores and Vegetative Cells. *Front. Microbiol.* **7**, (2016).
 104. Setlow, P. Spore Resistance Properties. *Microbiol. Spectr.* **2**, (2014).
 105. Paredes-Sabja, D. & Sarker, M. R. Adherence of *Clostridium difficile* spores to Caco-2 cells in culture. *J. Med. Microbiol.* **61**, 1208–1218 (2012).
 106. Theriot, C. M. & Young, V. B. Interactions Between the Gastrointestinal Microbiome and *Clostridium difficile*. *Annu. Rev. Microbiol.* **69**, 445–461 (2015).
 107. Crobach, M. J. T. *et al.* Understanding *Clostridium difficile* Colonization. *Clin. Microbiol. Rev.* **31**, (2018).
 108. Seekatz, A. M. & Young, V. B. *Clostridium difficile* and the microbiota. *J. Clin. Invest.* **124**, 4182–4189 (2014).
 109. Lynch, S. V. & Pedersen, O. The Human Intestinal Microbiome in Health and Disease. *N. Engl. J. Med.* **375**, 2369–2379 (2016).

110. Buffie, C. G. & Pamer, E. G. Microbiota-mediated colonization resistance against intestinal pathogens. *Nat. Rev. Immunol.* **13**, 790–801 (2013).
111. Ducarmon, Q. R. *et al.* Gut Microbiota and Colonization Resistance against Bacterial Enteric Infection. *Microbiol. Mol. Biol. Rev.* **83**, (2019).
112. Buffie, C. G. *et al.* Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. *Nature* **517**, 205–208 (2015).
113. Vedantam, G. *et al.* *Clostridium difficile* infection: Toxins and non-toxin virulence factors, and their contributions to disease establishment and host response. *Gut Microbes* **3**, 121–134 (2012).
114. Britton, R. A. & Young, V. B. Role of the Intestinal Microbiota in Resistance to Colonization by *Clostridium difficile*. *Gastroenterology* **146**, 1547–1553 (2014).
115. Theriot, C. M., Bowman, A. A. & Young, V. B. Antibiotic-Induced Alterations of the Gut Microbiota Alter Secondary Bile Acid Production and Allow for *Clostridium difficile* Spore Germination and Outgrowth in the Large Intestine. *mSphere* **1**, (2016).
116. Rea, M. C. *et al.* Thuricin CD, a posttranslationally modified bacteriocin with a narrow spectrum of activity against *Clostridium difficile*. *Proc. Natl. Acad. Sci.* **107**, 9352–9357 (2010).
117. Deng, H. *et al.* *Bacteroides fragilis* Prevents *Clostridium difficile* Infection in a Mouse Model by Restoring Gut Barrier and Microbiome Regulation. *Front. Microbiol.* **9**, 2976 (2018).
118. Reeves, A. E., Koenigsnecht, M. J., Bergin, I. L. & Young, V. B. Suppression of *Clostridium difficile* in the Gastrointestinal Tracts of Germfree Mice Inoculated with a Murine Isolate from the Family Lachnospiraceae. *Infect. Immun.* **80**, 3786–3794 (2012).
119. Hryckowian, A. J., Pruss, K. M. & Sonnenburg, J. L. The emerging metabolic view of *Clostridium difficile* pathogenesis. *Curr. Opin. Microbiol.* **35**, 42–47 (2017).
120. Fletcher, J. R. *et al.* *Clostridioides difficile* exploits toxin-mediated inflammation to alter the host nutritional landscape and exclude competitors from the gut microbiota. *Nat. Commun.* **12**, 462 (2021).
121. Zhang, L. *et al.* Insight into alteration of gut microbiota in *Clostridium difficile* infection and asymptomatic *C. difficile* colonization. *Anaerobe* **34**, 1–7 (2015).
122. Reeves, A. E. *et al.* The interplay between microbiome dynamics and pathogen dynamics in a murine model of *Clostridium difficile* Infection. *Gut Microbes* **2**, 145–158 (2011).
123. Slimings, C. & Riley, T. V. Antibiotics and hospital-acquired *Clostridium difficile* infection: update of systematic review and meta-analysis. *J. Antimicrob. Chemother.* **69**, 881–891 (2014).
124. Mullish, B. H. & Williams, H. R. *Clostridium difficile* infection and antibiotic-associated diarrhoea. *Clin. Med.* **18**, 237–241 (2018).
125. Voth, D. E. & Ballard, J. D. *Clostridium difficile* Toxins: Mechanism of Action and Role in Disease. *Clin. Microbiol. Rev.* **18**, 247–263 (2005).
126. Dobson, G., Hickey, C. & Trinder, J. *Clostridium difficile* colitis causing toxic megacolon, severe sepsis and multiple organ dysfunction syndrome. *Intensive Care Med.* **29**, 1030–1030 (2003).
127. Leffler, D. A. & Lamont, J. T. *Clostridium difficile* Infection. *N. Engl. J. Med.* **372**, 1539–1548 (2015).
128. Farrell, R. J. & LaMont, J. T. Pathogenesis and clinical manifestations of *Clostridium difficile*

- diarrhea and colitis. *Curr. Top. Microbiol. Immunol.* **250**, 109–125 (2000).
129. Kachrimanidou, M. & Malisiovas, N. *Clostridium difficile* Infection: A Comprehensive Review. *Crit. Rev. Microbiol.* **37**, 178–187 (2011).
 130. Awad, M. M., Johanesen, P. A., Carter, G. P., Rose, E. & Lyras, D. *Clostridium difficile* virulence factors: Insights into an anaerobic spore-forming pathogen. *Gut Microbes* **5**, 579–593 (2014).
 131. Janoir, C. Virulence factors of *Clostridium difficile* and their role during infection. *Anaerobe* **37**, 13–24 (2016).
 132. Sára, M. & Sleytr, U. B. S-Layer proteins. *J. Bacteriol.* **182**, 859–868 (2000).
 133. Desvaux, M., Dumas, E., Chafsey, I. & Hébraud, M. Protein cell surface display in Gram-positive bacteria: from single protein to macromolecular protein structure. *FEMS Microbiol. Lett.* **256**, 1–15 (2006).
 134. Calabi, E. *et al.* Molecular characterization of the surface layer proteins from *Clostridium difficile*. *Mol. Microbiol.* **40**, 1187–1199 (2001).
 135. Bradshaw, W. J., Roberts, A. K., Shone, C. C. & Acharya, K. R. The structure of the S-layer of *Clostridium difficile*. *J. Cell Commun. Signal.* **12**, 319–331 (2018).
 136. Dang, T. H. T. *et al.* Chemical Probes of Surface Layer Biogenesis in *Clostridium difficile*. *ACS Chem. Biol.* **5**, 279–285 (2010).
 137. Kirby, J. M. *et al.* Cwp84, a Surface-associated Cysteine Protease, Plays a Role in the Maturation of the Surface Layer of *Clostridium difficile*. *J. Biol. Chem.* **284**, 34666–34673 (2009).
 138. Fagan, R. P. *et al.* Structural insights into the molecular organization of the S-layer from *Clostridium difficile*. *Mol. Microbiol.* **71**, 1308–1322 (2009).
 139. Qazi, O. *et al.* Mass spectrometric analysis of the S-layer proteins from *Clostridium difficile* demonstrates the absence of glycosylation. *J. Mass Spectrom.* **44**, 368–374 (2009).
 140. Dingle, K. E. *et al.* Recombinational Switching of the *Clostridium difficile* S-Layer and a Novel Glycosylation Gene Cluster Revealed by Large-Scale Whole-Genome Sequencing. *J. Infect. Dis.* **207**, 675–686 (2013).
 141. Calabi, E., Calabi, F., Phillips, A. D. & Fairweather, N. F. Binding of *Clostridium difficile* Surface Layer Proteins to Gastrointestinal Tissues. *Infect. Immun.* **70**, 5770–5778 (2002).
 142. Merrigan, M. M. *et al.* Surface-Layer Protein A (SlpA) Is a Major Contributor to Host-Cell Adherence of *Clostridium difficile*. *PLoS ONE* **8**, e78404 (2013).
 143. Karjalainen, T., Saumier, N., Barc, M.-C., Delmée, M. & Collignon, A. *Clostridium difficile* genotyping based on slpA variable region in S-layer gene sequence: an alternative to serotyping. *J. Clin. Microbiol.* **40**, 2452–2458 (2002).
 144. Fagan, R. P. *et al.* A proposed nomenclature for cell wall proteins of *Clostridium difficile*. *J. Med. Microbiol.* **60**, 1225–1228 (2011).
 145. Willing, S. E. *et al.* *C. lostridium difficile* surface proteins are anchored to the cell wall using CWB 2 motifs that recognise the anionic polymer PSII. *Mol. Microbiol.* **96**, 596–608 (2015).
 146. Cafardi, V. *et al.* Identification of a Novel Zinc Metalloprotease through a Global Analysis of *Clostridium difficile* Extracellular Proteins. *PLoS ONE* **8**, e81306 (2013).
 147. Hensbergen, P. J. *et al.* *Clostridium difficile* secreted Pro-Pro endopeptidase PPEP-1 (ZMP1/CD2830) modulates adhesion through cleavage of the collagen binding protein CD2831. *FEBS Lett.* **589**, 3952–3958 (2015).

148. Kovacs-Simon, A. *et al.* Lipoprotein CD0873 Is a Novel Adhesin of *Clostridium difficile*. *J. Infect. Dis.* **210**, 274–284 (2014).
149. Collignon, A. *et al.* GroEL (Hsp60) of *Clostridium difficile* is involved in cell adherence. *Microbiology* **147**, 87–96 (2001).
150. Cerquetti, M., Serafino, A., Sebastianelli, A. & Mastrantonio, P. Binding of *Clostridium difficile* to Caco-2 epithelial cell line and to extracellular matrix proteins. *FEMS Immunol. Med. Microbiol.* **32**, 211–218 (2002).
151. Lin, Y.-P., Kuo, C.-J., Koleci, X., McDonough, S. P. & Chang, Y.-F. Manganese Binds to *Clostridium difficile* Fbp68 and Is Essential for Fibronectin Binding. *J. Biol. Chem.* **286**, 3957–3969 (2011).
152. Hennequin, C., Janoir, C., Barc, M.-C., Collignon, A. & Karjalainen, T. Identification and characterization of a fibronectin-binding protein from *Clostridium difficile*. *Microbiol. Read. Engl.* **149**, 2779–87 (2003).
153. Tulli, L. *et al.* CbpA: a novel surface exposed adhesin of *Clostridium difficile* targeting human collagen. *Cell. Microbiol.* **15**, 1674–1687 (2013).
154. Sebaihia, M. *et al.* The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nat. Genet.* **38**, 779–786 (2006).
155. Janvilisri, T. *et al.* Microarray Identification of *Clostridium difficile* Core Components and Divergent Regions Associated with Host Origin. *J. Bacteriol.* **191**, 3881–3891 (2009).
156. Piepenbrink, K. H. *et al.* Structural and Evolutionary Analyses Show Unique Stabilization Strategies in the Type IV Pili of *Clostridium difficile*. *Structure* **23**, 385–396 (2015).
157. Maldarelli, G. A., De Masi, L., von Rosenvinge, E. C., Carter, M. & Donnenberg, M. S. Identification, immunogenicity, and cross-reactivity of type IV pilin and pilin-like proteins from *Clostridium difficile*. *Pathog. Dis.* **71**, 302–314 (2014).
158. Purcell, E. B., McKee, R. W., Bordeleau, E., Burrus, V. & Tamayo, R. Regulation of Type IV Pili Contributes to Surface Behaviors of Historical and Epidemic Strains of *Clostridium difficile*. *J. Bacteriol.* **198**, 565–577 (2016).
159. Bordeleau, E. *et al.* Cyclic Di-GMP Riboswitch-Regulated Type IV Pili Contribute to Aggregation of *Clostridium difficile*. *J. Bacteriol.* **197**, 819–832 (2015).
160. Maldarelli, G. A. *et al.* Type IV pili promote early biofilm formation by *Clostridium difficile*. *Pathog. Dis.* **74**, ftw061 (2016).
161. Saujet, L., Monot, M., Dupuy, B., Soutourina, O. & Martin-Verstraete, I. The Key Sigma Factor of Transition Phase, SigH, Controls Sporulation, Metabolism, and Virulence Factor Expression in *Clostridium difficile*. *J. Bacteriol.* **193**, 3186–3196 (2011).
162. Tasteyre, A. *et al.* A *Clostridium difficile* gene encoding flagellin The GenBank accession numbers for the sequences reported in this paper are AF065259 (strain 79-685) and AF077341 (strain VPI 10463). *Microbiology* vol. 146 957–966 (2000).
163. Tasteyre, A. *et al.* Molecular Characterization of *fliD* Gene Encoding Flagellar Cap and Its Expression among *Clostridium difficile* Isolates from Different Serogroups. *J. Clin. Microbiol.* **39**, 1178–1183 (2001).
164. Twine, S. M. *et al.* Motility and Flagellar Glycosylation in *Clostridium difficile*. *J. BACTERIOL* **191**, 13 (2009).

165. Faulds-Pain, A. *et al.* The post-translational modification of the *C. lostridium difficile* flagellin affects motility, cell surface properties and virulence. *Mol. Microbiol.* **94**, 272–289 (2014).
166. Abt, M. C., McKenney, P. T. & Pamer, E. G. Clostridium difficile colitis: pathogenesis and host defence. *Nat. Rev. Microbiol.* **14**, 609–620 (2016).
167. Tasteyre, A., Barc, M.-C., Collignon, A., Boureau, H. & Karjalainen, T. Role of FliC and FliD Flagellar Proteins of *Clostridium difficile* in Adherence and Gut Colonization. *Infect. Immun.* **69**, 7937–7940 (2001).
168. Dingle, T. C., Mulvey, G. L. & Armstrong, G. D. Mutagenic Analysis of the Clostridium difficile Flagellar Proteins, FliC and FliD, and Their Contribution to Virulence in Hamsters. *Infect. Immun.* **79**, 4061–4067 (2011).
169. Baban, S. T. *et al.* The Role of Flagella in Clostridium difficile Pathogenesis: Comparison between a Non-Epidemic and an Epidemic Strain. *PLoS ONE* **8**, e73026 (2013).
170. Aubry, A. *et al.* Modulation of Toxin Production by the Flagellar Regulon in Clostridium difficile. *Infect. Immun.* **80**, 3521–3532 (2012).
171. Anjuwon-Foster, B. R. & Tamayo, R. A genetic switch controls the production of flagella and toxins in Clostridium difficile. *PLOS Genet.* **13**, e1006701 (2017).
172. Bordeleau, E. & Burrus, V. Cyclic-di-GMP signaling in the Gram-positive pathogen Clostridium difficile. *Curr. Genet.* **61**, 497–502 (2015).
173. Sudarsan, N. *et al.* Riboswitches in Eubacteria Sense the Second Messenger Cyclic Di-GMP. *Science* **321**, 411–413 (2008).
174. Soutourina, O. A. *et al.* Genome-Wide Identification of Regulatory RNAs in the Human Pathogen Clostridium difficile. *PLoS Genet.* **9**, e1003493 (2013).
175. Purcell, E. B., McKee, R. W., McBride, S. M., Waters, C. M. & Tamayo, R. Cyclic Diguanylate Inversely Regulates Motility and Aggregation in Clostridium difficile. *J. Bacteriol.* **194**, 3307–3316 (2012).
176. McKee, R. W., Mangalea, M. R., Purcell, E. B., Borchardt, E. K. & Tamayo, R. The Second Messenger Cyclic Di-GMP Regulates Clostridium difficile Toxin Production by Controlling Expression of *sigD*. *J. Bacteriol.* **195**, 5174–5185 (2013).
177. El Meouche, I. *et al.* Characterization of the SigD Regulon of C. difficile and Its Positive Control of Toxin Production through the Regulation of *tcdR*. *PLoS ONE* **8**, e83748 (2013).
178. Dapa, T. & Unnikrishnan, M. Biofilm formation by Clostridium difficile. *Gut Microbes* **4**, 397–402 (2013).
179. Vuotto, C., Donelli, G., Buckley, A. & Chilton, C. Clostridium difficile Biofilm. in *Updates on Clostridium difficile in Europe: Advances in Microbiology, Infectious Diseases and Public Health Volume 8* (eds. Mastrantonio, P. & Rupnik, M.) 97–115 (Springer International Publishing, 2018). doi:10.1007/978-3-319-72799-8_7.
180. Pantaléon, V. *et al.* Clostridium difficile forms variable biofilms on abiotic surface. *Anaerobe* **53**, 34–37 (2018).
181. Taggart, M. G. *et al.* Biofilm regulation in Clostridioides difficile: Novel systems linked to hypervirulence. *PLOS Pathog.* **17**, e1009817 (2021).
182. Semenyuk, E. G. *et al.* Spore Formation and Toxin Production in Clostridium difficile Biofilms. *PLoS ONE* **9**, e87757 (2014).

183. Pantaléon, V. *et al.* The *Clostridium difficile* Protease Cwp84 Modulates both Biofilm Formation and Cell-Surface Properties. *PLOS ONE* **10**, e0124971 (2015).
184. Dawson, L. F., Valiente, E., Faulds-Pain, A., Donahue, E. H. & Wren, B. W. Characterisation of *Clostridium difficile* Biofilm Formation, a Role for Spo0A. *PLoS ONE* **7**, e50527 (2012).
185. Thapa, T. *et al.* Multiple Factors Modulate Biofilm Formation by the Anaerobic Pathogen *Clostridium difficile*. *J. Bacteriol.* **195**, 545–555 (2013).
186. Lawley, T. D. *et al.* Antibiotic Treatment of *Clostridium difficile* Carrier Mice Triggers a Supershedder State, Spore-Mediated Transmission, and Severe Disease in Immunocompromised Hosts. *Infect. Immun.* **77**, 3661–3669 (2009).
187. Soavelomandroso, A. P. *et al.* Biofilm Structures in a Mono-Associated Mouse Model of *Clostridium difficile* Infection. *Front. Microbiol.* **8**, 2086 (2017).
188. Semenyuk, E. G. *et al.* Analysis of Bacterial Communities during *Clostridium difficile* Infection in the Mouse. *Infect. Immun.* **83**, 4383–4391 (2015).
189. Slater, R. T., Frost, L. R., Jossi, S. E., Millard, A. D. & Unnikrishnan, M. *Clostridioides difficile* LuxS mediates inter-bacterial interactions within biofilms. *Sci. Rep.* **9**, 9903 (2019).
190. Normington, C. *et al.* Biofilms harbour *Clostridioides difficile*, serving as a reservoir for recurrent infection. *Npj Biofilms Microbiomes* **7**, 16 (2021).
191. Crowther, G. S. *et al.* Development and Validation of a Chemostat Gut Model To Study Both Planktonic and Biofilm Modes of Growth of *Clostridium difficile* and Human Microbiota. *PLoS ONE* **9**, e88396 (2014).
192. Frost, L. R., Cheng, J. K. J. & Unnikrishnan, M. *Clostridioides difficile* biofilms: A mechanism of persistence in the gut? *PLoS Pathog.* **17**, e1009348 (2021).
193. Hatheway, C. L. Toxigenic clostridia. *Clin. Microbiol. Rev.* **3**, 66–98 (1990).
194. Hammond, G. A. & Johnson, J. L. The toxigenic element of *Clostridium difficile* strain VPI 10463. *Microb. Pathog.* **19**, 203–213 (1995).
195. Braun, V., Hundsberger, T., Leukel, P., Sauerborn, M. & von Eichel-Streiber, C. Definition of the single integration site of the pathogenicity locus in *Clostridium difficile*. *Gene* **181**, 29–38 (1996).
196. Govind, R. & Dupuy, B. Secretion of *Clostridium difficile* Toxins A and B Requires the Holin-like Protein TcdE. *PLoS Pathog.* **8**, e1002727 (2012).
197. TAN, K. S., WEE, B. Y. & SONG, K. P. Evidence for holin function of tcdE gene in the pathogenicity of *Clostridium difficile*. *Journal of Medical Microbiology* vol. 50 613–619 (2001).
198. Mani, N. & Dupuy, B. Regulation of toxin synthesis in *Clostridium difficile* by an alternative RNA polymerase sigma factor. *Proc. Natl. Acad. Sci.* **98**, 5844–5849 (2001).
199. Martin-Verstraete, I., Peltier, J. & Dupuy, B. The Regulatory Networks That Control *Clostridium difficile* Toxin Synthesis. *Toxins* **8**, 153 (2016).
200. Brouwer, M. S. M. *et al.* Horizontal gene transfer converts non-toxigenic *Clostridium difficile* strains into toxin producers. *Nat. Commun.* **4**, 2601 (2013).
201. Moncrief J S, Barroso L A, & Wilkins T D. Positive regulation of *Clostridium difficile* toxins. *Infect. Immun.* **65**, 1105–1108 (1997).
202. Mani, N. *et al.* Environmental Response and Autoregulation of *Clostridium difficile* TxeR, a Sigma Factor for Toxin Gene Expression. *J. Bacteriol.* **184**, 5971–5978 (2002).

203. Govind, R., VEDIYAPPAN, G., Rolfe, R. D. & Fralick, J. A. Evidence that *Clostridium difficile* TcdC Is a Membrane-Associated Protein. *J. Bacteriol.* **188**, 3716–3720 (2006).
204. Matamouros, S., England, P. & Dupuy, B. Clostridium difficile toxin expression is inhibited by the novel regulator TcdC: TcdC inhibits C. difficile toxin expression. *Mol. Microbiol.* **64**, 1274–1288 (2007).
205. van Leeuwen, H. C., Bakker, D., Steindel, P., Kuijper, E. J. & Corver, J. Clostridium difficile TcdC protein binds four-stranded G-quadruplex structures. *Nucleic Acids Res.* **41**, 2382–2393 (2013).
206. Hundsberger, T. *et al.* Transcription Analysis of the Genes tcdA-E of the Pathogenicity Locus of Clostridium Difficile. *Eur. J. Biochem.* **244**, 735–742 (1997).
207. Karlsson, S. *et al.* Expression of *Clostridium difficile* Toxins A and B and Their Sigma Factor TcdD Is Controlled by Temperature. *Infect. Immun.* **71**, 1784–1793 (2003).
208. Karlsson, S., Burman, L. G. & Åkerlund, T. Suppression of toxin production in Clostridium difficile VPI 10463 by amino acids. *Microbiology* **145**, 1683–1693 (1999).
209. Karlsson, S., Burman, L. G. & Åkerlund, T. Induction of toxins in Clostridium difficile is associated with dramatic changes of its metabolism. *Microbiology* **154**, 3430–3436 (2008).
210. Karlsson, S., Lindberg, A., Norin, E., Burman, L. G. & Åkerlund, T. Toxins, Butyric Acid, and Other Short-Chain Fatty Acids Are Coordinately Expressed and Down-Regulated by Cysteine in *Clostridium difficile*. *Infect. Immun.* **68**, 5881–5888 (2000).
211. Dineen, S. S., Villapakkam, A. C., Nordman, J. T. & Sonenshein, A. L. Repression of Clostridium difficile toxin gene expression by CodY. *Mol. Microbiol.* **66**, 206–219 (2007).
212. Dineen, S. S., McBride, S. M. & Sonenshein, A. L. Integration of Metabolism and Virulence by *Clostridium difficile* CodY. *J. Bacteriol.* **192**, 5350–5362 (2010).
213. Sonenshein, A. L. CodY, a global regulator of stationary phase and virulence in Gram-positive bacteria. *Curr. Opin. Microbiol.* **8**, 203–207 (2005).
214. Antunes, A., Martin-Verstraete, I. & Dupuy, B. CcpA-mediated repression of Clostridium difficile toxin gene expression: C. difficile toxin regulation by CcpA. *Mol. Microbiol.* **79**, 882–899 (2011).
215. Antunes, A. *et al.* Global transcriptional control by glucose and carbon regulator CcpA in Clostridium difficile. *Nucleic Acids Res.* **40**, 10701–10718 (2012).
216. Dupuy, B. & Sonenshein, A. L. Regulated transcription of *Clostridium difficile* toxin genes. *Mol. Microbiol.* **27**, 107–120 (1998).
217. Mackin, K. E., Carter, G. P., Howarth, P., Rood, J. I. & Lyras, D. Spo0A Differentially Regulates Toxin Production in Evolutionarily Diverse Strains of Clostridium difficile. *PLoS ONE* **8**, e79666 (2013).
218. Pettit, L. J. *et al.* Functional genomics reveals that Clostridium difficile Spo0A coordinates sporulation, virulence and metabolism. *BMC Genomics* **15**, 160 (2014).
219. Rosenbusch, K. E., Bakker, D., Kuijper, E. J. & Smits, W. K. C. difficile 630Δerm Spo0A Regulates Sporulation, but Does Not Contribute to Toxin Production, by Direct High-Affinity Binding to Target DNA. *PLoS ONE* **7**, e48608 (2012).
220. Edwards, A. N., Tamayo, R. & McBride, S. M. A novel regulator controls *Clostridium difficile* sporulation, motility and toxin production: A novel regulator of *C. difficile* physiology. *Mol.*

Microbiol. **100**, 954–971 (2016).

221. Papenfort, K. & Bassler, B. L. Quorum sensing signal–response systems in Gram-negative bacteria. *Nat. Rev. Microbiol.* **14**, 576–588 (2016).
222. Martin, M. J. *et al.* The *agr* Locus Regulates Virulence and Colonization Genes in *Clostridium difficile* 027. *J. Bacteriol.* **195**, 3672–3681 (2013).
223. Darkoh, C., DuPont, H. L., Norris, S. J. & Kaplan, H. B. Toxin Synthesis by *Clostridium difficile* Is Regulated through Quorum Signaling. *mBio* **6**, (2015).
224. Bouillaut, L., Dubois, T., Sonenshein, A. L. & Dupuy, B. Integration of metabolism and virulence in *Clostridium difficile*. *Res. Microbiol.* **166**, 375–383 (2015).
225. Mitchell, T. J. *et al.* Effect of toxin A and B of *Clostridium difficile* on rabbit ileum and colon. *Gut* **27**, 78 (1986).
226. Lyerly, D. M., Lockwood, D. E., Richardson, S. H. & Wilkins, T. D. Biological activities of toxins A and B of *Clostridium difficile*. *Infect. Immun.* **35**, 1147–1150 (1982).
227. Lyras, D. *et al.* Toxin B is essential for virulence of *Clostridium difficile*. *Nature* **458**, 1176–1179 (2009).
228. Kuehne, S. A. *et al.* The role of toxin A and toxin B in *Clostridium difficile* infection. *Nature* **467**, 711–713 (2010).
229. Chumbler, N. M., Farrow, M. A., Lapierre, L. A., Franklin, J. L. & Lacy, D. B. *Clostridium difficile* Toxins TcdA and TcdB Cause Colonic Tissue Damage by Distinct Mechanisms. *Infect. Immun.* **84**, 2871–2877 (2016).
230. Carter, G. P. *et al.* Defining the Roles of TcdA and TcdB in Localized Gastrointestinal Disease, Systemic Organ Damage, and the Host Response during *Clostridium difficile* Infections. *mBio* **6**, (2015).
231. Savidge, T. C. *et al.* *Clostridium difficile* toxin B is an inflammatory enterotoxin in human intestine. *Gastroenterology* **125**, 413–420 (2003).
232. Riegler, M. *et al.* *Clostridium difficile* toxin B is more potent than toxin A in damaging human colonic epithelium in vitro. *J. Clin. Invest.* **95**, 2004–2011 (1995).
233. Skinner, A. M. *et al.* The Relative Role of Toxins A and B in the Virulence of *Clostridioides difficile*. *J. Clin. Med.* **10**, 96 (2020).
234. von Eichel-Streiber, C., Laufenberg-Feldmann, R., Saringen, S., Schulze, J. & Sauerborn, M. Comparative sequence analysis of the *Clostridium difficile* toxins A and B. *Mol. Gen. Genet. MGG* **233**, 260–268 (1992).
235. Shen, A. *Clostridium difficile* Toxins: Mediators of Inflammation. *J. Innate Immun.* **4**, 149–158 (2012).
236. Chumbler, N. M. *et al.* Crystal structure of *Clostridium difficile* toxin A. *Nat. Microbiol.* **1**, 15002 (2016).
237. Pruitt, R. N. & Lacy, D. B. Toward a structural understanding of *Clostridium difficile* toxins A and B. *Front. Cell. Infect. Microbiol.* **2**, (2012).
238. Papatheodorou, P., Zamboglou, C., Genisyurek, S., Guttenberg, G. & Aktories, K. Clostridial Glucosylating Toxins Enter Cells via Clathrin-Mediated Endocytosis. *PLoS ONE* **5**, e10673 (2010).
239. Greco, A. *et al.* Carbohydrate recognition by *Clostridium difficile* toxin A. *Nat. Struct. Mol. Biol.* **13**, 460–461 (2006).

240. LaFrance, M. E. *et al.* Identification of an epithelial cell receptor responsible for *Clostridium difficile* TcdB-induced cytotoxicity. *Proc. Natl. Acad. Sci.* **112**, 7073–7078 (2015).
241. Yuan, P. *et al.* Chondroitin sulfate proteoglycan 4 functions as the cellular receptor for *Clostridium difficile* toxin B. *Cell Res.* **25**, 157–168 (2015).
242. Terada, N. *et al.* Immunohistochemical study of NG2 chondroitin sulfate proteoglycan expression in the small and large intestines. *Histochem. Cell Biol.* **126**, 483–490 (2006).
243. Olling, A. *et al.* The Repetitive Oligopeptide Sequences Modulate Cytopathic Potency but Are Not Crucial for Cellular Uptake of *Clostridium difficile* Toxin A. *PLoS ONE* **6**, e17623 (2011).
244. Genisyurek, S. *et al.* Structural determinants for membrane insertion, pore formation and translocation of *Clostridium difficile* toxin B: Pore formation of *C. difficile* toxin B. *Mol. Microbiol.* **79**, 1643–1654 (2011).
245. Sun, X. & Hirota, S. A. The roles of host and pathogen factors and the innate immune response in the pathogenesis of *Clostridium difficile* infection. *Mol. Immunol.* **63**, 193–202 (2015).
246. Gerding, D. N., Johnson, S., Rupnik, M. & Aktories, K. *Clostridium difficile* binary toxin CDT: Mechanism, epidemiology, and potential clinical importance. *Gut Microbes* **5**, 15–27 (2014).
247. Perelle, S., Gibert, M., Bourlioux, P., Corthier, G. & Popoff, M. R. Production of a complete binary toxin (actin-specific ADP-ribosyltransferase) by *Clostridium difficile* CD196. *Infect. Immun.* **65**, 1402–1407 (1997).
248. Gonçalves, C., Decré, D., Barbut, F., Burghoffer, B. & Petit, J.-C. Prevalence and Characterization of a Binary Toxin (Actin-Specific ADP-Ribosyltransferase) from *Clostridium difficile*. *J. Clin. Microbiol.* **42**, 1933–1939 (2004).
249. Stare, B. G., Delmée, M. & Rupnik, M. Variant forms of the binary toxin CDT locus and tcdC gene in *Clostridium difficile* strains. *J. Med. Microbiol.* **56**, 329–335 (2007).
250. Carter, G. P. *et al.* Binary Toxin Production in *Clostridium difficile* Is Regulated by CdtR, a LytTR Family Response Regulator. *J. Bacteriol.* **189**, 7290–7301 (2007).
251. Metcalf, D. S. & Weese, J. S. Binary toxin locus analysis in *Clostridium difficile*. *J. Med. Microbiol.* **60**, 1137–1145 (2011).
252. Schwan, C. *et al.* Cholesterol- and Sphingolipid-rich Microdomains Are Essential for Microtubule-based Membrane Protrusions Induced by *Clostridium difficile* Transferase (CDT). *J. Biol. Chem.* **286**, 29356–29365 (2011).
253. Eckert, C. *et al.* Prevalence and pathogenicity of binary toxin-positive *Clostridium difficile* strains that do not produce toxins A and B. *New Microbes New Infect.* **3**, 12–17 (2015).
254. Androga, G. O. *et al.* Infection with Toxin A-Negative, Toxin B-Negative, Binary Toxin-Positive *Clostridium difficile* in a Young Patient with Ulcerative Colitis. *J. Clin. Microbiol.* **53**, 3702–3704 (2015).
255. McGovern, A. M. *et al.* Prevalence of binary toxin positive *Clostridium difficile* in diarrhoeal humans in the absence of epidemic ribotype 027. *PLOS ONE* **12**, e0187658 (2017).
256. Peng, Z. *et al.* Genome characterization of a novel binary toxin-positive strain of *Clostridium difficile* and comparison with the epidemic 027 and 078 strains. *Gut Pathog.* **9**, 42 (2017).
257. Lawley, T. D. *et al.* Use of Purified *Clostridium difficile* Spores To Facilitate Evaluation of Health Care Disinfection Regimens. *Appl. Environ. Microbiol.* **76**, 6895–6900 (2010).

258. Rodriguez-Palacios, A. & LeJeune, J. T. Moist-Heat Resistance, Spore Aging, and Superdormancy in *Clostridium difficile*. *Appl. Environ. Microbiol.* **77**, 3085–3091 (2011).
259. Dawson, L. F., Valiente, E., Donahue, E. H., Birchenough, G. & Wren, B. W. Hypervirulent *Clostridium difficile* PCR-Ribotypes Exhibit Resistance to Widely Used Disinfectants. *PLoS ONE* **6**, e25754 (2011).
260. Edwards, A. N. & McBride, S. M. Initiation of sporulation in *Clostridium difficile*: a twist on the classic model. *FEMS Microbiol. Lett.* **358**, 110–118 (2014).
261. Underwood, S. *et al.* Characterization of the Sporulation Initiation Pathway of *Clostridium difficile* and Its Role in Toxin Production. *J. Bacteriol.* **191**, 7296–7305 (2009).
262. Galperin, M. Y. Genome Diversity of Spore-Forming Firmicutes. *Microbiol. Spectr.* **1**, TBS-0015-2012 (2013).
263. Wunderlin, T., Junier, T., Roussel-Delif, L., Jeanneret, N. & Junier, P. Stage 0 sporulation gene A as a molecular marker to study diversity of endospore-forming Firmicutes: Culture-independent study of endospore-forming Firmicutes. *Environ. Microbiol. Rep.* **5**, 911–924 (2013).
264. Deakin, L. J. *et al.* The *Clostridium difficile spo0A* Gene Is a Persistence and Transmission Factor. *Infect. Immun.* **80**, 2704–2711 (2012).
265. Fimlaid, K. A. & Shen, A. Diverse mechanisms regulate sporulation sigma factor activity in the Firmicutes. *Curr. Opin. Microbiol.* **24**, 88–95 (2015).
266. Saujet, L., Pereira, F. C., Henriques, A. O. & Martin-Verstraete, I. The regulatory network controlling spore formation in *Clostridium difficile*. *FEMS Microbiol. Lett.* **358**, 1–10 (2014).
267. Fimlaid, K. A. *et al.* Global Analysis of the Sporulation Pathway of *Clostridium difficile*. *PLoS Genet.* **9**, e1003660 (2013).
268. Pereira, F. C. *et al.* The Spore Differentiation Pathway in the Enteric Pathogen *Clostridium difficile*. *PLoS Genet.* **9**, e1003782 (2013).
269. Koenigsnecht, M. J. *et al.* Dynamics and Establishment of *Clostridium difficile* Infection in the Murine Gastrointestinal Tract. *Infect. Immun.* **83**, 934–941 (2015).
270. Wilson, K. H. Efficiency of various bile salt preparations for stimulation of *Clostridium difficile* spore germination. *J. Clin. Microbiol.* **18**, 1017–1019 (1983).
271. Sorg, J. A. & Sonenshein, A. L. Bile Salts and Glycine as Cogermnants for *Clostridium difficile* Spores. *J. Bacteriol.* **190**, 2505–2512 (2008).
272. Francis, M. B., Allen, C. A. & Sorg, J. A. Spore Cortex Hydrolysis Precedes Dipicolinic Acid Release during *Clostridium difficile* Spore Germination. *J. Bacteriol.* **197**, 2276–2283 (2015).
273. Wang, S., Shen, A., Setlow, P. & Li, Y. Characterization of the Dynamic Germination of Individual *Clostridium difficile* Spores Using Raman Spectroscopy and Differential Interference Contrast Microscopy. *J. Bacteriol.* **197**, 2361–2373 (2015).
274. Adams, C. M., Eckenroth, B. E., Putnam, E. E., Doublé, S. & Shen, A. Structural and Functional Analysis of the CspB Protease Required for *Clostridium* Spore Germination. *PLoS Pathog.* **9**, e1003165 (2013).
275. Kevorkian, Y., Shirley, D. J. & Shen, A. Regulation of *Clostridium difficile* spore germination by the CspA pseudoprotease domain. *Biochimie* **122**, 243–254 (2016).
276. Burns, D. A., Heap, J. T. & Minton, N. P. SleC Is Essential for Germination of *Clostridium*

- difficile* Spores in Nutrient-Rich Medium Supplemented with the Bile Salt Taurocholate. *J. Bacteriol.* **192**, 657–664 (2010).
277. Carlson, P. E. *et al.* Variation in germination of *Clostridium difficile* clinical isolates correlates to disease severity. *Anaerobe* **33**, 64–70 (2015).
278. Moore, P., Kyne, L., Martin, A. & Solomon, K. Germination efficiency of clinical *Clostridium difficile* spores and correlation with ribotype, disease severity and therapy failure. *J. Med. Microbiol.* **62**, 1405–1413 (2013).
279. Madan, R. & Petri, W. A., Jr. Immune responses to *Clostridium difficile* infection. *Trends Mol. Med.* **18**, 658–666 (2012).
280. Kelly, C. P. & Kyne, L. The host immune response to *Clostridium difficile*. *J. Med. Microbiol.* **60**, 1070–1079 (2011).
281. He, D. *et al.* *Clostridium difficile* toxin A triggers human colonocyte IL-8 release via mitochondrial oxygen radical generation. *Gastroenterology* **122**, 1048–1057 (2002).
282. Branka, J. *et al.* Early functional effects of *Clostridium difficile* toxin A on human colonocytes. *Gastroenterology* **112**, 1887–1894 (1997).
283. Ryan, A. *et al.* A Role for TLR4 in *Clostridium difficile* Infection and the Recognition of Surface Layer Proteins. *PLoS Pathog.* **7**, e1002076 (2011).
284. Hasegawa, M. *et al.* Nucleotide-Binding Oligomerization Domain 1 Mediates Recognition of *Clostridium difficile* and Induces Neutrophil Recruitment and Protection against the Pathogen. *J. Immunol.* **186**, 4872–4880 (2011).
285. Jarchum, I., Liu, M., Lipuma, L. & Pamer, E. G. Toll-Like Receptor 5 Stimulation Protects Mice from Acute *Clostridium difficile* Colitis. *Infect. Immun.* **79**, 1498–1503 (2011).
286. Kim, J. M. *et al.* Differential Expression and Polarized Secretion of CXC and CC Chemokines by Human Intestinal Epithelial Cancer Cell Lines in Response to *Clostridium difficile* Toxin A. *Microbiol. Immunol.* **46**, 333–342 (2002).
287. Ishida, Y. *et al.* Essential Involvement of IFN- γ in *Clostridium difficile* Toxin A-Induced Enteritis. *J. Immunol.* **172**, 3018–3025 (2004).
288. Castagliuolo, I. *et al.* *Clostridium difficile* Toxin A Stimulates Macrophage- Inflammatory Protein-2 Production in Rat Intestinal Epithelial Cells. *J. Immunol.* **160**, 6039 (1998).
289. Mykoniatis, A. *et al.* Leptin mediates *Clostridium difficile* toxin A-induced enteritis in mice. *Gastroenterology* **124**, 683–691 (2003).
290. Kelly, C. P. *et al.* Neutrophil recruitment in *Clostridium difficile* toxin A enteritis in the rabbit. *J. Clin. Invest.* **93**, 1257–1265 (1994).
291. Ng, J. *et al.* *Clostridium difficile* Toxin-Induced Inflammation and Intestinal Injury Are Mediated by the Inflammasome. *Gastroenterology* **139**, 542-552.e3 (2010).
292. Cowardin, C. A. *et al.* The binary toxin CDT enhances *Clostridium difficile* virulence by suppressing protective colonic eosinophilia. *Nat. Microbiol.* **1**, 16108 (2016).
293. Bianco, M. *et al.* Immunomodulatory activities of surface-layer proteins obtained from epidemic and hypervirulent *Clostridium difficile* strains. *J. Med. Microbiol.* **60**, 1162–1167 (2011).
294. Batah, J. *et al.* *Clostridium difficile* flagella predominantly activate TLR5-linked NF- κ B pathway in epithelial cells. *Anaerobe* **38**, 116–124 (2016).

295. Davis, K. M. & Weiser, J. N. Modifications to the Peptidoglycan Backbone Help Bacteria To Establish Infection. *Infect. Immun.* **79**, 562–570 (2011).
296. Peltier, J. *et al.* Clostridium difficile Has an Original Peptidoglycan Structure with a High Level of N-Acetylglucosamine Deacetylation and Mainly 3-3 Cross-links. *J. Biol. Chem.* **286**, 29053–29062 (2011).
297. Ho, T. D. *et al.* Clostridium difficile Extracytoplasmic Function σ Factor σ^V Regulates Lysozyme Resistance and Is Necessary for Pathogenesis in the Hamster Model of Infection. *Infect. Immun.* **82**, 2345–2355 (2014).
298. Revilla-Guarinos, A., Gebhard, S., Mascher, T. & Zúñiga, M. Defence against antimicrobial peptides: different strategies in *Firmicutes*. Antimicrobial peptide resistance in *Firmicutes*. *Environ. Microbiol.* **16**, 1225–1237 (2014).
299. Furci, L. *et al.* New Role for Human α -Defensin 5 in the Fight against Hypervirulent Clostridium difficile Strains. *Infect. Immun.* **83**, 986–995 (2015).
300. McQuade, R., Roxas, B., Viswanathan, V. K. & Vedantam, G. Clostridium difficile clinical isolates exhibit variable susceptibility and proteome alterations upon exposure to mammalian cationic antimicrobial peptides. *Anaerobe* **18**, 614–620 (2012).
301. Giesemann, T., Guttenberg, G. & Aktories, K. Human α -Defensins Inhibit Clostridium difficile Toxin B. *Gastroenterology* **134**, 2049–2058 (2008).
302. Hing, T. C. *et al.* The antimicrobial peptide cathelicidin modulates *Clostridium difficile* - associated colitis and toxin A-mediated enteritis in mice. *Gut* **62**, 1295–1305 (2013).
303. Clokie, M. R., Millard, A. D., Letarov, A. V. & Heaphy, S. Phages in nature. *Bacteriophage* **1**, 31–45 (2011).
304. Salmond, G. P. C. & Fineran, P. C. A century of the phage: past, present and future. *Nat. Rev. Microbiol.* **13**, 777–786 (2015).
305. E.V. Orlova ED1 - Ipek Kurtboke. Bacteriophages and Their Structural Organisation. in *Bacteriophages* Ch. 1 (IntechOpen, 2012). doi:10.5772/34642.
306. Ackermann, H.-W. Phage Classification and Characterization. in *Bacteriophages: Methods and Protocols, Volume 1: Isolation, Characterization, and Interactions* (eds. Clokie, M. R. J. & Kropinski, A. M.) 127–140 (Humana Press, 2009). doi:10.1007/978-1-60327-164-6_13.
307. Lefkowitz, E. J. *et al.* Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Res.* **46**, D708–D717 (2018).
308. Molineux, I. J. & Panja, D. Popping the cork: mechanisms of phage genome ejection. *Nat. Rev. Microbiol.* **11**, 194–204 (2013).
309. Young, R. Phage lysis: do we have the hole story yet? *Curr. Opin. Microbiol.* **16**, 790–797 (2013).
310. Rakonjac, J., Bennett, N. J., Spagnuolo, J., Gagic, D. & Russel, M. Filamentous bacteriophage: biology, phage display and nanotechnology applications. *Curr. Issues Mol. Biol.* **13**, 51–76 (2011).
311. Dy, R. L., Richter, C., Salmond, G. P. C. & Fineran, P. C. Remarkable Mechanisms in Microbes to Resist Phage Infections. *Annu. Rev. Virol.* **1**, 307–331 (2014).
312. Weinbauer, M. G. Ecology of prokaryotic viruses. *FEMS Microbiol. Rev.* **28**, 127–181 (2004).
313. Van Valen, L. Molecular evolution as predicted by natural selection. *J. Mol. Evol.* **3**, 89–101

(1974).

314. Labrie, S. J., Samson, J. E. & Moineau, S. Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.* **8**, 317–327 (2010).
315. Samson, J. E., Magadán, A. H., Sabri, M. & Moineau, S. Revenge of the phages: defeating bacterial defences. *Nat. Rev. Microbiol.* **11**, 675–687 (2013).
316. Sun, X., Göhler, A., Heller, K. J. & Neve, H. The *ltp* gene of temperate *Streptococcus thermophilus* phage TP-J34 confers superinfection exclusion to *Streptococcus thermophilus* and *Lactococcus lactis*. *Virology* **350**, 146–157 (2006).
317. Mahony, J., McGrath, S., Fitzgerald, G. F. & van Sinderen, D. Identification and Characterization of Lactococcal-Prophage-Carried Superinfection Exclusion Genes. *Appl. Environ. Microbiol.* **74**, 6206–6215 (2008).
318. Owen, S. V. *et al.* A window into lysogeny: revealing temperate phage biology with transcriptomics. *Microb. Genomics* **6**, (2020).
319. Bernheim, A. & Sorek, R. The pan-immune system of bacteria: antiviral defence as a community resource. *Nat. Rev. Microbiol.* **18**, 113–119 (2020).
320. Hille, F. *et al.* The Biology of CRISPR-Cas: Backward and Forward. *Cell* **172**, 1239–1259 (2018).
321. Koonin, E. V., Makarova, K. S. & Wolf, Y. I. Evolutionary Genomics of Defense Systems in Archaea and Bacteria. *Annu. Rev. Microbiol.* **71**, 233–261 (2017).
322. Oliveira, P. H., Touchon, M. & Rocha, E. P. C. The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res.* **42**, 10618–10631 (2014).
323. Tock, M. R. & Dryden, D. T. The biology of restriction and anti-restriction. *Curr. Opin. Microbiol.* **8**, 466–472 (2005).
324. Thiaville, J. J. *et al.* Novel genomic island modifies DNA with 7-deazaguanine derivatives. *Proc. Natl. Acad. Sci.* **113**, E1452–E1459 (2016).
325. Goldfarb, T. *et al.* BREX is a novel phage resistance system widespread in microbial genomes. *EMBO J.* **34**, 169–183 (2015).
326. Ofir, G. *et al.* DISARM is a widespread bacterial defence system with broad anti-phage activities. *Nat. Microbiol.* **3**, 90–98 (2018).
327. Amitai, G. & Sorek, R. CRISPR–Cas adaptation: insights into the mechanism of action. *Nat. Rev. Microbiol.* **14**, 67–76 (2016).
328. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR–Cas systems. *Nat. Rev. Microbiol.* **13**, 722–736 (2015).
329. Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Classification and Nomenclature of CRISPR–Cas Systems: Where from Here? *CRISPR J.* **1**, 325–336 (2018).
330. Malina, A. *et al.* PAM multiplicity marks genomic target sites as inhibitory to CRISPR–Cas9 editing. *Nat. Commun.* **6**, 10124 (2015).
331. Lopatina, A., Tal, N. & Sorek, R. Abortive Infection: Bacterial Suicide as an Antiviral Immune Strategy. *Annu. Rev. Virol.* **7**, 371–384 (2020).
332. Chopin, M.-C., Chopin, A. & Bidnenko, E. Phage abortive infection in lactococci: variations on a theme. *Curr. Opin. Microbiol.* **8**, 473–479 (2005).

333. Rostøl, J. T. & Marraffini, L. (Ph)ighting Phages: How Bacteria Resist Their Parasites. *Cell Host Microbe* **25**, 184–194 (2019).
334. Makarova, K. S., Anantharaman, V., Aravind, L. & Koonin, E. V. Live virus-free or die: coupling of antiviral immunity and programmed suicide or dormancy in prokaryotes. *Biol. Direct* **7**, 40 (2012).
335. Parma, D. H. *et al.* The Rex system of bacteriophage lambda: tolerance and altruistic cell death. *Genes Dev.* **6**, 497–510 (1992).
336. Snyder, L. Phage-exclusion enzymes: a bonanza of biochemical and cell biology reagents? *Mol. Microbiol.* **15**, 415–420 (1995).
337. Cohen, D. *et al.* Cyclic GMP–AMP signalling protects bacteria against viral infection. *Nature* **574**, 691–695 (2019).
338. Doron, S. *et al.* Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* **359**, eaar4120 (2018).
339. Kronheim, S. *et al.* A chemical defence against phage infection. *Nature* **564**, 283–286 (2018).
340. Fortier, L.-C. & Sekulovic, O. Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence* **4**, 354–365 (2013).
341. Brüssow, H., Canchaya, C. & Hardt, W.-D. Phages and the Evolution of Bacterial Pathogens: from Genomic Rearrangements to Lysogenic Conversion. *Microbiol. Mol. Biol. Rev.* **68**, 560–602 (2004).
342. Ohnishi, M., Kurokawa, K. & Hayashi, T. Diversification of Escherichia coli genomes: are bacteriophages the major contributors? *Trends Microbiol.* **9**, 481–485 (2001).
343. Boyd, E. F., Heilpern, A. J. & Waldor, M. K. Molecular Analyses of a Putative CTX ϕ Precursor and Evidence for Independent Acquisition of Distinct CTX ϕ s by Toxigenic Vibrio cholerae. *J BACTERIOL* **182**, 9 (2000).
344. Mills, S. *et al.* Movers and shakers: Influence of bacteriophages in shaping the mammalian gut microbiota. *Gut Microbes* **4**, 4–16 (2013).
345. Taylor, V. L., Fitzpatrick, A. D., Islam, Z. & Maxwell, K. L. Chapter One - The Diverse Impacts of Phage Morons on Bacterial Fitness and Virulence. in *Advances in Virus Research* (eds. Kielian, M., Mettenleiter, T. C. & Roossinck, M. J.) vol. 103 1–31 (Academic Press, 2019).
346. Cumby, N., Davidson, A. R. & Maxwell, K. L. The moron comes of age. *Bacteriophage* **2**, e23146 (2012).
347. Juhala, R. J. *et al.* Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages 1 Edited by M. Gottesman. *J. Mol. Biol.* **299**, 27–51 (2000).
348. Fortier, L.-C. Bacteriophages Contribute to Shaping Clostridioides (Clostridium) difficile Species. *Front. Microbiol.* **9**, 2033 (2018).
349. Fortier, L.-C. Chapter Five - The Contribution of Bacteriophages to the Biology and Virulence of Pathogenic Clostridia. in *Advances in Applied Microbiology* (eds. Sariaslani, S. & Gadd, G. M.) vol. 101 169–200 (Academic Press, 2017).
350. Goh, S., Chang, B. J. & Riley, T. V. Effect of phage infection on toxin production by Clostridium difficile. *J. Med. Microbiol.* **54**, 129–135 (2005).
351. Sekulovic, O., Meessen-Pinard, M. & Fortier, L.-C. Prophage-Stimulated Toxin Production in

- Clostridium difficile* NAP1/027 Lysogens. *J. Bacteriol.* **193**, 2726–2734 (2011).
352. Govind, R., Vedyappan, G., Rolfe, R. D., Dupuy, B. & Fralick, J. A. Bacteriophage-Mediated Toxin Gene Regulation in *Clostridium difficile*. *J. Virol.* **83**, 12037–12045 (2009).
353. Heuler, J., Fortier, L.-C. & Sun, X. Clostridioides difficile phage biology and application. *FEMS Microbiol. Rev.* **45**, fuab012 (2021).
354. Garneau, J. R. *et al.* High Prevalence and Genetic Diversity of Large phiCD211 (phiCDIF1296T)-Like Prophages in Clostridioides difficile. *Appl. Environ. Microbiol.* **84**, (2018).
355. Maikova, A., Severinov, K. & Soutourina, O. New Insights Into Functions and Possible Applications of Clostridium difficile CRISPR-Cas System. *Front. Microbiol.* **9**, 1740 (2018).
356. Sekulovic, O., Ospina Bedoya, M., Fivian-Hughes, A. S., Fairweather, N. F. & Fortier, L. The *Clostridium difficile* cell wall protein CWPV confers phase-variable phage resistance. *Mol. Microbiol.* **98**, 329–342 (2015).
357. Riedel, T. *et al.* A Clostridioides difficile bacteriophage genome encodes functional binary toxin-associated genes. *J. Biotechnol.* **250**, 23–28 (2017).
358. Sekulovic, O. & Fortier, L.-C. Global Transcriptional Response of Clostridium difficile Carrying the ϕ CD38-2 Prophage. *Appl. Environ. Microbiol.* **81**, 1364–1374 (2015).
359. Li, T. *et al.* Isolation and Characterization of the Novel Phage JD032 and Global Transcriptomic Response during JD032 Infection of Clostridioides difficile Ribotype 078. *mSystems* **5**, (2020).
360. Hargreaves, K. R., Flores, C. O., Lawley, T. D. & Clokie, M. R. J. Abundant and Diverse Clustered Regularly Interspaced Short Palindromic Repeat Spacers in Clostridium difficile Strains and Prophages Target Multiple Phage Types within This Pathogen. *mBio* **5**, (2014).
361. Boudry, P. *et al.* Function of the CRISPR-Cas System of the Human Pathogen Clostridium difficile. *mBio* **6**, (2015).
362. Wagner, E. G. H. & Romby, P. Chapter Three - Small RNAs in Bacteria and Archaea: Who They Are, What They Do, and How They Do It. in *Advances in Genetics* (eds. Friedmann, T., Dunlap, J. C. & Goodwin, S. F.) vol. 90 133–208 (Academic Press, 2015).
363. Mattick, J. S. & Makunin, I. V. Non-coding RNA. *Hum. Mol. Genet.* **15**, R17–R29 (2006).
364. Cech, T. R. & Steitz, J. A. The Noncoding RNA Revolution—Trashing Old Rules to Forge New Ones. *Cell* **157**, 77–94 (2014).
365. Gottesman, S. Micros for microbes: non-coding regulatory RNAs in bacteria. *Trends Genet.* **21**, 399–404 (2005).
366. Tomizawa, J., Itoh, T., Selzer, G. & Som, T. Inhibition of ColE1 RNA primer formation by a plasmid-specified small RNA. *Proc. Natl. Acad. Sci.* **78**, 1421–1425 (1981).
367. Tomizawa, J. Control of colE1 plasmid replication: The process of binding of RNA I to the primer transcript. *Cell* **38**, 861–870 (1984).
368. Simons, R. W. & Kleckner, N. Translational control of IS10 transposition. *Cell* **34**, 683–691 (1983).
369. Mizuno, T., Chou, M. Y. & Inouye, M. A unique mechanism regulating gene expression: translational inhibition by a complementary RNA transcript (micRNA). *Proc. Natl. Acad. Sci.* **81**, 1966–1970 (1984).
370. Waters, L. S. & Storz, G. Regulatory RNAs in Bacteria. *Cell* **136**, 615–628 (2009).

371. Gottesman, S. The Small RNA Regulators of *Escherichia coli*: Roles and Mechanisms. *Annu. Rev. Microbiol.* **58**, 303–328 (2004).
372. Hershberg, R. A survey of small RNA-encoding genes in *Escherichia coli*. *Nucleic Acids Res.* **31**, 1813–1820 (2003).
373. Xu, P., Guo, M. & Hay, B. A. MicroRNAs and the regulation of cell death. *Trends Genet.* **20**, 617–624 (2004).
374. Lagos-Quintana, M. Identification of Novel Genes Coding for Small Expressed RNAs. **294**, 7 (2001).
375. Vanderpool, C. K., Balasubramanian, D. & Lloyd, C. R. Dual-function RNA regulators in bacteria. *Biochimie* **93**, 1943–1949 (2011).
376. Bronesky, D. *et al.* *Staphylococcus aureus* RNAIII and Its Regulon Link Quorum Sensing, Stress Responses, Metabolic Adaptation, and Regulation of Virulence Gene Expression. *Annu. Rev. Microbiol.* **70**, 299–316 (2016).
377. Carrier, M.-C., Lalaouna, D. & Massé, E. Broadening the Definition of Bacterial Small RNAs: Characteristics and Mechanisms of Action. *Annu. Rev. Microbiol.* **72**, 141–161 (2018).
378. Lalaouna, D., Carrier, M.-C. & Massé, E. Every little piece counts: the many faces of tRNA transcripts. *Transcription* **6**, 74–77 (2015).
379. Desgranges, E., Marzi, S., Moreau, K., Romby, P. & Caldelari, I. Noncoding RNA. *Microbiol. Spectr.* **7**, (2019).
380. Gottesman, S. & Storz, G. Bacterial Small RNA Regulators: Versatile Roles and Rapidly Evolving Variations. *Cold Spring Harb. Perspect. Biol.* **3**, a003798–a003798 (2011).
381. Richards, G. R. & Vanderpool, C. K. Molecular call and response: The physiology of bacterial small RNAs. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* **1809**, 525–531 (2011).
382. Gripenland, J. *et al.* RNAs: regulators of bacterial virulence. *Nat. Rev. Microbiol.* **8**, 857–866 (2010).
383. Masse, E., Vanderpool, C. K. & Gottesman, S. Effect of RyhB Small RNA on Global Iron Use in *Escherichia coli*. *J BACTERIOL* **187**, 10 (2005).
384. Massé, E. & Gottesman, S. A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc. Natl. Acad. Sci.* **99**, 4620 (2002).
385. Chen, S., Zhang, A., Blyn, L. B. & Storz, G. MicC, a second small-RNA regulator of Omp protein expression in *Escherichia coli*. *J. Bacteriol.* **186**, 6689–6697 (2004).
386. Andersen, J., Forst, S. A., Zhao, K., Inouye, M. & Delihias, N. The function of micF RNA: micF RNA is a major factor in the thermal regulation of OmpF protein in *Escherichia coli**. *J. Biol. Chem.* **264**, 17961–17970 (1989).
387. Gogol, E. B., Rhodius, V. A., Papenfort, K., Vogel, J. & Gross, C. A. Small RNAs endow a transcriptional activator with essential repressor functions for single-tier control of a global stress regulon. *Proc. Natl. Acad. Sci.* **108**, 12875 (2011).
388. Guo, M. S. *et al.* MicL, a new σ E-dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein. *Genes Dev.* **28**, 1620–1634 (2014).
389. Mika, F. & Hengge, R. Small Regulatory RNAs in the Control of Motility and Biofilm Formation in *E. coli* and *Salmonella*. *Int. J. Mol. Sci.* **14**, 4560–4579 (2013).
390. Gorke, B. & Vogel, J. Noncoding RNA control of the making and breaking of sugars. *Genes*

- Dev.* **22**, 2914–2925 (2008).
391. Moller, T. Spot 42 RNA mediates discoordinate expression of the *E. coli* galactose operon. *Genes Dev.* **16**, 1696–1706 (2002).
392. Beisel, C. L. & Storz, G. The Base-Pairing RNA Spot 42 Participates in a Multioutput Feedforward Loop to Help Enact Catabolite Repression in *Escherichia coli*. *Mol. Cell* **41**, 286–297 (2011).
393. Vanderpool, C. K. & Gottesman, S. Involvement of a novel transcriptional activator and small RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system: RNA regulator of phosphosugar stress response. *Mol. Microbiol.* **54**, 1076–1089 (2004).
394. Kawamoto, H. Implication of membrane localization of target mRNA in the action of a small RNA: mechanism of post-transcriptional regulation of glucose transporter in *Escherichia coli*. *Genes Dev.* **19**, 328–338 (2005).
395. Wadler, C. S. & Vanderpool, C. K. A dual function for a bacterial small RNA: SgrS performs base pairing-dependent regulation and encodes a functional polypeptide. *Proc. Natl. Acad. Sci.* **104**, 20454–20459 (2007).
396. Urban, J. H. & Vogel, J. Two Seemingly Homologous Noncoding RNAs Act Hierarchically to Activate *glmS* mRNA Translation. *PLoS Biol.* **6**, e64 (2008).
397. Göpel, Y., Khan, M. A. & Görke, B. Ménage à trois: Post-transcriptional control of the key enzyme for cell envelope synthesis by a base-pairing small RNA, an RNase adaptor protein, and a small RNA mimic. *RNA Biol.* **11**, 433–442 (2014).
398. Mandin, P. & Guillier, M. Expanding control in bacteria: interplay between small RNAs and transcriptional regulators to control gene expression. *Curr. Opin. Microbiol.* **16**, 125–132 (2013).
399. Storz, G., Vogel, J. & Wassarman, K. M. Regulation by Small RNAs in Bacteria: Expanding Frontiers. *Mol. Cell* **43**, 880–891 (2011).
400. Battesti, A., Majdalani, N. & Gottesman, S. The RpoS-Mediated General Stress Response in *Escherichia coli*. *Annu. Rev. Microbiol.* **65**, 189–213 (2011).
401. Holmqvist, E. *et al.* Two antisense RNAs target the transcriptional regulator CsgD to inhibit curli synthesis. *EMBO J.* **29**, 1840–1850 (2010).
402. Sharma, C. M. *et al.* Pervasive post-transcriptional control of genes involved in amino acid metabolism by the Hfq-dependent GcvB small RNA: GcvB regulon. *Mol. Microbiol.* **81**, 1144–1165 (2011).
403. Rutherford, S. T., van Kessel, J. C., Shao, Y. & Bassler, B. L. AphA and LuxR/HapR reciprocally control quorum sensing in vibrios. *Genes Dev.* **25**, 397–408 (2011).
404. Altuvia, S., Storz, G. & Papenfort, K. Cross-Regulation between Bacteria and Phages at a Posttranscriptional Level. *Microbiol. Spectr.* **6**, (2018).
405. Vial, L. & Hommais, F. Plasmid-chromosome cross-talks. *Environ. Microbiol.* **22**, 540–556 (2020).
406. Bloch, S., Lewandowska, N., Węgrzyn, G. & Nejman-Faleńczyk, B. Bacteriophages as sources of small non-coding RNA molecules. *Plasmid* **113**, 102527 (2021).
407. Fröhlich, K. S. & Papenfort, K. Interplay of regulatory RNAs and mobile genetic elements in enteric pathogens: Regulatory RNAs and mobile genetic elements. *Mol. Microbiol.* **101**, 701–

713 (2016).

408. Hershko-Shalev, T. *et al.* Gifsy-1 Prophage IsrK with Dual Function as Small and Messenger RNA Modulates Vital Bacterial Machineries. *PLoS Genet.* **12**, e1005975 (2016).
409. Westermann, A. J. *et al.* Dual RNA-seq unveils noncoding RNA functions in host–pathogen interactions. *Nature* **529**, 496–501 (2016).
410. Correia Santos, S., Bischler, T., Westermann, A. J. & Vogel, J. MAPS integrates regulation of actin-targeting effector SteC into the virulence control network of Salmonella small RNA PinT. *Cell Rep.* **34**, 108722 (2021).
411. Sy, B. M., Lan, R. & Tree, J. J. Early termination of the Shiga toxin transcript generates a regulatory small RNA. *Proc. Natl. Acad. Sci.* **117**, 25055–25065 (2020).
412. Tree, J. J., Granneman, S., McAteer, S. P., Tollervey, D. & Gally, D. L. Identification of Bacteriophage-Encoded Anti-sRNAs in Pathogenic Escherichia coli. *Mol. Cell* **55**, 199–213 (2014).
413. Dedrick, R. M., Mavrich, T. N., Ng, W. L. & Hatfull, G. F. Expression and evolutionary patterns of mycobacteriophage D29 and its temperate close relatives. *BMC Microbiol.* **17**, 225 (2017).
414. Chevallereau, A. *et al.* Next-Generation “-omics” Approaches Reveal a Massive Alteration of Host RNA Metabolism during Bacteriophage Infection of Pseudomonas aeruginosa. *PLoS Genet.* **12**, e1006134 (2016).
415. Harms, A., Brodersen, D. E., Mitarai, N. & Gerdes, K. Toxins, Targets, and Triggers: An Overview of Toxin-Antitoxin Biology. *Mol. Cell* **70**, 768–784 (2018).
416. Hayes, F. Toxins-Antitoxins: Plasmid Maintenance, Programmed Cell Death, and Cell Cycle Arrest. **301**, 5 (2003).
417. Fraikin, N., Goormaghtigh, F. & Van Melderen, L. Type II Toxin-Antitoxin Systems: Evolution and Revolutions. *J. Bacteriol.* **202**, (2020).
418. Sarpong, D. D. & Murphy, E. R. RNA Regulated Toxin-Antitoxin Systems in Pathogenic Bacteria. *Front. Cell. Infect. Microbiol.* **11**, 661026 (2021).
419. Li, M. *et al.* Toxin-antitoxin RNA pairs safeguard CRISPR-Cas systems. *Science* **372**, eabe5601 (2021).
420. Jørgensen, M. G., Pettersen, J. S. & Kallipolitis, B. H. sRNA-mediated control in bacteria: An increasing diversity of regulatory mechanisms. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* **1863**, 194504 (2020).
421. Carvalho Barbosa, C., Calhoun, S. H. & Wieden, H.-J. Non-coding RNAs: what are we missing? *Biochem. Cell Biol.* **98**, 23–30 (2020).
422. Fröhlich, K. S. & Vogel, J. Activation of gene expression by small RNA. *Curr. Opin. Microbiol.* **12**, 674–682 (2009).
423. Brantl, S. Regulatory mechanisms employed by cis-encoded antisense RNAs. *Curr. Opin. Microbiol.* **10**, 102–109 (2007).
424. Wagner, E. G. H., Altuvia, S. & Romby, P. 12 - Antisense RNAs in Bacteria and Their Genetic Elements. in *Advances in Genetics* (eds. Dunlap, J. C. & Wu, C. -ting) vol. 46 361–398 (Academic Press, 2002).
425. Thomason, M. K. & Storz, G. Bacterial Antisense RNAs: How Many Are There, and What Are They Doing? *Annu. Rev. Genet.* **44**, 167–188 (2010).

426. Gerdes, K. & Wagner, E. G. H. RNA antitoxins. *Curr. Opin. Microbiol.* **10**, 117–124 (2007).
427. Fozo, E. M., Hemm, M. R. & Storz, G. Small Toxic Proteins and the Antisense RNAs That Repress Them. *Microbiol. Mol. Biol. Rev.* **72**, 579–589 (2008).
428. Fozo, E. M. *et al.* Abundance of type I toxin–antitoxin systems in bacteria: searches for new candidates and discovery of novel families. *Nucleic Acids Res.* **38**, 3743–3759 (2010).
429. Pinel-Marie, M.-L., Brielle, R. & Felden, B. Dual Toxic-Peptide-Coding *Staphylococcus aureus* RNA under Antisense Regulation Targets Host Cells and Bacterial Rivals Unequally. *Cell Rep.* **7**, 424–435 (2014).
430. Opdyke, J. A., Kang, J.-G. & Storz, G. GadY, a Small-RNA Regulator of Acid Response Genes in *Escherichia coli*. *J. Bacteriol.* **186**, 6698–6705 (2004).
431. Lejars, M., Kobayashi, A. & Hajsndorf, E. Physiological roles of antisense RNAs in prokaryotes. *Biochimie* **164**, 3–16 (2019).
432. Lejars, M. & Hajsndorf, E. The world of asRNAs in Gram-negative and Gram-positive bacteria. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* **1863**, 194489 (2020).
433. Shearwin, K., Callen, B. & Egan, J. Transcriptional interference – a crash course. *Trends Genet.* **21**, 339–345 (2005).
434. André, G. *et al.* S-box and T-box riboswitches and antisense RNA control a sulfur metabolic operon of *Clostridium acetobutylicum*. *Nucleic Acids Res.* **36**, 5955–5969 (2008).
435. Stork, M., Di Lorenzo, M., Welch, T. J. & Crosa, J. H. Transcription Termination within the Iron Transport-Biosynthesis Operon of *Vibrio anguillarum* Requires an Antisense RNA. *J. Bacteriol.* **189**, 3479–3488 (2007).
436. Trinquier, A., Durand, S., Braun, F. & Condon, C. Regulation of RNA processing and degradation in bacteria. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* **1863**, 194505 (2020).
437. Carpousis, A. J., Luisi, B. F. & McDowall, K. J. Chapter 3 Endonucleolytic Initiation of mRNA Decay in *Escherichia coli*. in *Progress in Molecular Biology and Translational Science* vol. 85 91–135 (Academic Press, 2009).
438. Durand, S., Tomasini, A., Braun, F., Condon, C. & Romby, P. sRNA and mRNA turnover in Gram-positive bacteria. *FEMS Microbiol. Rev.* **39**, 316–330 (2015).
439. Vogel, J. RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res.* **31**, 6435–6443 (2003).
440. Kawano, M., Aravind, L. & Storz, G. An antisense RNA controls synthesis of an SOS-induced toxin evolved from an antitoxin. *Mol. Microbiol.* **64**, 738–754 (2007).
441. Ruiz de los Mozos, I. *et al.* Base Pairing Interaction between 5'- and 3'-UTRs Controls *icaR* mRNA Translation in *Staphylococcus aureus*. *PLoS Genet.* **9**, e1004001 (2013).
442. Mandin, P., Repoila, F., Vergassola, M., Geissmann, T. & Cossart, P. Identification of new noncoding RNAs in *Listeria monocytogenes* and prediction of mRNA targets. *Nucleic Acids Res.* **35**, 962–974 (2007).
443. Arnvig, K. B. & Young, D. B. Identification of small RNAs in *Mycobacterium tuberculosis*. *Mol. Microbiol.* **73**, 397–408 (2009).
444. Franch, T. & Gerdes, K. U-turns and regulatory RNAs. *Curr. Opin. Microbiol.* **3**, 159–164 (2000).
445. Miyakoshi, M., Chao, Y. & Vogel, J. Regulatory small RNAs from the 3' regions of bacterial

- mRNAs. *Curr. Opin. Microbiol.* **24**, 132–139 (2015).
446. Večerek, B., Moll, I. & Bläsi, U. Control of Fur synthesis by the non-coding RNA RyhB and iron-responsive decoding. *EMBO J.* **26**, 965–975 (2007).
447. Johansen, J., Eriksen, M., Kallipolitis, B. & Valentin-Hansen, P. Down-regulation of Outer Membrane Proteins by Noncoding RNAs: Unraveling the cAMP–CRP- and σ E-Dependent CyaR–ompX Regulatory Case. *J. Mol. Biol.* **383**, 1–9 (2008).
448. Urbanowski, M. L., Stauffer, L. T. & Stauffer, G. V. The *gcvB* gene encodes a small untranslated RNA involved in expression of the dipeptide and oligopeptide transport systems in *Escherichia coli*. *Mol. Microbiol.* **37**, 856–868 (2000).
449. De Lay, N. & Gottesman, S. The Crp-Activated Small Noncoding Regulatory RNA CyaR (RyeE) Links Nutritional Status to Group Behavior. *J. Bacteriol.* **191**, 461–476 (2009).
450. Fröhlich, K. S., Papenfort, K., Fekete, A. & Vogel, J. A small RNA activates CFA synthase by isoform-specific mRNA stabilization: Small RNA activates CFA synthesis. *EMBO J.* **32**, 2963–2979 (2013).
451. Morita, T., Mochizuki, Y. & Aiba, H. Translational repression is sufficient for gene silencing by bacterial small noncoding RNAs in the absence of mRNA destruction. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 4858 (2006).
452. Chevalier, C. *et al.* Staphylococcus aureus RNAIII Binds to Two Distant Regions of *coa* mRNA to Arrest Translation and Promote mRNA Degradation. *PLoS Pathog.* **6**, e1000809 (2010).
453. Kawamoto, H., Koide, Y., Morita, T. & Aiba, H. Base-pairing requirement for RNA silencing by a bacterial small RNA and acceleration of duplex formation by Hfq. *Mol. Microbiol.* **61**, 1013–1022 (2006).
454. Bejerano-Sagie, M. & Xavier, K. B. The role of small RNAs in quorum sensing. *Curr. Opin. Microbiol.* **10**, 189–198 (2007).
455. Massé, E., Salvail, H., Desnoyers, G. & Arguin, M. Small RNAs controlling iron metabolism. *Curr. Opin. Microbiol.* **10**, 140–145 (2007).
456. Valentin-Hansen, P., Johansen, J. & Rasmussen, A. A. Small RNAs controlling outer membrane porins. *Curr. Opin. Microbiol.* **10**, 152–155 (2007).
457. Aiba, H. Mechanism of RNA silencing by Hfq-binding small RNAs. *Curr. Opin. Microbiol.* **10**, 134–139 (2007).
458. Brennan, R. G. & Link, T. M. Hfq structure, function and ligand binding. *Curr. Opin. Microbiol.* **10**, 125–133 (2007).
459. Valentin-Hansen, P., Eriksen, M. & Udesen, C. The bacterial Sm-like protein Hfq: a key player in RNA transactions: Hfq, an Sm-related protein. *Mol. Microbiol.* **51**, 1525–1533 (2004).
460. Vogel, J. & Luisi, B. F. Hfq and its constellation of RNA. *Nat. Rev. Microbiol.* **9**, 578–589 (2011).
461. Hermann, H. *et al.* snRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein–protein interactions. *EMBO J.* **14**, 2076–2088 (1995).
462. Schu, D. J., Zhang, A., Gottesman, S. & Storz, G. Alternative Hfq- sRNA interaction modes dictate alternative mRNA recognition. *EMBO J.* **34**, 2557–2573 (2015).
463. Gorski, S. A., Vogel, J. & Doudna, J. A. RNA-based recognition and targeting: sowing the seeds of specificity. *Nat. Rev. Mol. Cell Biol.* **18**, 215–228 (2017).

464. Updegrave, T. B., Zhang, A. & Storz, G. Hfq: the flexible RNA matchmaker. *Curr. Opin. Microbiol.* **30**, 133–138 (2016).
465. Otaka, H., Ishikawa, H., Morita, T. & Aiba, H. PolyU tail of rho-independent terminator of bacterial small RNAs is essential for Hfq action. *Proc. Natl. Acad. Sci.* **108**, 13059–13064 (2011).
466. Salvail, H., Caron, M.-P., Bélanger, J. & Massé, E. Antagonistic functions between the RNA chaperone Hfq and an sRNA regulate sensitivity to the antibiotic colicin. *EMBO J.* **32**, 2764–2778 (2013).
467. Chao, Y. & Vogel, J. The role of Hfq in bacterial pathogens. *Curr. Opin. Microbiol.* **13**, 24–33 (2010).
468. Oliva, G., Sahr, T. & Buchrieser, C. Small RNAs, 5' UTR elements and RNA-binding proteins in intracellular bacteria: impact on metabolism and virulence. *FEMS Microbiol. Rev.* **39**, 331–349 (2015).
469. Bohn, C., Rigoulay, C. & Bouloc, P. No detectable effect of RNA-binding protein Hfq absence in *Staphylococcus aureus*. *BMC Microbiol.* **7**, 10 (2007).
470. Christiansen, J. K. Identification of small Hfq-binding RNAs in *Listeria monocytogenes*. *RNA* **12**, 1383–1396 (2006).
471. Kim, S., Reyes, D., Beaume, M., Francois, P. & Cheung, A. Contribution of teg49 Small RNA in the 5' Upstream Transcriptional Region of *sarA* to Virulence in *Staphylococcus aureus*. *Infect. Immun.* **82**, 4369–4379 (2014).
472. Manna, A. C. *et al.* Small RNA teg49 Is Derived from a *sarA* Transcript and Regulates Virulence Genes Independent of SarA in *Staphylococcus aureus*. *Infect. Immun.* **86**, (2018).
473. Kortmann, J. & Narberhaus, F. Bacterial RNA thermometers: molecular zippers and switches. *Nat. Rev. Microbiol.* **10**, 255–265 (2012).
474. Johansson, J. *et al.* An RNA Thermosensor Controls Expression of Virulence Genes in *Listeria monocytogenes*. *Cell* **110**, 551–561 (2002).
475. Loh, E. *et al.* An unstructured 5'-coding region of the *prfA* mRNA is required for efficient translation. *Nucleic Acids Res.* **40**, 1818–1827 (2012).
476. Grundy, F. J. & Henkin, T. M. From Ribosome to Riboswitch: Control of Gene Expression in Bacteria by RNA Structural Rearrangements. *Crit. Rev. Biochem. Mol. Biol.* **41**, 329–338 (2006).
477. Breaker, R. R. Riboswitches and the RNA World. *Cold Spring Harb. Perspect. Biol.* **4**, a003566–a003566 (2012).
478. Mandal, M. & Breaker, R. R. Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.* **5**, 451–463 (2004).
479. Nudler, E. The riboswitch control of bacterial metabolism. *Trends Biochem. Sci.* **29**, 11–17 (2004).
480. Montange, R. K. & Batey, R. T. Riboswitches: Emerging Themes in RNA Structure and Function. *Annu. Rev. Biophys.* **37**, 117–133 (2008).
481. Henkin, T. M. Riboswitch RNAs: using RNA to sense cellular metabolism. *Genes Dev.* **22**, 3383–3390 (2008).
482. Roth, A. & Breaker, R. R. The Structural and Functional Diversity of Metabolite-Binding Riboswitches. *Annu. Rev. Biochem.* **78**, 305–334 (2009).
483. Tamayo, R. Cyclic diguanylate riboswitches control bacterial pathogenesis mechanisms.

PLOS Pathog. **15**, e1007529 (2019).

484. Toledo-Arana, A. *et al.* The *Listeria* transcriptional landscape from saprophytism to virulence. *Nature* **459**, 950–956 (2009).
485. Babitzke, P. & Romeo, T. CsrB sRNA family: sequestration of RNA-binding regulatory proteins. *Curr. Opin. Microbiol.* **10**, 156–163 (2007).
486. Suzuki, K. Identification of a novel regulatory protein (CsrD) that targets the global regulatory RNAs CsrB and CsrC for degradation by RNase E. *Genes Dev.* **20**, 2605–2617 (2006).
487. Lapouge, K., Schubert, M., Allain, F. H.-T. & Haas, D. Gac/Rsm signal transduction pathway of γ -proteobacteria: from RNA recognition to regulation of social behaviour: Regulation of RsmA/CsrA binding to RNA. *Mol. Microbiol.* **67**, 241–253 (2007).
488. Lucchetti-Miganeh, C., Burrowes, E., Baysse, C. & Ermel, G. The post-transcriptional regulator CsrA plays a central role in the adaptation of bacterial pathogens to different stages of infection in animal hosts. *Microbiology* **154**, 16–29 (2008).
489. Wassarman, K. M. 6S RNA: a small RNA regulator of transcription. *Curr. Opin. Microbiol.* **10**, 164–168 (2007).
490. Trotochaud, A. E. & Wassarman, K. M. A highly conserved 6S RNA structure is required for regulation of transcription. *Nat. Struct. Mol. Biol.* **12**, 313–319 (2005).
491. Reichenbach, B., Maes, A., Kalamorz, F., Hajnsdorf, E. & Görke, B. The small RNA GlmY acts upstream of the sRNA GlmZ in the activation of glmS expression and is subject to regulation by polyadenylation in *Escherichia coli*. *Nucleic Acids Res.* **36**, 2570–2580 (2008).
492. Marraffini, L. A. CRISPR-Cas immunity in prokaryotes. *Nature* **526**, 55–61 (2015).
493. Sorek, R., Kunin, V. & Hugenholtz, P. CRISPR — a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat. Rev. Microbiol.* **6**, 181–186 (2008).
494. Westra, E. R., Buckling, A. & Fineran, P. C. CRISPR–Cas systems: beyond adaptive immunity. *Nat. Rev. Microbiol.* **12**, 317–326 (2014).
495. Sampson, T. R. & Weiss, D. S. CRISPR-Cas systems: new players in gene regulation and bacterial physiology. *Front. Cell. Infect. Microbiol.* **4**, (2014).
496. Louwen, R., Staals, R. H. J., Endtz, H. P., van Baarlen, P. & van der Oost, J. The Role of CRISPR-Cas Systems in Virulence of Pathogenic Bacteria. *Microbiol. Mol. Biol. Rev.* **78**, 74–88 (2014).
497. Koonin, E. V. & Makarova, K. S. Origins and evolution of CRISPR-Cas systems. *Philos. Trans. R. Soc. B Biol. Sci.* **374**, 20180087 (2019).
498. Barrangou, R. CRISPR-Cas systems and RNA-guided interference: CRISPR-Cas systems and RNA-guided interference. *Wiley Interdiscip. Rev. RNA* **4**, 267–278 (2013).
499. Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71 (2010).
500. Wiedenheft, B., Sternberg, S. H. & Doudna, J. A. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**, 331–338 (2012).
501. Westra, E. R. *et al.* The CRISPRs, They Are A-Changin': How Prokaryotes Generate Adaptive Immunity. *Annu. Rev. Genet.* **46**, 311–339 (2012).
502. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**,

733–740 (2009).

503. Wang, R., Preamplume, G., Terns, M. P., Terns, R. M. & Li, H. Interaction of the Cas6 Riboendonuclease with CRISPR RNAs: Recognition and Cleavage. *Structure* **19**, 257–264 (2011).
504. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607 (2011).
505. Brouns, S. J. J. *et al.* Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science* **321**, 960–964 (2008).
506. Rouillon, C. *et al.* Structure of the CRISPR Interference Complex CSM Reveals Key Similarities with Cascade. *Mol. Cell* **52**, 124–134 (2013).
507. Spilman, M. *et al.* Structure of an RNA Silencing Complex of the CRISPR-Cas Immune System. *Mol. Cell* **52**, 146–152 (2013).
508. Shmakov, S. *et al.* Diversity and evolution of class 2 CRISPR–Cas systems. *Nat. Rev. Microbiol.* **15**, 169–182 (2017).
509. Wang, F. *et al.* Advances in CRISPR-Cas systems for RNA targeting, tracking and editing. *Biotechnol. Adv.* **37**, 708–729 (2019).
510. Liu, Z., Dong, H., Cui, Y., Cong, L. & Zhang, D. Application of different types of CRISPR/Cas-based systems in bacteria. *Microb. Cell Factories* **19**, 172 (2020).
511. Fang, F. C., Frawley, E. R., Tapscott, T. & Vázquez-Torres, A. Bacterial Stress Responses during Host Infection. *Cell Host Microbe* **20**, 133–143 (2016).
512. Fuchs, T. M., Eisenreich, W., Heesemann, J. & Goebel, W. Metabolic adaptation of human pathogenic and related nonpathogenic bacteria to extra- and intracellular habitats. *FEMS Microbiol. Rev.* **36**, 435–462 (2012).
513. Stewart, P. S. Diffusion in Biofilms. *J. Bacteriol.* **185**, 1485–1491 (2003).
514. Vestby, L. K., Grønseth, T., Simm, R. & Nesse, L. L. Bacterial Biofilm and its Role in the Pathogenesis of Disease. *Antibiotics* **9**, 59 (2020).
515. Antunes, L. C. M., Ferreira, R. B. R., Buckner, M. M. C. & Finlay, B. B. Quorum sensing in bacterial virulence. *Microbiology* **156**, 2271–2282 (2010).
516. Papenfort, K. & Vogel, J. Regulatory RNA in Bacterial Pathogens. *Cell Host Microbe* **8**, 116–127 (2010).
517. Chakravarty, S. & Massé, E. RNA-Dependent Regulation of Virulence in Pathogenic Bacteria. *Front. Cell. Infect. Microbiol.* **9**, 337 (2019).
518. Quereda, J. J. & Cossart, P. Regulating Bacterial Virulence with RNA. *Annu. Rev. Microbiol.* **71**, 263–280 (2017).
519. Caldelari, I., Chao, Y., Romby, P. & Vogel, J. RNA-Mediated Regulation in Pathogenic Bacteria. *Cold Spring Harb. Perspect. Med.* **3**, a010298–a010298 (2013).
520. Westermann, A. J. Regulatory RNAs in Virulence and Host-Microbe Interactions. *Microbiol. Spectr.* **6**, (2018).
521. Terlizzi, M. E., Gribaudo, G. & Maffei, M. E. UroPathogenic *Escherichia coli* (UPEC) Infections: Virulence Factors, Bladder Responses, Antibiotic, and Non-antibiotic Antimicrobial Strategies. *Front. Microbiol.* **8**, 1566 (2017).
522. Kulesus, R. R., Diaz-Perez, K., Slechta, E. S., Eto, D. S. & Mulvey, M. A. Impact of the RNA Chaperone Hfq on the Fitness and Virulence Potential of Uropathogenic *Escherichia coli*. *Infect.*

Immun. **76**, 3019–3026 (2008).

523. Khandige, S., Kronborg, T., Uhlin, B. E. & Møller-Jensen, J. sRNA-Mediated Regulation of P-Fimbriae Phase Variation in Uropathogenic *Escherichia coli*. *PLoS Pathog.* **11**, e1005109 (2015).
524. Porcheron, G. *et al.* The Small RNA RyhB Contributes to Siderophore Production and Virulence of Uropathogenic *Escherichia coli*. *Infect. Immun.* **82**, 5056–5068 (2014).
525. Murphy, E. R. & Payne, S. M. RyhB, an Iron-Responsive Small RNA Molecule, Regulates *Shigella dysenteriae* Virulence. *Infect. Immun.* **75**, 3470–3477 (2007).
526. Oglesby-Sherrouse, A. G. & Murphy, E. R. Iron-responsive bacterial small RNAs: variations on a theme. *Metallomics* **5**, 276 (2013).
527. Wroblewski, L. E., Peek, R. M. & Wilson, K. T. *Helicobacter pylori* and Gastric Cancer: Factors That Modulate Disease Risk. *Clin. Microbiol. Rev.* **23**, 713–739 (2010).
528. Libânio, D. *Helicobacter pylori* and microRNAs: Relation with innate immunity and progression of preneoplastic conditions. *World J. Clin. Oncol.* **6**, 111 (2015).
529. Pernitzsch, S. R. & Sharma, C. M. Transcriptome Complexity and Riboregulation in the Human Pathogen *Helicobacter pylori*. *Front. Cell. Infect. Microbiol.* **2**, (2012).
530. Hanyu, H. *et al.* *Helicobacter pylori* Uses the TlpB Receptor To Sense Sites of Gastric Injury. *Infect. Immun.* **87**, (2019).
531. Croxen, M. A., Sisson, G., Melano, R. & Hoffman, P. S. The *Helicobacter pylori* Chemotaxis Receptor TlpB (HP0103) Is Required for pH Taxis and for Colonization of the Gastric Mucosa. *J. Bacteriol.* **188**, 2656–2665 (2006).
532. Graham, D. Y. & Miftahussurur, M. *Helicobacter pylori* urease for diagnosis of *Helicobacter pylori* infection: A mini review. *J. Adv. Res.* **13**, 51–57 (2018).
533. Pflock, M., Kennard, S., Delany, I., Scarlato, V. & Beier, D. Acid-Induced Activation of the Urease Promoters Is Mediated Directly by the ArsRS Two-Component System of *Helicobacter pylori*. *Infect. Immun.* **73**, 6437–6445 (2005).
534. Wen, Y., Feng, J. & Sachs, G. *Helicobacter pylori* 5'ureB-sRNA, a cis-Encoded Antisense Small RNA, Negatively Regulates ureAB Expression by Transcription Termination. *J. Bacteriol.* **195**, 444–452 (2013).
535. Smith, W. D. *et al.* Current and future therapies for *Pseudomonas aeruginosa* infection in patients with cystic fibrosis. *FEMS Microbiol. Lett.* **364**, (2017).
536. Emerson, J., Rosenfeld, M., McNamara, S., Ramsey, B. & Gibson, R. L. *Pseudomonas aeruginosa* and other predictors of mortality and morbidity in young children with cystic fibrosis. *Pediatr. Pulmonol.* **34**, 91–100 (2002).
537. Bjarnsholt, T. *et al.* *Pseudomonas aeruginosa* biofilms in the respiratory tract of cystic fibrosis patients: *Pseudomonas aeruginosa* Biofilms in the Cystic Fibrosis Lung. *Pediatr. Pulmonol.* **44**, 547–558 (2009).
538. Matsui, H. *et al.* A physical linkage between cystic fibrosis airway surface dehydration and *Pseudomonas aeruginosa* biofilms. *Proc. Natl. Acad. Sci.* **103**, 18131–18136 (2006).
539. Conese, M., Copreni, E., Gioia, S. D., Rinaldis, P. D. & Fumarulo, R. Neutrophil recruitment and airway epithelial cell involvement in chronic cystic fibrosis lung disease. *J. Cyst. Fibros.* **2**, 129–135 (2003).
540. Tolker-Nielsen, T. & Høiby, N. Extracellular DNA and F-actin as targets in antibiofilm cystic

- fibrosis therapy. *Future Microbiol.* **4**, 645–647 (2009).
541. Fuxman Bass, J. I. *et al.* Extracellular DNA: A Major Proinflammatory Component of *Pseudomonas aeruginosa* Biofilms. *J. Immunol.* **184**, 6386–6395 (2010).
542. Pita, T., Feliciano, J. & Leitão, J. Small Noncoding Regulatory RNAs from *Pseudomonas aeruginosa* and *Burkholderia cepacia* Complex. *Int. J. Mol. Sci.* **19**, 3759 (2018).
543. Vakulskas, C. A., Potts, A. H., Babitzke, P., Ahmer, B. M. M. & Romeo, T. Regulation of Bacterial Virulence by Csr (Rsm) Systems. *Microbiol. Mol. Biol. Rev.* **79**, 193–224 (2015).
544. Romero, M. *et al.* Genome-wide mapping of the RNA targets of the *Pseudomonas aeruginosa* riboregulatory protein RsmN. *Nucleic Acids Res.* **46**, 6823–6840 (2018).
545. Janssen, K. H., Diaz, M. R., Gode, C. J., Wolfgang, M. C. & Yahr, T. L. RsmV, a Small Noncoding Regulatory RNA in *Pseudomonas aeruginosa* That Sequesters RsmA and RsmF from Target mRNAs. *J. Bacteriol.* **200**, (2018).
546. Brencic, A. *et al.* The GacS/GacA signal transduction system of *Pseudomonas aeruginosa* acts exclusively through its control over the transcription of the RsmY and RsmZ regulatory small RNAs. *Mol. Microbiol.* **73**, 434–445 (2009).
547. Chakravarty, S., Melton, C. N., Bailin, A., Yahr, T. L. & Anderson, G. G. *Pseudomonas aeruginosa* Magnesium Transporter MgtE Inhibits Type III Secretion System Gene Expression by Stimulating *rsmYZ* Transcription. *J. Bacteriol.* **199**, (2017).
548. Broder, U. N., Jaeger, T. & Jenal, U. LadS is a calcium-responsive kinase that induces acute-to-chronic virulence switch in *Pseudomonas aeruginosa*. *Nat. Microbiol.* **2**, 16184 (2017).
549. Chen, R. *et al.* Polynucleotide Phosphorylase Regulates Multiple Virulence Factors and the Stabilities of Small RNAs RsmY/Z in *Pseudomonas aeruginosa*. *Front. Microbiol.* **7**, (2016).
550. Williams McMackin, E. A., Djapgne, L., Corley, J. M. & Yahr, T. L. Fitting Pieces into the Puzzle of *Pseudomonas aeruginosa* Type III Secretion System Gene Expression. *J. Bacteriol.* **201**, (2019).
551. Carloni, S., Macchi, R., Sattin, S., Ferrara, S. & Berton, G. The small RNA Real: a novel regulatory element embedded in the *Pseudomonas aeruginosa* quorum sensing networks: Real connects *las* to *pqs* quorum sensing systems. *Environ. Microbiol.* **19**, 4220–4237 (2017).
552. Meyer, J. M., Neely, A., Stintzi, A., Georges, C. & Holder, I. A. Pyoverdine is essential for virulence of *Pseudomonas aeruginosa*. *INFECT IMMUN* **64**, 6 (1996).
553. Lau Gee W., Ran Huimin, Kong Fansheng, Hasset Daniel J., & Mavrodi Dimitri. *Pseudomonas aeruginosa* Pyocyanin Is Critical for Lung Infection in Mice. *Infect. Immun.* **72**, 4275–4278 (2004).
554. Sonnleitner, E. *et al.* The small RNA PhrS stimulates synthesis of the *Pseudomonas aeruginosa* quinolone signal: The small RNA PhrS activates translation of *pqsR*. *Mol. Microbiol.* **80**, 868–885 (2011).
555. Zimmermann, A., Reimann, C., Galimand, M. & Haas, D. Anaerobic growth and cyanide synthesis of *Pseudomonas aeruginosa* depend on *anr*, a regulatory gene homologous with *fnr* of *Escherichia coli*. *Mol. Microbiol.* **5**, 1483–1490 (1991).
556. Brouwer, S. *et al.* The PqsR and RhlR Transcriptional Regulators Determine the Level of *Pseudomonas* Quinolone Signal Synthesis in *Pseudomonas aeruginosa* by Producing Two Different *pqsABCDE* mRNA Isoforms. *J. Bacteriol.* **196**, 4163–4171 (2014).

557. Reinhart, A. A. *et al.* The *Pseudomonas aeruginosa* PrrF Small RNAs Regulate Iron Homeostasis during Acute Murine Lung Infection. *Infect. Immun.* **85**, (2017).
558. Djapgne, L. *et al.* The *Pseudomonas aeruginosa* PrrF1 and PrrF2 Small Regulatory RNAs Promote 2-Alkyl-4-Quinolone Production through Redundant Regulation of the *antR* mRNA. *J. Bacteriol.* **200**, (2018).
559. Nelson, C. E. *et al.* Proteomic Analysis of the *Pseudomonas aeruginosa* Iron Starvation Response Reveals PrrF Small Regulatory RNA-Dependent Iron Regulation of Twitching Motility, Amino Acid Metabolism, and Zinc Homeostasis Proteins. *J. Bacteriol.* **201**, (2019).
560. Reinhart, A. A. *et al.* The *prfF*-Encoded Small Regulatory RNAs Are Required for Iron Homeostasis and Virulence of *Pseudomonas aeruginosa*. *Infect. Immun.* **83**, 863–875 (2015).
561. Oglesby, A. G. *et al.* The Influence of Iron on *Pseudomonas aeruginosa* Physiology. *J. Biol. Chem.* **283**, 15558–15567 (2008).
562. Falcone, M. *et al.* The Small RNA *ErsA* of *Pseudomonas aeruginosa* Contributes to Biofilm Development and Motility through Post-transcriptional Modulation of *AmrZ*. *Front. Microbiol.* **9**, 238 (2018).
563. Okkotsu, Y., Little, A. S. & Schurr, M. J. The *Pseudomonas aeruginosa* AlgZR two-component system coordinates multiple phenotypes. *Front. Cell. Infect. Microbiol.* **4**, (2014).
564. Li, H., Luo, Y.-F., Williams, B. J., Blackwell, T. S. & Xie, C.-M. Structure and function of *OprD* protein in *Pseudomonas aeruginosa*: From antibiotic resistance to novel therapies. *Int. J. Med. Microbiol.* **302**, 63–68 (2012).
565. Zhang, Y.-F. *et al.* Probing the sRNA regulatory landscape of *P. aeruginosa*: post-transcriptional control of determinants of pathogenicity and antibiotic susceptibility: sRNA regulatory landscape of *P. aeruginosa*. *Mol. Microbiol.* **106**, 919–937 (2017).
566. Tong, S. Y. C., Davis, J. S., Eichenberger, E., Holland, T. L. & Fowler, V. G. *Staphylococcus aureus* Infections: Epidemiology, Pathophysiology, Clinical Manifestations, and Management. *Clin. Microbiol. Rev.* **28**, 603–661 (2015).
567. Lister, J. L. & Horswill, A. R. *Staphylococcus aureus* biofilms: recent developments in biofilm dispersal. *Front. Cell. Infect. Microbiol.* **4**, (2014).
568. Wang, L. & Ruan, S. Modeling Nosocomial Infections of Methicillin-Resistant *Staphylococcus aureus* with Environment Contamination*. *Sci. Rep.* **7**, 580 (2017).
569. Santajit, S. & Indrawattana, N. Mechanisms of Antimicrobial Resistance in ESKAPE Pathogens. *BioMed Res. Int.* **2016**, 2475067 (2016).
570. Powers, M. E. & Wardenburg, J. B. Igniting the Fire: *Staphylococcus aureus* Virulence Factors in the Pathogenesis of Sepsis. *PLoS Pathog.* **10**, e1003871 (2014).
571. Fechter, P., Caldelari, I., Lioliou, E. & Romby, P. Novel aspects of RNA regulation in *Staphylococcus aureus*. *FEBS Lett.* **588**, 2523–2529 (2014).
572. Tomasini, A. *et al.* The importance of regulatory RNAs in *Staphylococcus aureus*. *Infect. Genet. Evol.* **21**, 616–626 (2014).
573. Boisset, S. *et al.* *Staphylococcus aureus* RNAIII coordinately represses the synthesis of virulence factors and the transcription regulator *Rot* by an antisense mechanism. *Genes Dev.* **21**, 1353–1366 (2007).
574. Felden, B., Vandenesch, F., Bouloc, P. & Romby, P. The *Staphylococcus aureus* RNome and

- Its Commitment to Virulence. *PLoS Pathog.* **7**, e1002006 (2011).
575. Guillet, J., Hallier, M. & Felden, B. Emerging Functions for the Staphylococcus aureus RNome. *PLoS Pathog.* **9**, e1003767 (2013).
576. Geisinger, E., Adhikari, R. P., Jin, R., Ross, H. F. & Novick, R. P. Inhibition of rot translation by RNAIII, a key feature of agr function. *Mol. Microbiol.* **61**, 1038–1048 (2006).
577. Toledo-Arana, A., Repoila, F. & Cossart, P. Small noncoding RNAs controlling pathogenesis. *Curr. Opin. Microbiol.* **10**, 182–188 (2007).
578. Geissmann, T. *et al.* A search for small noncoding RNAs in Staphylococcus aureus reveals a conserved sequence motif for regulation. *Nucleic Acids Res.* **37**, 7239–7257 (2009).
579. Romilly, C. *et al.* A Non-Coding RNA Promotes Bacterial Persistence and Decreases Virulence by Regulating a Regulator in Staphylococcus aureus. *PLoS Pathog.* **10**, e1003979 (2014).
580. Tomasini, A. *et al.* The RNA targetome of Staphylococcus aureus non-coding RNA RsaA: impact on cell surface properties and defense mechanisms. *Nucleic Acids Res.* **45**, 6746–6760 (2017).
581. Luong, T. T., Dunman, P. M., Murphy, E., Projan, S. J. & Lee, C. Y. Transcription Profiling of the *mgrA* Regulon in *Staphylococcus aureus*. *J. Bacteriol.* **188**, 1899–1910 (2006).
582. Trottonda, M. P., Tamber, S., Memmi, G. & Cheung, A. L. MgrA Represses Biofilm Formation in *Staphylococcus aureus*. *Infect. Immun.* **76**, 5645–5654 (2008).
583. Chabelskaya, S., Gaillot, O. & Felden, B. A Staphylococcus aureus Small RNA Is Required for Bacterial Virulence and Regulates the Expression of an Immune-Evasion Molecule. *PLoS Pathog.* **6**, e1000927 (2010).
584. Haupt, K. *et al.* The Staphylococcus aureus Protein Sbi Acts as a Complement Inhibitor and Forms a Tripartite Complex with Host Complement Factor H and C3b. *PLoS Pathog.* **4**, e1000250 (2008).
585. Radoshevich, L. & Cossart, P. *Listeria monocytogenes*: towards a complete picture of its physiology and pathogenesis. *Nat. Rev. Microbiol.* **16**, 32–46 (2018).
586. Lebreton, A. & Cossart, P. RNA- and protein-mediated control of *Listeria monocytogenes* virulence gene expression. *RNA Biol.* **14**, 460–470 (2017).
587. Quereda, J. J., Ortega, Á. D., Pucciarelli, M. G. & García-del Portillo, F. The *Listeria* Small RNA Rli27 Regulates a Cell Wall Protein inside Eukaryotic Cells by Targeting a Long 5'-UTR Variant. *PLoS Genet.* **10**, e1004765 (2014).
588. García-del Portillo, F., Calvo, E., D'Orazio, V. & Pucciarelli, M. G. Association of ActA to Peptidoglycan Revealed by Cell Wall Proteomics of Intracellular *Listeria monocytogenes*. *J. Biol. Chem.* **286**, 34675–34689 (2011).
589. Sievers, S. *et al.* A multicopy sRNA of *Listeria monocytogenes* regulates expression of the virulence adhesin LapB. *Nucleic Acids Res.* **42**, 9383–9398 (2014).
590. dos Santos, P. T. *et al.* The Small Regulatory RNAs LhrC1–5 Contribute to the Response of *Listeria monocytogenes* to Heme Toxicity. *Front. Microbiol.* **9**, 599 (2018).
591. Sievers, S. *et al.* The multicopy sRNA LhrC controls expression of the oligopeptide-binding protein OppA in *Listeria monocytogenes*. *RNA Biol.* **12**, 985–997 (2015).
592. Ross, J. A., Thorsing, M., Lillebæk, E. M. S., Teixeira Dos Santos, P. & Kallipolitis, B. H. The

- LhrC sRNAs control expression of T cell-stimulating antigen TcsA in *Listeria monocytogenes* by decreasing *tcsA* mRNA stability. *RNA Biol.* **16**, 270–281 (2019).
593. Garsin, D. A. Ethanolamine utilization in bacterial pathogens: roles and regulation. *Nat. Rev. Microbiol.* **8**, 290–295 (2010).
594. Mellin, J. R. *et al.* Sequestration of a two-component response regulator by a riboswitch-regulated noncoding RNA. *Science* **345**, 940–943 (2014).
595. Mellin, J. R. *et al.* A riboswitch-regulated antisense RNA in *Listeria monocytogenes*. *Proc. Natl. Acad. Sci.* **110**, 13132–13137 (2013).
596. Kim, E. Y., Jakobson, C. M. & Tullman-Ercek, D. Engineering Transcriptional Regulation to Control Pdu Microcompartment Formation. *PLoS ONE* **9**, e113814 (2014).
597. de las Heras, A., Cain, R. J., Bielecka, M. K. & Vázquez-Boland, J. A. Regulation of *Listeria* virulence: PrfA master and commander. *Curr. Opin. Microbiol.* **14**, 118–127 (2011).
598. Loh, E. *et al.* A trans-Acting Riboswitch Controls Expression of the Virulence Regulator PrfA in *Listeria monocytogenes*. *Cell* **139**, 770–779 (2009).
599. Sesto, N., Wurtzel, O., Archambaud, C., Sorek, R. & Cossart, P. The excludon: a new concept in bacterial antisense RNA-mediated gene regulation. *Nat. Rev. Microbiol.* **11**, 75–82 (2013).
600. Soutourina, O. RNA-based control mechanisms of *Clostridium difficile*. *Curr. Opin. Microbiol.* **36**, 62–68 (2017).
601. Soutourina, O. *et al.* Genome-Wide Transcription Start Site Mapping and Promoter Assignments to a Sigma Factor in the Human Enteropathogen *Clostridioides difficile*. *Front. Microbiol.* **11**, 1939 (2020).
602. Boudry, P. *et al.* Identification of RNAs bound by Hfq reveals widespread RNA partners and a sporulation regulator in the human pathogen *Clostridioides difficile*. *RNA Biol.* **18**, 1931–1952 (2021).
603. Chen, Y., Indurthi, D. C., Jones, S. W. & Papoutsakis, E. T. Small RNAs in the Genus *Clostridium*. *mBio* **2**, (2011).
604. Fuchs, M. *et al.* An RNA-centric global view of *Clostridioides difficile* reveals broad activity of Hfq in a clinically important gram-positive bacterium. *Proc. Natl. Acad. Sci.* **118**, e2103579118 (2021).
605. Bordeleau, E., Fortier, L.-C., Malouin, F. & Burrus, V. c-di-GMP Turn-Over in *Clostridium difficile* Is Controlled by a Plethora of Diguanylate Cyclases and Phosphodiesterases. *PLoS Genet.* **7**, e1002039 (2011).
606. McKee, R. W., Harvest, C. K. & Tamayo, R. Cyclic Diguanylate Regulates Virulence Factor Genes via Multiple Riboswitches in *Clostridium difficile*. *mSphere* **3**, (2018).
607. Lee, E. R., Baker, J. L., Weinberg, Z., Sudarsan, N. & Breaker, R. R. An Allosteric Self-Splicing Ribozyme Triggered by a Bacterial Second Messenger. *Science* **329**, 845–848 (2010).
608. Peltier, J. *et al.* Type I toxin-antitoxin systems contribute to the maintenance of mobile genetic elements in *Clostridioides difficile*. *Commun. Biol.* **3**, 718 (2020).
609. Maikova, A., Kreis, V., Boutserin, A., Severinov, K. & Soutourina, O. Using an Endogenous CRISPR-Cas System for Genome Editing in the Human Pathogen *Clostridium difficile*. *Appl. Environ. Microbiol.* **85**, (2019).
610. Boudry, P. *et al.* Pleiotropic Role of the RNA Chaperone Protein Hfq in the Human Pathogen

- Clostridium difficile*. *J. Bacteriol.* **196**, 3234–3248 (2014).
611. Maikova, A. *et al.* Discovery of new type I toxin–antitoxin systems adjacent to CRISPR arrays in *Clostridium difficile*. *Nucleic Acids Res.* **46**, 4733–4751 (2018).
612. Soutourina, O. Type I Toxin-Antitoxin Systems in Clostridia. *Toxins* **11**, 253 (2019).
613. Lamm-Schmidt, V. *et al.* Grad-seq identifies KhpB as a global RNA-binding protein in *Clostridioides difficile* that regulates toxin production. *microLife* **2**, uqab004 (2021).
614. Andersen, J. M. *et al.* CRISPR Diversity and Microevolution in *Clostridium difficile*. *Genome Biol. Evol.* **8**, 2841–2855 (2016).
615. Maikova, A. *et al.* Protospacer-Adjacent Motif Specificity during *Clostridioides difficile* Type I-B CRISPR-Cas Interference and Adaptation. *mBio* **12**, (2021).
616. Morris, K. V. & Mattick, J. S. The rise of regulatory RNA. *Nat. Rev. Genet.* **15**, 423–437 (2014).
617. Felden, B. & Paillard, L. When eukaryotes and prokaryotes look alike: the case of regulatory RNAs. *FEMS Microbiol. Rev.* **41**, 624–639 (2017).
618. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
619. Kim, V. N. & Nam, J.-W. Genomics of microRNA. *Trends Genet.* **22**, 165–173 (2006).
620. Kim, V. N. MicroRNA biogenesis: coordinated cropping and dicing. *Nat. Rev. Mol. Cell Biol.* **6**, 376–385 (2005).
621. Carthew, R. W. & Sontheimer, E. J. Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136**, 642–655 (2009).
622. Golden, D. E., Gerbasi, V. R. & Sontheimer, E. J. An Inside Job for siRNAs. *Mol. Cell* **31**, 309–312 (2008).
623. Meister, G. & Tuschl, T. Mechanisms of gene silencing by double-stranded RNA. *Nature* **431**, 343–349 (2004).
624. Mello, C. C. & Conte, D. Revealing the world of RNA interference. *Nature* **431**, 338–342 (2004).
625. Lippman, Z. & Martienssen, R. The role of RNA interference in heterochromatic silencing. *Nature* **431**, 364–370 (2004).
626. Slotkin, R. K. & Martienssen, R. Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* **8**, 272–285 (2007).
627. Allen, E., Xie, Z., Gustafson, A. M. & Carrington, J. C. microRNA-Directed Phasing during Trans-Acting siRNA Biogenesis in Plants. *Cell* **121**, 207–221 (2005).
628. Girard, A., Sachidanandam, R., Hannon, G. J. & Carmell, M. A. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442**, 199–202 (2006).
629. Aravin, A. A., Sachidanandam, R., Girard, A., Fejes-Toth, K. & Hannon, G. J. Developmentally Regulated piRNA Clusters Implicate MILI in Transposon Control. *Science* **316**, 744–747 (2007).
630. Chapman, E. J. & Carrington, J. C. Specialization and evolution of endogenous small RNA pathways. *Nat. Rev. Genet.* **8**, 884–896 (2007).
631. Mercer, T. R., Dinger, M. E. & Mattick, J. S. Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.* **10**, 155–159 (2009).
632. Feng, J. The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. *Genes Dev.* **20**, 1470–1484 (2006).

633. Martianov, I., Ramadass, A., Serra Barros, A., Chow, N. & Akoulitchev, A. Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature* **445**, 666–670 (2007).
634. Wang, X. *et al.* Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature* **454**, 126–130 (2008).
635. Li, Y., Lu, J., Han, Y., Fan, X. & Ding, S.-W. RNA Interference Functions as an Antiviral Immunity Mechanism in Mammals. *Science* **342**, 231–234 (2013).
636. Maillard, P. V. *et al.* Antiviral RNA Interference in Mammalian Cells. *Science* **342**, 235–238 (2013).
637. García, M. A., Meurs, E. F. & Esteban, M. The dsRNA protein kinase PKR: Virus and cell control. *Biochimie* **89**, 799–811 (2007).
638. Vogt, A., Goldman, A. D., Mochizuki, K. & Landweber, L. F. Transposon Domestication versus Mutualism in Ciliate Genome Rearrangements. *PLoS Genet.* **9**, e1003659 (2013).
639. Ross, J. A., Ellis, M. J., Hossain, S. & Haniford, D. B. Hfq restructures RNA-IN and RNA-OUT and facilitates antisense pairing in the Tn10/IS10 system. *RNA* **19**, 670–684 (2013).
640. Diribarne, G. & Bensaude, O. 7SK RNA, a non-coding RNA regulating P-TEFb, a general transcription factor. *RNA Biol.* **6**, 122–128 (2009).
641. Miska, E. A. *et al.* Most *Caenorhabditis elegans* microRNAs are individually not essential for development or viability. *PLoS Genet.* **3**, e215 (2007).
642. Chen, Y.-W. *et al.* Systematic Study of *Drosophila* MicroRNA Functions Using a Collection of Targeted Knockout Mutations. *Dev. Cell* **31**, 784–800 (2014).
643. Shirahama, S., Miki, A., Kaburaki, T. & Akimitsu, N. Long Non-coding RNAs Involved in Pathogenic Infection. *Front. Genet.* **11**, 454 (2020).
644. Hoe, C.-H., Raabe, C. A., Rozhdestvensky, T. S. & Tang, T.-H. Bacterial sRNAs: Regulation in stress. *Int. J. Med. Microbiol.* **303**, 217–229 (2013).
645. Le Pabic, H., Germain-Amiot, N., Bordeau, V. & Felden, B. A bacterial regulatory RNA attenuates virulence, spread and human host cell phagocytosis. *Nucleic Acids Res.* **43**, 9232–9248 (2015).
646. Westermann, A. J., Gorski, S. A. & Vogel, J. Dual RNA-seq of pathogen and host. *Nat. Rev. Microbiol.* **10**, 618–630 (2012).
647. Westermann, A. J., Barquist, L. & Vogel, J. Resolving host–pathogen interactions by dual RNA-seq. *PLOS Pathog.* **13**, e1006033 (2017).
648. Westermann, A. J. & Vogel, J. Host-Pathogen Transcriptomics by Dual RNA-Seq. in *Bacterial Regulatory RNA* (eds. Arluison, V. & Valverde, C.) vol. 1737 59–75 (Springer New York, 2018).
649. Fletcher, J. R., Erwin, S., Lanzas, C. & Theriot, C. M. Shifts in the Gut Metabolome and *Clostridium difficile* Transcriptome throughout Colonization and Infection in a Mouse Model. *mSphere* **3**, (2018).
650. Pruss, K. M. & Sonnenburg, J. L. *C. difficile* exploits a host metabolite produced during toxin-mediated disease. *Nature* **593**, 261–265 (2021).
651. D’Auria, K. M. *et al.* *In Vivo* Physiological and Transcriptional Profiling Reveals Host Responses to *Clostridium difficile* Toxin A and Toxin B. *Infect. Immun.* **81**, 3814–3824 (2013).
652. Kansau, I. *et al.* Deciphering Adaptation Strategies of the Epidemic *Clostridium difficile* 027

- Strain during Infection through In Vivo Transcriptional Analysis. *PLOS ONE* **11**, e0158204 (2016).
653. Janoir, C. *et al.* Adaptive Strategies and Pathogenesis of *Clostridium difficile* from *In Vivo* Transcriptomics. *Infect. Immun.* **81**, 3757–3769 (2013).
654. Chen, X. *et al.* A Mouse Model of *Clostridium difficile*–Associated Disease. *Gastroenterology* **135**, 1984–1992 (2008).
655. Peltier, J. & Soutourina, O. Identification of c-di-GMP-Responsive Riboswitches. in *c-di-GMP Signaling* (ed. Sauer, K.) vol. 1657 377–402 (Springer New York, 2017).
656. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
657. Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211–3217 (2012).
658. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
659. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
660. Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15**, R46 (2014).
661. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
662. Varet, H., Brillet-Guéguen, L., Coppée, J.-Y. & Dillies, M.-A. SARTools: A DESeq2- and EdgeR-Based R Pipeline for Comprehensive Differential Analysis of RNA-Seq Data. *PLOS ONE* **11**, e0157022 (2016).
663. Jenior, M. L., Leslie, J. L., Young, V. B. & Schloss, P. D. *Clostridium difficile* Colonizes Alternative Nutrient Niches during Infection across Distinct Murine Gut Microbiomes. *mSystems* **2**, e00063-17 (2017).
664. Scaria, J. *et al.* *Clostridium difficile* Transcriptome Analysis Using Pig Ligated Loop Model Reveals Modulation of Pathways Not Modulated In Vitro. *J. Infect. Dis.* **203**, 1613–1620 (2011).
665. Reynolds, C. B., Emerson, J. E., de la Riva, L., Fagan, R. P. & Fairweather, N. F. The *Clostridium difficile* cell wall protein CwpV is antigenically variable between strains, but exhibits conserved aggregation-promoting function. *PLoS Pathog.* **7**, e1002024–e1002024 (2011).
666. Neumann-Schaal, M., Jahn, D. & Schmidt-Hohagen, K. Metabolism the Difficile Way: The Key to the Success of the Pathogen *Clostridioides difficile*. *Front. Microbiol.* **10**, 219–219 (2019).
667. Nawrocki, K. L., Wetzels, D., Jones, J. B., Woods, E. C. & McBride, S. M. Ethanolamine is a valuable nutrient source that impacts *Clostridium difficile* pathogenesis. *Environ. Microbiol.* **20**, 1419–1435 (2018).
668. Cernat, R. C. & Scott, K. P. Evaluation of novel assays to assess the influence of different iron sources on the growth of *Clostridium difficile*. *Anaerobe* **18**, 298–304 (2012).
669. Boudry, P. *et al.* Identification of RNAs bound by Hfq reveals widespread RNA partners and a sporulation regulator in the human pathogen *Clostridioides difficile*. *RNA Biol.* 1–22 (2021) doi:10.1080/15476286.2021.1882180.
670. Dennis, G. *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery.

Genome Biol. **4**, P3 (2003).

671. Heap, J. T. *et al.* The CloStron: Mutagenesis in Clostridium refined and streamlined. *J. Microbiol. Methods* **80**, 49–55 (2010).
672. Kuehne, S. A., Heap, J. T., Cooksley, C. M., Cartman, S. T. & Minton, N. P. CloStron-Mediated Engineering of Clostridium. in *Strain Engineering* (ed. Williams, J. A.) vol. 765 389–407 (Humana Press, 2011).
673. Cartman, S. T., Kelly, M. L., Heeg, D., Heap, J. T. & Minton, N. P. Precise Manipulation of the Clostridium difficile Chromosome Reveals a Lack of Association between the *tcdC* Genotype and Toxin Production. *Appl. Environ. Microbiol.* **78**, 4683–4690 (2012).
674. Ng, Y. K. *et al.* Expanding the Repertoire of Gene Tools for Precise Manipulation of the Clostridium difficile Genome: Allelic Exchange Using pyrE Alleles. *PLoS ONE* **8**, e56051 (2013).
675. Christiansen, J. K., Larsen, M. H., Ingmer, H., Søgaard-Andersen, L. & Kallipolitis, B. H. The RNA-Binding Protein Hfq of *Listeria monocytogenes*: Role in Stress Tolerance and Virulence. *J. Bacteriol.* **186**, 3355–3362 (2004).
676. Vuotto, C., Donelli, G., Buckley, A. & Chilton, C. Clostridium difficile Biofilm. in *Updates on Clostridium difficile in Europe* (eds. Mastrantonio, P. & Rupnik, M.) vol. 1050 97–115 (Springer International Publishing, 2018).
677. Buckley, A. M., Ewin, D., Moura, I. B., Wilcox, M. H. & Douce, G. R. *Insights into the regulatory mechanisms of Clostridioides difficile biofilm formation.*
<http://biorxiv.org/lookup/doi/10.1101/2021.02.19.431970> (2021)
doi:10.1101/2021.02.19.431970.
678. Sobrero, P. & Valverde, C. The bacterial protein Hfq: much more than a mere RNA-binding factor. *Crit. Rev. Microbiol.* **38**, 276–299 (2012).
679. Ridlon, J. M., Kang, D.-J. & Hylemon, P. B. Bile salt biotransformations by human intestinal bacteria. *J. Lipid Res.* **47**, 241–259 (2006).
680. Chowdhury, R., Sahu, G. K. & Das, J. Stress response in pathogenic bacteria. 12.
681. Ilhan, Z. E., Marcus, A. K., Kang, D.-W., Rittmann, B. E. & Krajmalnik-Brown, R. pH-Mediated Microbial and Metabolic Interactions in Fecal Enrichment Cultures. *mSphere* **2**, (2017).
682. Feliciano, J. R., Grilo, A. M., Guerreiro, S. I., Sousa, S. A. & Leitão, J. H. Hfq: a multifaceted RNA chaperone involved in virulence. *Future Microbiol.* **11**, 137–151 (2016).
683. Goh, S., Ong, P. F., Song, K. P., Riley, T. V. & Chang, B. J. The complete genome sequence of Clostridium difficile phage ϕ C2 and comparisons to ϕ CD119 and inducible prophages of CD630. *Microbiology* **153**, 676–685 (2007).
684. Moura, I., Spigaglia, P., Barbanti, F. & Mastrantonio, P. Analysis of metronidazole susceptibility in different Clostridium difficile PCR ribotypes. *J. Antimicrob. Chemother.* **68**, 362–365 (2013).
685. Mann, M., Wright, P. R. & Backofen, R. IntaRNA 2.0: enhanced and customizable prediction of RNA–RNA interactions. *Nucleic Acids Res.* **45**, W435–W439 (2017).
686. Eggenhofer, F., Tafer, H., Stadler, P. F. & Hofacker, I. L. RNApredator: fast accessibility-based prediction of sRNA targets. *Nucleic Acids Res.* **39**, W149–154 (2011).
687. Garvey, P., Fitzgerald, G. F. & Hill, C. Cloning and DNA sequence analysis of two abortive infection phage resistance determinants from the lactococcal plasmid pNP40. *Appl. Environ.*

- Microbiol.* **61**, 4321–4328 (1995).
688. Weinberg, Z. *et al.* Detection of 224 candidate structured RNAs by comparative analysis of specific subsets of intergenic regions. *Nucleic Acids Res.* **45**, 10811–10823 (2017).
689. Bouillaut, L., McBride, S. M. & Sorg, J. A. Genetic Manipulation of *Clostridium difficile*. Firmicutes (Low G+C Gram Positive). *Curr. Protoc. Microbiol.* **20**, (2011).
690. Kirk, J. A. & Fagan, R. P. Heat shock increases conjugation efficiency in *Clostridium difficile*. *Anaerobe* **42**, 1–5 (2016).
691. Woods, C. *et al.* A novel conjugal donor strain for improved DNA transfer into *Clostridium* spp. *Anaerobe* **59**, 184–191 (2019).
692. Bidnenko, E., Chopin, A., Ehrlich, S. D. & Chopin, M.-C. Activation of mRNA translation by phage protein and low temperature: the case of *Lactococcus lactis* abortive infection system AbiD1. *BMC Mol. Biol.* **10**, 4 (2009).
693. Rio, D. C., Ares, M., Hannon, G. J. & Nilsen, T. W. Purification of RNA Using TRIzol (TRI Reagent). *Cold Spring Harb. Protoc.* **2010**, pdb.prot5439 (2010).
694. Wolf, T., Kämmer, P., Brunke, S. & Linde, J. Two's company: studying interspecies relationships with dual RNA-seq. *Curr. Opin. Microbiol.* **42**, 7–12 (2018).
695. Yazdi, M., Bouzari, M. & Ghaemi, E. A. Genomic analyses of a novel bacteriophage (VB_PmiS-Isfahan) within Siphoviridae family infecting *Proteus mirabilis*. *Genomics* **111**, 1283–1291 (2019).
696. van Dam, S., Vösa, U., van der Graaf, A., Franke, L. & de Magalhães, J. P. Gene co-expression analysis for functional classification and gene–disease predictions. *Brief. Bioinform.* bbw139 (2017) doi:10.1093/bib/bbw139.
697. Galperin, M. Y. *et al.* COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res.* **49**, D274–D281 (2021).
698. Leonard, S. *et al.* APERO: a genome-wide approach for identifying bacterial small RNAs from RNA-Seq data. *Nucleic Acids Res.* **47**, e88–e88 (2019).
699. Schmerer, N. & Schulte, L. N. Long noncoding RNAs in bacterial infection. *WIREs RNA* **12**, (2021).
700. Lalaouna, D., Prévost, K., Eyraud, A. & Massé, E. Identification of unknown RNA partners using MAPS. *Methods San Diego Calif* **117**, 28–34 (2017).