

Skolkovo Institute of Science and Technology
Doctoral Program in Life Sciences

Université de Paris

Ecole doctorale 562 Bio Sorbonne Paris Cité (BioSPC)

**The CRISPR-Cas system of human pathogen
Clostridium difficile : function and regulation**

By **Anna Maikova**

Doctoral thesis

Supervised by Olga Soutourina and Konstantin Severinov

Date of the defense : 30/09/2019

Jury composition :

Dr. Harald Putzer	Université Paris Diderot - Paris 7	Co-chair
Prof. Mikhail Gelfand	Skoltech	Co-chair
Dr. Mart Krupovic	Sorbonne Université, Pasteur Institut	Jury member
Prof. Claire Janoir	Université Paris Sud - Paris 11	Jury member
Dr. Ekaterina Semenova	Waksman Institute for Microbiology, Rutgers	Jury member
Prof. Konstantin Lukyanov	Skoltech	Jury member
Prof. Olga Soutourina	Université Paris Sud - Paris 11	Supervisor
Prof. Konstantin Severinov	Skoltech	Supervisor

Université de Paris
En cotutelle avec Institut des Sciences et de la Technologie
de Skolkovo (Skoltech)

Ecole doctorale 562 Bio Sorbonne Paris Cité (BioSPC)
Et Programme Doctoral en Sciences de la Vie

**La fonction et la régulation du système CRISPR-
Cas chez un pathogène humain *Clostridium
difficile***

Par Anna Maikova

Thèse de doctorat de microbiologie et biologie moléculaire

Dirigée par Olga Soutourina et par Konstantin Severinov

Présentée et soutenue publiquement le 30/09/2019

Devant un jury composé de :

Dr. Harald Putzer Prof. Mikhail Gelfand	Université Paris Diderot - Paris 7 Skoltech	Président du jury Président du jury
Dr. Mart Krupovic Prof. Claire Janoir	Sorbonne Université, Pasteur Institut Université Paris Sud - Paris 11	Rapporteur de thèse Rapporteur de thèse
Dr. Ekaterina Semenova Prof. Konstantin Lukyanov	Waksman Institute for Microbiology, Rutgers Skoltech	Membre du jury Membre du jury
Prof. Olga Soutourina Prof. Konstantin Severinov	Université Paris Sud - Paris 11 Skoltech	Directrice de thèse Directeur de thèse

Table of Contents

Acknowledgments	5
Abstract.....	6
Publications	10
Abbreviations	11
List of Figures.....	13
List of Tables	16
Chapter 1. Literature review	17
1.1 Human enteropathogen <i>Clostridium difficile</i>	17
1.1.1 General characterization of <i>Clostridium difficile</i>	17
1.1.2 <i>C. difficile</i> epidemiology and infection cycle	19
1.1.3 <i>C. difficile</i> virulence factors.....	21
1.1.4 Regulatory small noncoding RNAs in <i>C. difficile</i>	24
1.2 CRISPR-Cas systems: functional aspects and diversity	27
1.2.1 Discovery and general description of CRISPR-Cas systems.....	27
1.2.2 Classification of CRISPR-Cas systems	28
1.2.3 General description of CRISPR-Cas systems defense mechanisms	33
1.2.4 Interference in type I CRISPR-Cas systems	34
1.2.5 Adaptation in type I CRISPR-Cas systems.....	36
1.3 <i>C. difficile</i> CRISPR-Cas system	39
1.3.1 Characterization of <i>C. difficile</i> CRISPR-Cas system.....	39
1.3.2 Regulation and potential alternative functions of <i>C. difficile</i> CRISPR-Cas system....	42
1.3.3 Potential applications of <i>C. difficile</i> CRISPR-Cas system.....	44
Chapter 2. Functionality of <i>C. difficile</i> CRISPR-Cas system.....	48
2.1 Introduction.....	48
2.2 Materials and methods	49
2.2.1 Bacterial strains and growth conditions	49
2.2.2 Construction of plasmids and conjugation into <i>C. difficile</i>	49
2.2.3 PAM libraries high throughput sequencing and data analysis.....	50
2.2.4 Plasmid conjugation efficiency assays.....	50
2.2.5 High throughput sequencing and data analysis of newly acquired spacers	51

2.2.5 Construction of the <i>C. difficile</i> 630 Δ erm Δ CD2975-2982 mutant and its verification by qRT-PCR	51
2.3 Results.....	52
2.3.1 Determining of PAM-sequences in <i>C. difficile</i>	52
2.3.1.1 Experimental model and construction of PAM libraries	52
2.3.1.2 PAM sequences determination	53
2.3.2 The functionality of CRISPR interference in <i>C. difficile</i>	56
2.3.2.1 Plasmid interference assays in <i>C. difficile</i> 630 Δ erm strain.....	56
2.3.2.2 Plasmid interference assays in <i>C. difficile</i> R20291 strain.....	59
2.3.3 The functionality of CRISPR-Cas system for adaptation in <i>C. difficile</i>	60
2.3.3.1 Experimental model of naïve adaptation assays in <i>C. difficile</i> 630 Δ erm	60
2.3.3.2 Detection of new spacer acquisition	62
2.3.3.3 General analysis of newly acquired spacers	64
2.3.3.4 Analysis of the distribution of spacer lengths and frequencies.....	65
2.3.3.5 Definition of PAM sequences, corresponding to acquired spacers	67
2.3.4 Role of multiple <i>cas</i> operons	68
2.3.4.1 Construction of the <i>C. difficile</i> 630 Δ erm full <i>cas</i> operon deletion mutant	68
2.3.4.2 Growth of the <i>C. difficile</i> 630 Δ erm Δ CD2975-2982 mutant.....	69
2.3.4.3 Plasmid interference assays in <i>C. difficile</i> 630 Δ erm Δ CD2975-2982 mutant	70
2.4 Discussion.....	71
2.5 Supplementary materials.....	74
Chapter 3. Regulation of <i>C. difficile</i> CRISPR-Cas system.....	77
3.1 Introduction.....	77
3.2 Materials and methods	79
3.2.1 Plasmid and bacterial strain construction and growth conditions.....	79
3.2.3 Light microscopy	80
3.2.4 RNA extraction, quantitative real-time PCR, and 5'/3'RACE	80
3.2.5 Subcellular localization of HA-tagged toxins by cell fractionation and Western blotting	80
3.2.6 <i>In silico</i> screening for potential new TA genes and CRISPR arrays co-localization ..	81

3.2.7 Construction of the <i>C. difficile</i> CDIP634 strain, overexpressing <i>dccA</i> gene	81
3.2.8 Plasmid conjugation efficiency assays.....	81
3.3 Results.....	82
3.3.1 Co-regulation of <i>C. difficile</i> CRISPR-Cas system and type I toxin-antitoxin systems adjacent to CRISPR arrays	82
3.3.1.1 Identification of toxin-antitoxin system candidates in <i>C. difficile</i> genome ...	82
3.3.1.2 Functionality of toxin-antitoxin systems in <i>C. difficile</i>	85
3.3.1.3 Expression analysis of TA and CRISPR-Cas systems.....	91
3.3.1.4 Genomic analysis of TA and CRISPR arrays co-localization	93
3.3.2 Regulation of <i>C. difficile</i> CRISPR-Cas system by c-di-GMP	94
3.3.2.1 Identification of the c-di-GMP-I riboswitch adjacent to <i>C. difficile</i> 630 Δ <i>erm</i> CRISPR12 array.....	94
3.3.2.2 Role of high c-di-GMP intracellular levels on <i>C. difficile</i> 630 Δ <i>erm</i> CRISPR12 array functionality	95
3.3.2.3 <i>C. difficile</i> 630 Δ <i>erm</i> CRISPR-Cas system expression under high c-di-GMP intracellular levels.....	97
3.4 Discussion.....	98
3.4.1 Discovery of new type I TA modules, associated and co-regulated with <i>C. difficile</i> CRISPR-Cas system	98
3.4.2 Role of c-di-GMP in <i>C. difficile</i> CRISPR-Cas system regulation	101
3.5 Supplementary materials.....	103
Chapter 4. Using endogenous CRISPR-Cas system for genome editing in the human pathogen <i>C. difficile</i>	113
4.1 Introduction.....	113
4.2 Materials and methods	116
4.2.1 Bacterial strains, plasmids and growth conditions.....	116
4.2.2 Plasmid construction and conjugation into <i>C. difficile</i>	117
4.2.3 Deletion of the <i>hfq</i> gene and validation of Δ <i>hfq</i> mutants	118
4.2.4 RNA extraction and qRT-PCR	118
4.2.5 Protein extract preparation and Western blotting	118

4.2.6 Sporulation assay	119
4.3 Results.....	121
4.3.1 Construction of targeting mini-array plasmids and verification of their functionality	121
4.3.2 Construction of the genome editing plasmid and deletion of the <i>hfq</i> gene of <i>C. difficile</i> 630 Δ <i>erm</i> and R20291	124
4.3.3 Validation and complementation of <i>hfq</i> deletion strains	126
4.3.4 Sporulation assay of <i>C. difficile</i> 630 Δ <i>erm</i> Δ <i>hfq</i> mutants.....	127
4.4 Discussion.....	128
4.5 Supplementary materials.....	133
Chapter 5. Conclusions and perspectives	135
Bibliography	139
Résumé.....	163
Annex. Article 1.....	167
Annex. Article 2.....	189
Annex. Article 3.....	199

Acknowledgments

First of all, I would like to thank my supervisors, Prof. Konstantin Severinov and Prof. Olga Soutourina, for giving me an opportunity to undertake this long exciting journey of the joint PhD program. Without their guidance, supportive advice and help it would be impossible to finish this interesting Thesis. I also want to thank all the jury members: Prof. Harald Putzer, Prof. Mikhail Gelfand, Dr. Mart Krupovic, Prof. Claire Janoir, Dr. Ekaterina Semenova and Prof. Konstantin Lukyanov for their time and critical review of the Thesis manuscript.

No less I would like to acknowledge my colleagues from two wonderful countries: Russia and France, and from three beautiful cities: Paris, Saint-Petersburg, and Moscow. I wish to thank Skoltech PhD student Anna Shiriaeva and research intern Aleksandra Vasileva for their inestimable help in bioinformatic analysis of my experimental data. I would like to commend Skoltech Center of Life Sciences team for support during all my PhD studies. I am thankful for my colleagues from the Nanobiotechnologies Center of Peter the Great St. Petersburg Polytechnic University: Dr. Mikhail Khodorkovsky, Dr. Sergey Murashov, Dr. Maria Yakunina, Aleksey Vedyajkin, Natalia Morozova, Dr. Anton Sabantsev, Iana Fedorova, Dr. Maria Sokolova and Tatiana Zubko for the opportunity to start my work and continue it during the periods of my stay in Russia. I would like to thank the Laboratory of Pathogenesis of Bacterial Anaerobes team from the Pasteur Institute: Dr. Bruno Dupuy, Dr. Pierre Boudry, Prof. Isabelle Martin-Verstraete, Dr. Nicolas Kint, Dr. Marc Monot, Dr. Elodie Cuenot, Dr. Carolina Feliciano, Dr. Isabelle Poquet and Audrey Hamiot for their help and useful discussions in the beginning of my work with a tricky anaerobic bacterium *C. difficile*. I also wish to thank the Regulatory RNAs in Clostridia team from the I2BC Institute: Anaïs Boutserin, Victor Kreis, Dr. Johann Peltier and Emma Piattelli for the inestimable support during my further experimental work with *C. difficile*.

Finally, I want to thank my family and friends from Russia and France. Without their support and patience, it would be very difficult to go all the way.

Abstract

In English

Title: The CRISPR-Cas system of human pathogen *Clostridium difficile*: function and regulation

Clostridium difficile (the novel name – *Clostridioides difficile*) is a Gram-positive, strictly anaerobic spore forming bacterium, found in soil and aquatic environments as well as in mammalian intestinal tracts. *C. difficile* is one of the major pathogenic clostridia. This bacterium has become a key public health issue associated with antibiotic therapy in industrialized countries. *C. difficile*-associated diarrhoea is currently the most frequently occurring nosocomial diarrhoea in Europe and worldwide. Since the last decade the number of severe infection forms has been rising due to emergence of the hypervirulent and epidemic strains as ribotype 027 R20291 strain. *C. difficile* infection causes diarrhoea, colitis and even death. Many aspects of *C. difficile* pathogenesis remain poorly understood. Particularly, the molecular mechanisms of its adaptation to changing conditions inside the host are to be scrutinized.

During the infection cycle *C. difficile* survives in bacteriophage-rich gut communities possibly by relying on some special systems that control the genetic exchanges favored within these complex environments. During the last decade, CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR-associated) systems of adaptive prokaryotic immunity against exogenic genetic elements has become the center of interest among various anti-invader bacterial defense systems.

Previous studies revealed the presence of abundant and diverse CRISPR RNAs in *C. difficile*. *C. difficile* has an original CRISPR system, which is characterized by the presence of an unusually large set of CRISPR arrays (12 arrays in the laboratory 630 Δ erm strain and 9 ones in the hypervirulent R20291 strain), of two or three sets of *cas* genes conserved in the majority of sequenced *C. difficile* genomes and the prophage location of several CRISPR arrays. However, the role CRISPR-Cas plays in the physiology and infectious cycle of this important pathogen remains obscure.

The general aims of this work run as follows:

1) to investigate the role and the functionality of *C. difficile* CRISPR-Cas system in the interactions with foreign DNA elements (such as plasmids), 2) to reveal the way *C.*

difficile CRISPR-Cas system expression is regulated and functions in different states of bacterial culture, including its response to stresses.

In the present PhD thesis the functionality of *C. difficile* CRISPR-Cas system was investigated (Chapter 2). Through conjugation efficiency assays defensive function (CRISPR interference) of *C. difficile* CRISPR-Cas system was demonstrated. The correlation between the previously known levels of expression of CRISPR RNAs and the observed levels of interference has also been shown. Moreover, through the series of interference experiments the functionality of PAMs (protospacer adjacent motifs) was confirmed, which have already been predicted *in silico*. Additionally, the general functional PAM consensus was determined using PAM libraries experiments. Furthermore, an adaptive function of *C. difficile* CRISPR-Cas system was shown for laboratory strain. The role of multiple *cas* operons in *C. difficile* CRISPR functionality is also demonstrated in this Chapter.

In Chapter 3 the link between *C. difficile* CRISPR-Cas system and a new type I toxin-antitoxin system is demonstrated, as well as a possible co-regulation under biofilm and stress conditions of CRISPR-Cas system and these toxin-antitoxin modules. This Chapter also defines a possible role of c-di-GMP in regulation of *C. difficile* CRISPR-Cas system.

Additionally, Chapter 4 describes the utilization of endogenous *C. difficile* CRISPR-Cas system as a novel tool for genome editing in *C. difficile*.

Altogether, the obtained data highlight the original features of active *C. difficile* CRISPR-Cas system and demonstrate its biotechnological potential.

Keywords: *Clostridium difficile*, CRISPR, CRISPR interference, CRISPR adaptation, I-B subtype CRISPR-Cas system, CRISPR regulation, toxin-antitoxin system, genome editing

En français

Titre: La fonction et la régulation du système CRISPR-Cas chez un pathogène humain
Clostridium difficile

Clostridium difficile (nouveau nom *Clostridioides difficile*) est une bactérie à Gram-positif, sporulante, anaérobie stricte, présente dans le sol et les environnements aquatiques, ainsi que dans le tractus intestinal des mammifères. *C. difficile* est l'un des principaux clostridies pathogènes. Cette bactérie est devenue un vrai problème de santé publique associé à l'antibiothérapie dans les pays industrialisés. La diarrhée associée à *C. difficile* est actuellement la diarrhée nosocomiale la plus fréquente en Europe et dans le monde. Depuis la dernière décennie, la proportion de formes d'infections graves a augmentée en raison de l'émergence des souches hypervirulantes et épidémiques comme la souche R20291 de ribotype 027. L'infection à *C. difficile* provoque la diarrhée, la colite et même la mort. De nombreux aspects de la pathogenèse de *C. difficile* restent mal compris. En particulier, les mécanismes moléculaires de son adaptation aux conditions changeantes de l'hôte doivent être examinés.

Durant le cycle d'infection, *C. difficile* survit dans des communautés intestinales riches en bactériophages, en utilisant des systèmes qui contrôlent les échanges génétiques favorisés dans ces environnements complexes. Au cours de la dernière décennie, les systèmes CRISPR (*clustered regularly interspaced short palindromic repeats*)-Cas (associés aux CRISPR) d'immunité adaptative chez les procaryotes contre des éléments génétiques exogènes sont devenus le centre d'intérêt scientifique parmi les divers systèmes de défense bactérienne.

Des études antérieures ont révélé la présence d'ARN CRISPR abondants chez *C. difficile*. Cette bactérie possède un système CRISPR original, caractérisé par la présence d'un grand nombre de cassettes CRISPR (12 dans la souche 630 Δ *erm* et 9 dans la souche hypervirulante R20291), de deux ou trois opérons *cas* conservés dans la majorité des génomes séquencés de *C. difficile* et la localisation au sein des prophages de plusieurs cassettes CRISPR. Cependant, le rôle de CRISPR-Cas dans la physiologie et le cycle infectieux de cet important pathogène reste obscur.

Les objectifs de ce travail sont les suivants:

1) étudier le rôle et la fonctionnalité du système CRISPR-Cas de *C. difficile* dans les interactions avec des éléments d'ADN étrangers (tels que les plasmides), 2) révéler la

manière dont le système CRISPR-Cas de *C. difficile* est régulé et fonctionne dans des conditions de culture bactérienne différentes, incluant la réponse aux stress.

Dans la présente thèse, la fonctionnalité du système CRISPR-Cas de *C. difficile* a été étudiée (chapitre 2). Grâce à des tests d'efficacité de conjugaison, la fonction défensive (en interférence) du système CRISPR-Cas a été démontrée. La corrélation entre les niveaux d'expression des ARN CRISPR et les niveaux d'interférence observés a également été montrée. De plus, grâce à la série d'expériences d'interférence, la fonctionnalité des motifs PAM (*protospacer adjacent motifs*) a été confirmée en accord avec des prédictions *in silico*. Le consensus fonctionnel de PAM a été déterminé expérimentalement avec les bibliothèques des plasmides. La fonction adaptative du système CRISPR-Cas de *C. difficile* a été également démontrée pour la souche de laboratoire. Le rôle de plusieurs opérons *cas* dans la fonctionnalité du système CRISPR de *C. difficile* est démontré aussi dans ce chapitre.

Le chapitre 3 montre le lien entre le système CRISPR-Cas et un nouveau système toxine-antitoxine de type I, ainsi que leur possible co-régulation dans des conditions de biofilm et de stress. Ce chapitre définit également le rôle possible du c-di-GMP dans la régulation du système CRISPR-Cas de *C. difficile*. De plus, le chapitre 4 décrit l'utilisation du système CRISPR-Cas endogène comme nouvel outil pour la rédaction du génome de *C. difficile*.

En conclusion, les données obtenues mettent en évidence les caractéristiques originales du système CRISPR-Cas actif de *C. difficile* et démontrent son potentiel biotechnologique.

Mots clefs: *Clostridium difficile*, CRISPR, interférence CRISPR, adaptation CRISPR, système CRISPR-Cas de sous-type I-B, régulation de CRISPR, système toxine-antitoxine, rédaction du génome

Publications

1. **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. Oxford University Press*. 2018; 46: 4733-4751.
2. **Maikova A.**, Severinov K., Soutourina O. New insights into functions and possible applications of *Clostridium difficile* CRISPR-Cas system. *Front. Microbiol. Frontiers*. 2018; 9: 1740. (Minireview)
3. **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*. 2019. *Applied and Environmental Microbiology*. 85 (20). In press.
4. **Maikova A.**, Boudry P., Shiriaeva A., Vasileva A., Boutserin A., Semenova E., Severinov K., Soutourina O. New aspects of *Clostridium difficile* CRISPR-Cas system functionality. 2019. In preparation.
5. Peltier J., Hamiot A., Garneau J., Boudry P., **Maikova A.**, Fortier L., Dupuy B., Soutourina O. Type I toxin-antitoxin systems stabilize prophage regions in the human pathogen *Clostridium difficile*. 2019. In preparation.

equal contribution

Abbreviations

Amp	ampicillin
ATc	anhydrotetracycline
BHI	brain heart infusion medium
BHIS	BHI medium with supplements
bp	base pairs
CDI	<i>Clostridium difficile</i> infection
c-di-GMP	cyclic di-guanosine monophosphate
CDMM	<i>Clostridium difficile</i> minimal medium
cDNA	complementary DNA
Cfx	cefoxitin
Cm	chloramphenicol
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	CRISPR RNA
crRNP	CRISPR ribonucleoprotein complex
Cs	D-cycloserine
DNA	deoxyribonucleic acid
<i>E. coli</i>	bacteria <i>Escherichia coli</i>
HGT	horizontal gene transfer
LB	Luria-Bertani medium
MGE	mobile genetic elements
mRNA	messenger RNA
ncRNAs	non-coding RNAs
PAM	protospacer adjacent motif
PCR	polymerase chain reaction
pre-crRNA	precursor crRNA
P _{tet}	tetracycline-inducible promoter
qRT-PCR	quantitative reverse transcription polymerase chain reaction
RBS	ribosome-binding site
RNA	ribonucleic acid
RNAseq	RNA sequencing
TA	toxin-antitoxin system

TAP	tabacco acid pyrophosphotase
TSS	transcriptional start site
TY	tryptone-yeast extract medium

List of Figures

Figure 1.1. <i>C. difficile</i> morphology and pseudomembranous colitis caused by CDI	18
Figure 1.2. <i>C. difficile</i> infection cycle.....	21
Figure 1.3. Coordinated control of motility, biofilm formation, and other related processes via c-di-GMP-I and di-GMP-II-type riboswitches in <i>C. difficile</i>	27
Figure 1.4. General organization of a CRISPR-Cas locus.....	28
Figure 1.5. Classification of main CRISPR-Cas systems subtypes and the architecture of their loci	30
Figure 1.6. CRISPR-based defensive functions: immunity (interference) and immunization (adaptation)	34
Figure 1.7. Type I CRISPR-Cas systems molecular mechanism of interference	35
Figure 1.8. CRISPR adaptation in type I CRISPR-Cas systems	37
Figure 1.9. A model for protospacer integration into the CRISPR array	38
Figure 1.10. Schematic view of the chromosomal location of CRISPR arrays and the organization of <i>cas</i> operons in <i>C. difficile</i> strains 630 and R20291	42
Figure 2.1. PAM libraries experimental strategy for <i>C. difficile</i>	53
Figure 2.2. PAM-consensus WebLogos for <i>C. difficile</i> 630 Δ <i>erm</i> and R202091 strains ..	54
Figure 2.3. PAM wheels for <i>C. difficile</i> 630 Δ <i>erm</i> and R202091 strains	55
Figure 2.4. Experimental strategy for plasmid interference assays	57
Figure 2.5. Plasmid conjugation efficiency in <i>C. difficile</i> 630 Δ <i>erm</i> strain and expression levels of 630 Δ <i>erm</i> strain CRISPR arrays detected by RNAseq	58
Figure 2.6. Plasmid conjugation efficiency in <i>C. difficile</i> R20291 strain and expression level of R20291 strain CRISPR13 array, detected by RNAseq	60
Figure 2.7. Experimental plan of naïve adaptation assays in <i>C. difficile</i> 630 Δ <i>erm</i>	62
Figure 2.8. PCR analysis of naïve adaptation in <i>C. difficile</i> 630 Δ <i>erm</i>	63
Figure 2.9. The distribution of spacer lengths	69
Figure 2.10. The distribution of spacers, aligned to the chromosome, pCas1-2-4 plasmid and pCD630 plasmid	67
Figure 2.11. The distribution of PAM sequences, corresponding to acquired spacers	68
Figure 2.12. PCR analysis of obtained clones after the <i>codA</i> allele exchange genome editing to verify <i>C. difficile</i> 630 Δ <i>erm</i> Δ <i>CD2975-2982</i> mutants	69

Figure 2.13. Growth curves of <i>C. difficile</i> 630 Δ <i>erm</i> and <i>C. difficile</i> 630 Δ <i>erm</i> Δ CD2975-2982 strains in BHI medium at 37°C	69
Figure 2.14. Plasmid conjugation efficiency in <i>C. difficile</i> 630 Δ <i>erm</i> strain and <i>C. difficile</i> 630 Δ <i>erm</i> Δ CD2975-2982 mutant.	70
Figure 3.1. Genomic map of potential type I TA loci in association with CRISPR arrays in <i>Clostridium difficile</i> strain 630 Δ <i>erm</i>	84
Figure 3.2. Schematic representation of potential type I TA locus in <i>C. difficile</i> chromosome	85
Figure 3.3. Potential type I toxin proteins alignment and analysis	88
Figure 3.4. Effect of inducible toxin and TA overexpression for CD2517.1-RCd8 TA module near CRISPR 12 on growth in solid and liquid medium	89
Figure 3.5. Effect of inducible toxin and TA overexpression for CD2907.1-RCd9/CD0956.2-RCd10 TA module near CRISPR 16/15 (CRISPR 3/4) on growth in solid and liquid medium	90
Figure 3.6. Alignment of small proteins at the near proximity of CRISPR arrays in <i>C. difficile</i> strains	94
Figure 3.7. c-di-GMP-I riboswitch and a <i>cis</i> -antisense RNA adjacent to <i>C. difficile</i> 630 Δ <i>erm</i> CRISPR12 array	95
Figure 3.8. <i>C. difficile</i> CDIP634 strain, carrying <i>dccA</i> gene under the control of inducible P _{tet} promoter	96
Figure 3.9. Plasmid conjugation efficiency in <i>C. difficile</i> 630 Δ <i>erm</i> and <i>C. difficile</i> CDIP634 strains after their growth in ATc-containing medium	97
Figure 3.10. qRT-PCR analysis of the <i>C. difficile</i> 630 Δ <i>erm</i> CRISPR-Cas system expression in high c-di-GMP levels conditions	98
Figure S3.1. RNA-seq and TSS mapping profiles for the TA loci in <i>C. difficile</i> strain 630 Δ <i>erm</i>	103
Figure S3.2. Alignment of genomic TA regions	109
Figure S3.3. Effect of inducible HA-tagged proteins CD2517.1 and CD2907.1/CD0956.2 overexpression on growth in liquid medium	110
Figure S3.4. Light microscopy analysis of morphology changes induced by toxin overexpression for CD2517.1 TA module	110
Figure S3.5. RNA-seq profiles for the TA loci in <i>C. difficile</i> strain R20291	111

Figure S3.6. Number of the cells with lengths above the value of 630/p mean length with 2 standard deviations (10.5 μm) in strains, overexpressing CD2907.1 and CD2517.1 TA modules	111
Figure 4.1. General scheme of using endogenous CRISPR-Cas systems for genome editing in bacteria and archaea.....	116
Figure 4.2. Strategy for the design of the editing plasmids to delete the <i>hfq</i> gene in <i>C. difficile</i> 630 Δerm and R20291 strains	123
Figure 4.3. Different effects of the conjugation of mini-array and editing plasmids into <i>C. difficile</i> cells	125
Figure 4.4. Validation of <i>hfq</i> deletion mutants	126
Figure 4.5. Sporulation levels in <i>C. difficile</i> 630 Δerm wt/p, Δhfq /p, and Δhfq /p- <i>hfq</i> strains and the total amount of bacteria in CFUs	127
Figure 4.6. The general workflow for application of endogenous CRISPR-Cas-based genome editing method in <i>C. difficile</i>	132
Figure S4.1. Sequences used to construct mini-array and editing plasmids to delete <i>hfq</i> gene in <i>C. difficile</i>	133
Figure S4.2. Protein gels stained with the InstantBlue and used as a loading control for the Western blot analysis of wt/p, Δhfq -p, and Δhfq /p- <i>hfq</i> <i>C. difficile</i> strains	134
Figure 5.1. Conclusions and perspectives of <i>C. difficile</i> CRISPR-Cas system research	138

List of Tables

Table 1.1. Main subtypes of CRISPR-Cas systems and examples of system-harboring microorganisms and clostridial species	32
Table 2.1 Statistics of spacers acquired into CRISPR8 (CR8) and CRISPR9 (CR9) arrays	65
Table S2.1. Bacterial strains and plasmids used in Chapter 2	74
Table 3.1. Differential expression of TA and CRISPR-Cas systems revealed by transcriptome and/or qRT-PCR analysis	92
Table S3.1. Bacterial strains and plasmids used in Chapter 3	112
Table 4.1. Bacterial strains and plasmids used in Chapter 4	120

Chapter 1. Literature review

Several parts of this Chapter are published (Maikova et al., 2018b):

Maikova A., Severinov K., Soutourina O. New insights into functions and possible applications of *Clostridium difficile* CRISPR-Cas system. *Front. Microbiol. Frontiers*. 2018; 9: 1740.

1.1 Human enteropathogen *Clostridium difficile*

1.1.1 General characterization of *Clostridium difficile*

For the first time, *Clostridium difficile* (the novel name *Clostridioides difficile* (Oren and Rupnik, 2018)) was identified in 1935 by Hall and O'Toole (Hall and O'Toole, 1935). The bacterium was found in the normal gut microflora of newborns and named *Bacillus difficilis* as a result of difficulties of its insulation. *Clostridium difficile* is a Gram-positive, strictly anaerobic, motile and spore-forming bacterium. This species is ubiquitous and can be found in soil and aquatic environments, and in mammalian intestinal tracts. *C. difficile* is asymptotically carried by about 3-5% of the human population. *C. difficile* cells are characterized by size from 3 to 15 μm long, and they may have a small bulge at the end of the cell, which corresponds to the subterminal spore (Figure 1.1A). *C. difficile* colonies are opaque, non-hemolytic, irregular, or even slightly rhizoid (Figure 1.1B). In 1978 Barlett et al. showed the correlation between the presence of *C. difficile* and the emergence of pseudomembranous colitis (acute inflammation of the colon (Figure 1.1C)), which is often associated with antibiotic therapy (Bartlett et al., 1978).

The complete genome sequence of reference strain *C. difficile* 630 was obtained in 2006 (Sebahia et al., 2006). The chromosome is composed of 4,290,252 base pairs (bp) with 29% of G + C content (Sebahia et al., 2006) and contains 3,897 coding sequences (Monot et al., 2011). *C. difficile* genome is mosaic, where mobile genetic elements (MGE) constitute 11% of the chromosome (Sebahia et al., 2006). These MGE are potentially responsible for the acquisition of new genes for a better adaptation to the gut environment inside the host, including antimicrobial resistance, virulence, and the production of bacterial cell surface structures. In 2009, the comparative genomic analysis of *C. difficile* R20291 ribotype 027 strain provided a basis for the evolution of hypervirulence of this epidemic strain (Stabler et al., 2009). In particular, epidemic

R20291 strain contains five unique genomic regions, including a prophage region, and a significant number of genes, encoding transcriptional regulators and two-component systems. Using high throughput sequencing methods allowed to expand *C. difficile* genome databases (approximately 3,000 sequenced strains are currently available) and to undertake comparative and scalable genome analyzes. The available *C. difficile* genomes can be found at following databases: <https://www.ebi.ac.uk/ena> (Cairns et al., 2017; Dingle et al., 2014; Eyre et al., 2013; He et al., 2013; Kurka et al., 2014), <https://www.ncbi.nlm.nih.gov/> (He et al., 2010; Moura et al., 2014; Sebaihia et al., 2006), <http://www.genoscope.cns.fr/> (Moura et al., 2014). In one of these studies, the core genome of *C. difficile* was defined as about 3,000 genes (Stabler et al., 2009). Thus, the first complete sequencing of the *C. difficile* genome has opened a new genomic era in studies of this enteropathogen and enabled the development of new global approaches, including transcriptomic analysis.

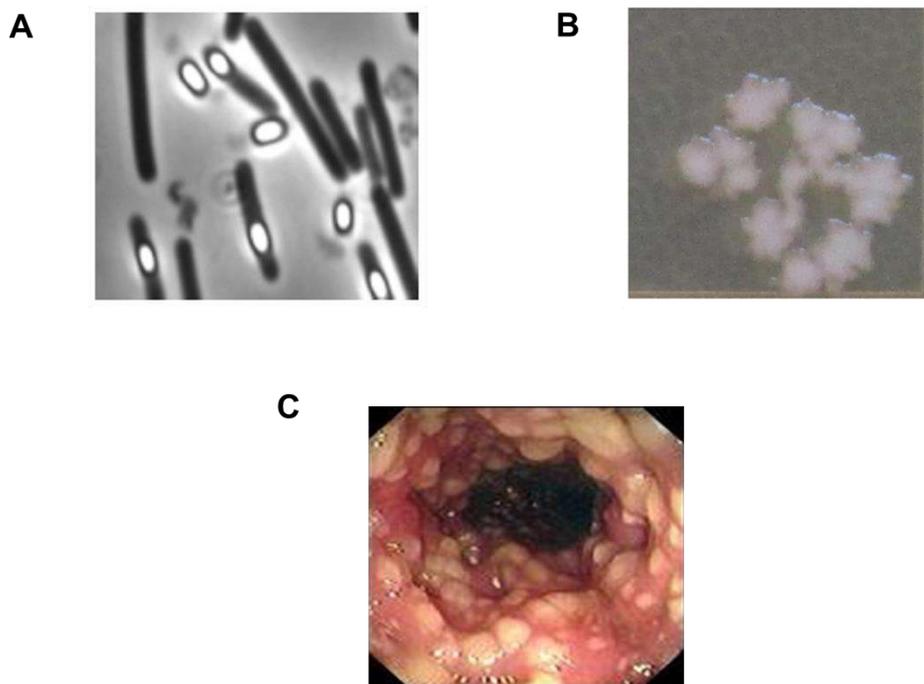


Figure 1.1. *C. difficile* morphology and pseudomembranous colitis caused by CDI. **A** – a light microscopy image of *C. difficile* cells, **B** – *C. difficile* colonies, **C** – an endoscopy image of the colon, affected by pseudomembranous colitis.

1.1.2 *C. difficile* epidemiology and infection cycle

C. difficile is an opportunistic pathogen that causes antibiotic therapy-associated diarrhea and pseudomembranous colitis in adults. This bacterium is found in the intestinal tract of 66% of newborns and young children without any symptoms of diarrhea. The general distribution of *C. difficile* carries is 3% in adults, and it increases in case of long-term hospitalization (Kazanowski et al., 2014). The epidemiology of *C. difficile* infection (CDI) has changed since the early 2000s. *C. difficile* infection cases have significantly increased in North America and Europe, with the emergence of clinical severity forms, characterized by septic shock, and megacolon perforation. Moreover, increased incidence of *C. difficile* antibiotic resistance has been recently reported (Banawas, 2018).

Data, provided by the Agency for Healthcare Research and Quality Healthcare and Utilization Project, demonstrates a total number of 336,600 hospitalizations with CDI diagnosis in the United States of America in 2009 (Lucado et al., 2006). The rate of hospitalization with a diagnosis of CDI increased from 5.6‰ to 12.7‰ between 2001 and 2011 in this country (Steiner et al., 2012). 93% of deaths associated with CDI occurred in patients over 65 years old (Miniño et al., 2010).

A retrospective study of CDI rates in Canada revealed that between 1991 and 2003, the number of pseudomembranous colitis diarrhea cases has increased by four times among the general population and by ten times since 1938. This study also showed that the appearance of symptoms in hospitalized patients had increased from 3‰ to 12‰ at the same period of time, and it was 43‰ in 2004 (Pepin et al., 2004).

In Europe, the progression of CDI is similar to other industrialized countries. The average incidence is 4.1‰ patients per hospital, and the virulence is increasing (Bauer et al., 2011). This phenomenon is partly related to the emergence and dissemination of so-called "hyper-virulent" or "epidemic" and fluoroquinolone-resistant strains such as the ribotype 027 strain, appeared in the north of France in 2006 (Coignard et al., 2006).

CDI is caused by oral transmission of *C. difficile* spores, which are resistant to heat, acids, and antibiotics (Figure 1.2). Spores are present in small quantities in different environments and in large amounts inside the health services that causes nosocomial or community-based infections. After the germination of spores inside the stomach, *C. difficile* colonizes the intestinal tract. The colonization process depends on the properties and the conditions of the colon microbiota. Alteration of microbiota following by the antibiotic therapy is a major risk factor for CDI; age and immunodepression are CDI risk

factors as well (Rupnik et al., 2009). All antibiotics are associated with an increased risk of CDI. However, some broad-spectrum drugs such as clindamycin, some β -lactams (cephalosporins, and carbapenem) and quinolones demonstrated a stronger association with the infection (Brown et al., 2013; Owens et al., 2008). *C. difficile* adheres to the mucus layer, covering the gut epithelial cells, using multiple adhesins. Then the bacterium penetrates this mucus layer using proteases (Denève et al., 2009). After this, *C. difficile* produces two toxins, TcdA and TcdB, which are the major virulence factors of this enteropathogen (Carroll and Bartlett, 2011). These two toxins cause lysis of enterocytes and a robust inflammatory reaction, which leads to diarrhea, pseudomembranous colitis, and even the colon perforation and a patient's death (Just et al., 1995; Vedantam et al., 2012). Additionally, *C. difficile* forms spores inside the gut, which will be released into the environment where they can potentially infect more persons. During the infection cycle, *C. difficile* metabolically adapts to changing environments and various stresses inside the host, such as hyperosmolarity, pH variations, and exposure to bile acids and antibiotics, or to cationic antimicrobial peptides produced by the host and/or by the microbiota (Abt et al., 2016). These adaptations allow this pathogen to successfully colonize the colon and survive in unfavorable conditions inside the gut. Additionally, this enteropathogen forms biofilms (Dapa et al., 2013; Nale et al., 2016; Soavelomandroso et al., 2017). *C. difficile* vegetative cells also interact with phages in phage-rich gut communities (Mick et al., 2013; Stern et al., 2012) (Figure 1.2).

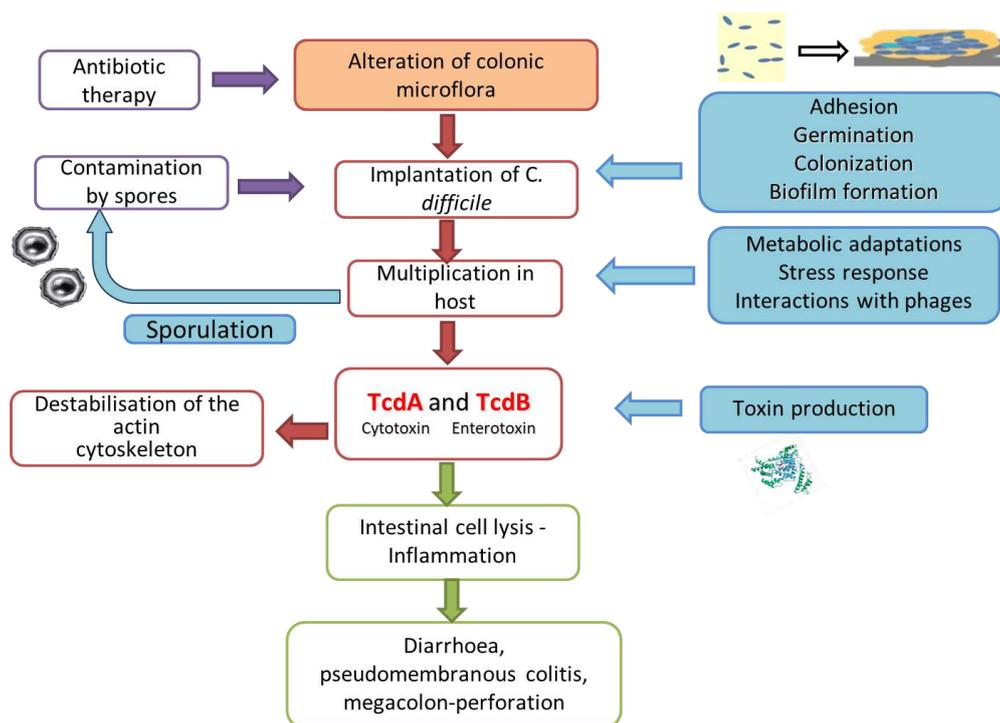


Figure 1.2. *C. difficile* infection cycle. The disruption of the colonic microflora by antimicrobial therapy allows colonization of the intestinal tract by pre-existing or newly acquired *C. difficile* spores that ultimately leads to infection. After spore germination, vegetative *C. difficile* cells multiply, and major virulence factors (TcdA and TcdB toxins) are produced causing alterations in the actin cytoskeleton of intestinal epithelial cells. This results in diarrhea, pseudomembranous colitis and megacolon-perforation in severe cases. During its infection cycle *C. difficile* metabolically adapts to changing environments inside the host, responds to various stresses, forms biofilms and interacts with bacteriophages.

1.1.3 *C. difficile* virulence factors

Several factors are involved in *C. difficile* virulence. They can be divided into 2 groups: factors, involved in colonization of the host and factors, involved in toxigenic stages. During the colonization step *C. difficile* develops inside colonic environments, and it includes avoidance of host immunity defense, multiplication of vegetative cells and adherence to epithelium mucosa. These events are often associated with bacterial surface components. *C. difficile* expresses a large number of surface proteins (Wright et al., 2005) and some of them play key roles in the interaction with host.

Pathogenic bacteria often utilize surface proteins, associated with the peptidoglycan, to recognize elements of the extracellular matrix of the host cells. In Gram-positive bacteria, these structures are designated by the acronym MSCRAMMS

(microbial surface components recognizing adhesive matrix molecules). MSCRAMMS can bind the host extracellular matrix proteins such as fibronectin, fibrinogen, and collagen (Vengadesan and Narayana, 2011). *C. difficile* genome analysis identified several genes encoding extracellular host cell matrix binding proteins. Among them there was *fbp68* gene (Fibronectin-binding protein 68 kDa, also annotated as *fbpA*) encoding a fibronectin-binding protein, which could play an important role in *C. difficile* cells adhesion (Hennequin et al., 2003; Lin et al., 2011).

Many bacteria possess a surface layer called "S-layer," composed of glycoproteins, which completely covers the cell. S-layers are important for many biological functions such as cell adhesion or protection against phagocytosis. All *C. difficile* strains have an S-layer composed of two superimposed protein layers, where SlpA is the dominant protein (Calabi et al., 2002). SlpA plays a significant role in *C. difficile* cells adhesion *in vitro* (Calabi et al., 2002). It is also suggested that SlpA is essential for *C. difficile* growth (Dembek et al., 2015). Several other proteins with adhesion properties were also characterized in *C. difficile*: Cwp66 protein (Waligora et al., 2001), GroEL heat shock protein (Collignon et al., 2001) and CbpA collagen-binding protein (Tulli et al., 2013). Interestingly, *cwp66* and *slpA* genes are variable and contain a 10-kb cassette, which is capable of recombinational switching (Dingle et al., 2013).

To successfully colonize the host and to adapt to the new conditions after the colonization pathogenic bacteria usually form biofilms. *C. difficile* forms biofilms on different surfaces during its infection cycle (Dapa et al., 2013; Nale et al., 2016; Soavelomandroso et al., 2017). Within biofilms, *C. difficile* cells are immersed in the matrix composed of DNA, polysaccharides, and proteins, including the toxin A (Semenyuk et al., 2014; Soavelomandroso et al., 2017). *C. difficile* biofilms formation could play an important role in CDI development and its recurrence since they have been shown to promote the persistence for this enteropathogen (Crowther et al., 2014) and enhance the resistance of *C. difficile* to oxygen stress and antibiotics (Dawson et al., 2012; Semenyuk et al., 2014). *C. difficile* biofilm formation is controlled by central regulators Spo0A (a sporulation initiation regulator) and LuxS (a quorum sensing regulator) (Dapa et al., 2013; Dawson et al., 2012). Moreover, the second messenger c-di-GMP (cyclic di-guanosine monophosphate) has been shown to positively regulate the biofilm development in *C. difficile* 630 strain (Soutourina et al., 2013). c-di-GMP is a small signaling molecule with functions in controlling lifestyle changes in bacteria, like biofilms formation and switch to the virulence state (Römling, 2012). c-di-GMP is

synthesized from two GTP molecules by diguanylate cyclases (Ryjenkov et al., 2005) and hydrolyzed into pGpG or two GMP molecules by c-di-GMP phosphodiesterases (Schmidt et al., 2005; Tamayo et al., 2005). *C. difficile* encodes up to 37 c-di-GMP turnover enzymes, suggesting an important function of c-di-GMP for this bacterium (Bordeleau et al., 2011).

The main *C. difficile* virulence factors, involved in the toxigenic stage of the infection, are two large toxins A and B (TcdA and TcdB), which are members of a large clostridial toxins family. This family is characterized by 250 to 300 kDa toxins, and it includes the lethal and hemorrhagic toxins of *Clostridium sordelii*, the α -toxin of *Clostridium novyi* and the Tpel cytotoxin of *Clostridium perfringens* (Carter et al., 2012). Both toxins have four similar functional domains: the N-terminal domain with the glucosyltransferase activity, a highly conserved cysteine protease domain, the central domain with a hydrophobic region and the C-terminal domain, which consists of repetitive sequences and binds to the surface receptor of the host cells (Pruitt and Lacy, 2012). The targets of TcdA and TcdB glucosyltransferase activity are Rho and Ras family GTPases of the host cells, which are involved in maintenance of the cytoskeleton and cell-to-cell adhesion. TcdA and TcdB inactivate these GTPases leading to alteration of the actin cytoskeleton and the cell death by apoptosis and necrosis.

TcdA and TcdB are encoded by *tcdA* and *tcdB* genes, localized in the 19.6 kb chromosomal region, called pathogenicity locus (PaLoc). PaLoc has been found in only toxigenic *C. difficile* strains (Braun et al., 1996). In addition to *tcdA* and *tcdB*, PaLoc contains *tcdR* and *tcdC* genes, which encode transcriptional regulators of the toxin genes (Mani and Dupuy, 2001; Matamouros et al., 2007). TcdR is an alternative RNA polymerase sigma factor that induces toxin genes expression (Mani and Dupuy, 2001). The role of TcdC remains unclear, though it has been proposed to act as a negative regulator of TcdA and TcdB synthesis (Janoir, 2016). PaLoc also contains *tcdE* gene, encoding a putative holin, involved in TcdA and TcdB secretion (Govind and Dupuy, 2012). Furthermore, the expression of TcdA and TcdB toxins is also regulated by environmental conditions and global regulators. A pleiotropic regulator CcpA, which is involved in utilization of alternative carbon sources, mediates glucose-dependent repression of TcdA and TcdB expression (Antunes et al., 2012). In nutrient-deficient conditions, the expression of PaLoc genes is also repressed by the global transcriptional regulator CodY, which plays an important role in response to nutrient availability in the environment (Dineen et al., 2007). A transcriptional regulator Spo0A, which is necessary

for the initiation of sporulation, negatively controls toxin expression in *C. difficile* (Mackin et al., 2013; Pettit et al., 2014). Additionally, the expression of *tcdA* and *tcdB* is affected indirectly by c-di-GMP being repressed in high c-di-GMP intracellular levels through the control of flagellar-specific sigma factor SigD expression (McKee et al., 2013).

Some strains of *C. difficile* produce a binary toxin belonging to the family of clostridial ADP-ribosylating toxins. The toxin composed of two subunits: the catalytic subunit CDTa with an actin specific ADP ribosyltransferase activity, and the binding CDTb part. The binary toxin is encoded by a CdtLoc locus, which includes *cdtA* and *cdtB* genes and their positive transcriptional regulator *cdtR* (Carter et al., 2007). *C. difficile* CdtA inhibits the actin polymerization in the cytosol of the target cell by transferring ADP-ribose. The CdtB subunit is involved in the reception of the host cell and it allows the uptake of CdtA into host cells (Gerding et al., 2014). The role of binary toxin in CDI still remains poorly understood. To date, the correlation between the presence of binary toxin and severe CDI cases has been supposed, and this toxin might potentiate the toxicity of TcdA and TcdB (Gerding et al., 2014).

1.1.4 Regulatory small noncoding RNAs in *C. difficile*

Non-coding RNAs (ncRNAs) are present in all Domains of Life and play various roles in living organisms. The first ncRNAs (transfer RNAs (tRNAs) and the rRNA ribosomal RNAs) were identified in the 1960s. Since that time a huge amount of different ncRNAs and their associated physiological roles have been discovered. In 2000s, new sequencing techniques allowed to demonstrate the existence of transgenic intergenic regions in all studied species and characterized a great diversity of ncRNAs mechanisms of actions and functions. Until recent time, almost all studies of cellular functions had been focused mainly on protein regulators. Nevertheless, RNA molecules have been shown to have adaptive and physiological functions. Research on regulatory RNAs has resulted in the awarding of the Nobel Prize in Physiology and Medicine to Andrew Z. Fire and Craig C Mello in 2006 for the discovery of the mechanisms of RNA interference (Fire and Mello, 2006).

In bacteria, ncRNAs play a crucial role in diverse metabolic, physiological, and pathogenic processes and adaptive responses (Wagner and Romby, 2015). In particular, small ncRNAs have been recently found in many pathogenic bacteria (Gripenland et al., 2010; Romby and Charpentier, 2010). Regulatory ncRNAs may contribute to several

steps during the infection cycle. These small ncRNAs control their targets using various mechanisms such as RNA/RNA duplex formation with mRNA targets, binding to proteins, interaction with double-stranded DNA or RNA (CRISPR, clustered regularly interspaced short palindromic repeats RNAs), and direct binding to low-molecular-weight effector molecules for riboswitches (Brantl, 2012a). In a recent study, more than 200 ncRNAs were identified in *C. difficile*, by combining *in silico* analysis, RNAseq and genome-wide promoter mapping (Soutourina et al., 2013). They include *cis*-acting and *trans*-acting riboregulators, i.e. antisense RNAs, CRISPR RNAs and riboswitches.

The most well-studied type of small regulatory ncRNA acts by modulating the translation and/or stability of specific mRNA targets in response to changes in the environment (Waters and Storz, 2009). These ncRNAs are divided into *cis*-encoded RNAs and *trans*-encoded RNAs. *Cis*-encoded RNAs are thus fully complementary to their targets and transcribed from a DNA strand opposite to the one from which the target mRNA is transcribed (Brantl, 2007). In contrast, *trans*-encoded RNAs are transcribed from separate loci and are only partially complementary to target mRNAs (Waters and Storz, 2009). The RNA-chaperone Hfq protein is frequently required for *trans*-encoded sRNA-mediated control (Vogel and Luisi, 2011). A recent study of Hfq depletion suggested the pleiotropic role of this protein in *C. difficile* physiology, including cell morphology, sporulation, biofilm formation and response to stresses, a unique feature in Gram-positive bacteria (Boudry et al., 2014). The accumulation of several ncRNA was altered under Hfq depletion and Hfq can bind selected ncRNAs supporting its involvement in their function (Boudry et al., 2014).

Protein-binding small ncRNAs directly modify the function of their targets (Pichon and Felden, 2007). For instance, the widely distributed 6S RNA imitates an open promoter complex and acts as a promoter decoy for RNA polymerase holoenzyme containing major sigma 70 factor. In this way, 6S RNA globally regulates transcription during adaptation to stationary phase of growth (Wassarman, 2007).

CRISPR RNAs (crRNAs) contain short regions, which are complementary to bacteriophage and other MGE sequences. crRNAs form a complex with Cas (CRISPR-associated) proteins that interferes with foreign DNA invasion by its recognizing and targeting for destruction (Bhaya et al., 2011). The more detailed description of crRNAs and CRISPR-Cas systems features and functions is presented in the part 1.2 of this Chapter.

Riboswitches are *cis*-acting elements, commonly located in the 5'-untranslated region of some mRNAs. Through binding of specific ligands, riboswitches conformationally change, which leads to modifications of transcription termination or translation (Nudler and Mironov, 2004). These regulatory ncRNAs are often involved in the control of vitamin, amino acid, and nucleotide base biosynthesis gene expression in bacteria (Soutourina et al., 2013). One of the most abundant riboswitch classes found in *C. difficile* is c-di-GMP-binding ones. There are two types of c-di-GMP-responsive riboswitches that differ in structure and the recognition of ligand. Type I c-di-GMP-dependent riboswitches (c-di-GMP-I) contain a conservative “GEMM” RNA domain and positively or negatively control their target gene expression through termination of transcription or anti-termination (Sudarsan et al., 2008). The second type of these riboswitches (c-di-GMP-II) carry a distinct RNA motif, and they are widespread in the Clostridiales family (Lee et al., 2010). c-di-GMP-II riboswitches positively regulate target gene expression through antitermination of transcription or modulation of the translation in association with the group I self-splicing intron (Bordeleau et al., 2015; Chen et al., 2011; Lee et al., 2010). To date, active expression of 12 c-di-GMP-I riboswitches and 4 c-di-GMP-II riboswitches was detected in *C. difficile* (Soutourina et al., 2013). These regulatory ncRNAs are suggested to be involved in coordinated control of motility and biofilm formation in *C. difficile* depending on c-di-GMP levels inside cells (Figure 1.3). High c-di-GMP levels repress type I riboswitches, which positively regulate motility (Purcell et al., 2012), toxin production (McKee et al., 2013) and proteolysis of adhesins (Peltier et al., 2015). Conversely, c-di-GMP-II riboswitches, positively controlling adhesion, aggregation, and biofilm formation (Bordeleau et al., 2015), are activated by high c-di-GMP levels (Figure 1.3).

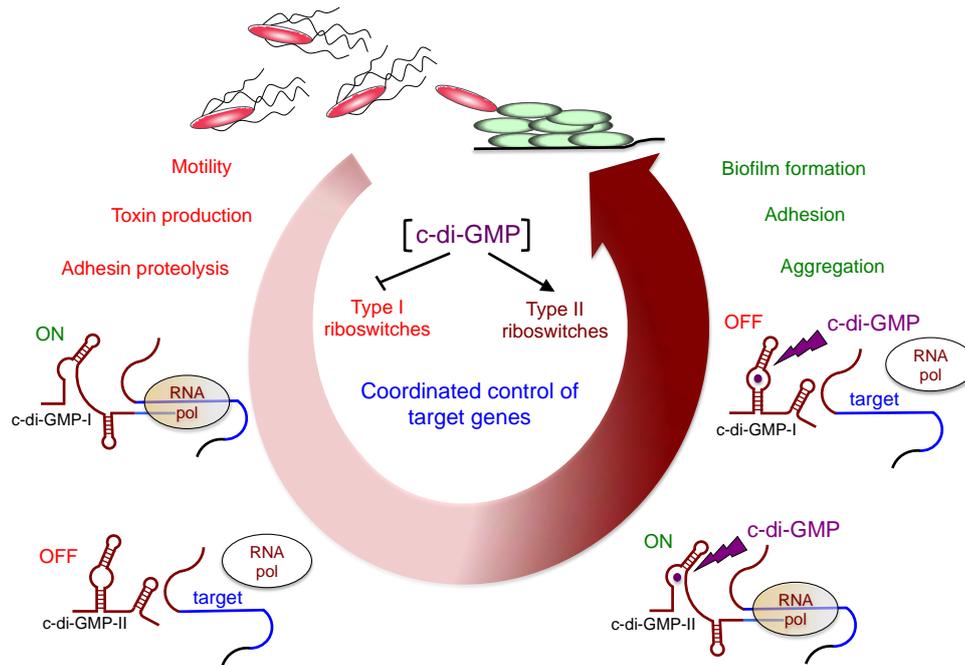


Figure 1.3. Coordinated control of motility, biofilm formation, and other related processes via c-di-GMP-I and di-GMP-II-type riboswitches in *C. difficile*. High intracellular c-di-GMP levels (green color) are associated with non-motile forms within biofilm communities, and low intracellular levels (green color) are associated with motile planktonic forms. In the presence of c-di-GMP (right-hand part), binding of c-di-GMP to c-di-GMP-I riboswitches result in premature termination of transcription and the “OFF” state of target gene expression. In contrast, the c-di-GMP-II target genes switch to the “ON” state through c-di-GMP-binding leading to the synthesis of full-length mRNA transcripts by RNA polymerase (RNA pol). Under low levels of c-di-GMP conditions (left part), c-di-GMP-I riboswitches transfer the target gene read-through transcription to the “ON” mode. Respectively, the expression of c-di-GMP-II targets is turned “OFF”. Adapted from (Soutourina, 2017) with permission.

1.2 CRISPR-Cas systems: functional aspects and diversity

1.2.1 Discovery and general description of CRISPR-Cas systems

The CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR-associated) systems are adaptive prokaryotic immune systems against phages and other MGE, such as plasmids and transposons (Sorek et al., 2013). These defensive systems are found almost in all sequenced archaeal genomes and in about a half of bacterial genomes (Grissa et al., 2007). CRISPR-Cas systems are made up of CRISPR arrays and *cas* gene operons. CRISPR arrays, in their turn, consist of short direct repeat

sequences (20-40 bp) separated by variable spacers. Spacers are often complementary to viral and other MGE (Shmakov et al., 2017). CRISPR arrays also contain leader regions, which dispose transcriptional start sites (TSS) for their expression (Figure 1.4).

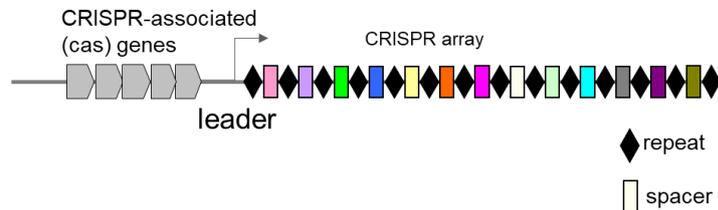


Figure 1.4. General organization of a CRISPR-Cas locus. CRISPR-Cas systems comprise two main components: a *cas* gene set and a CRISPR array, composed of spacers and repeats. The broken arrow indicates a leader region.

For the first time, CRISPR arrays were described in *E. coli* by Japanese scientists in 1987 (Ishino et al., 1987). They found an unusual set of 29-nucleotide palindromic repeats, separated by unique sequences (spacers) of 32 nucleotides in length with an unknown function. Further, similar genomic regions were discovered in other bacteria and archaea (Groenen et al., 1993; Mojica et al., 1995, 2000). In 2002 these interspaced repetitive genomic structures were defined as CRISPR (Jansen et al., 2002b), and *cas* genes were also described in the same year (Jansen et al., 2002a). In 2005, almost after 20 years after the first detection of CRISPR loci, several studies reported that CRISPR spacers derive from bacteriophages and other foreign genetic elements (Bolotin et al., 2005; Mojica et al., 2005; Pourcel et al., 2005). Additionally, the role of CRISPR loci in providing immunity against DNA invaders was suggested (Bolotin et al., 2005; Mojica et al., 2005). In 2006, the RNA-interference-based mechanism of CRISPR-Cas systems action and classification of CRISPR-Cas systems, based on amino acid sequences of Cas proteins, was proposed by Koonin's laboratory (Makarova et al., 2006). Finally, in 2007, an adaptive immunity function through new spacer acquisition was demonstrated for CRISPR-Cas systems (Barrangou et al., 2007).

1.2.2 Classification of CRISPR-Cas systems

CRISPR-Cas systems are highly diverse; moreover, they are able to undergo rapid coevolution with viruses. This peculiarity influences CRISPR loci architectures, especially *cas* gene sets (Takeuchi et al., 2012). The variability of *cas* gene sets is

considered to be the basis of CRISPR-Cas systems classification (Makarova et al., 2011b). According to the recent classification, all investigated CRISPR-Cas systems are divided into two entirely different classes characterized by a certain interference (effector) *cas* gene module structure (Koonin et al., 2017). These classes, in their turn, are divided into six types and 33 subtypes (Figure 1.5, only main CRISPR types are presented).

Class 1 includes the most abundant and diverse type I and type III CRISPR-Cas systems as well as quite a rare type IV lacking adaptation genes; all these types of CRISPR-Cas systems are found in both archaeal and bacterial genomes. Effector complexes of type I and type III include Cas5, Cas7, Cas8 (in type I) and Cas10 (in type III) proteins (Koonin, 2017) and bind mature crRNA. For crRNA processing Cas6 family proteins are necessarily required to be present in the class 1 CRISPR-Cas systems (Figure 1.5) (Charpentier et al., 2015). Type I systems are also characterized by Cas3 proteins, which degrade foreign DNA at the last step of interference (Brouns et al., 2008). Class 2 includes type II, type V and type VI CRISPR-Cas systems, all of which possess the effector modules consisting of only one multi-domain protein (Figure 1.5). Cas9 protein of type II is the most characterized one; this protein is widely used in genome editing techniques (Wang et al., 2016). This class of CRISPR systems are present in only 10% of bacteria and cannot be found in archaea (Makarova et al., 2015).

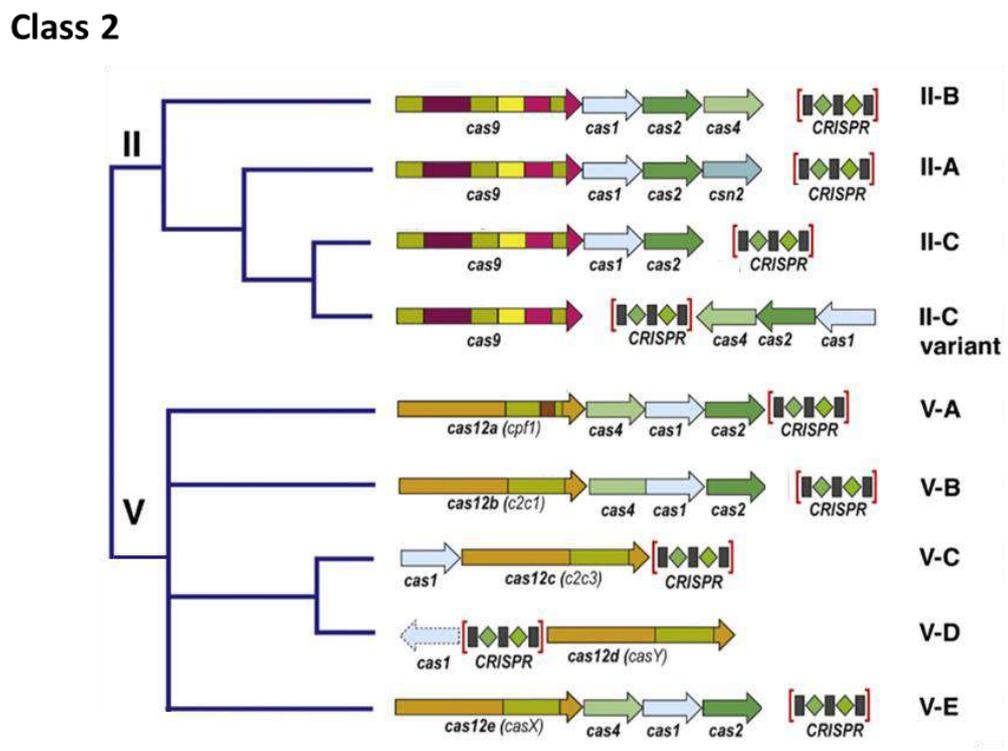
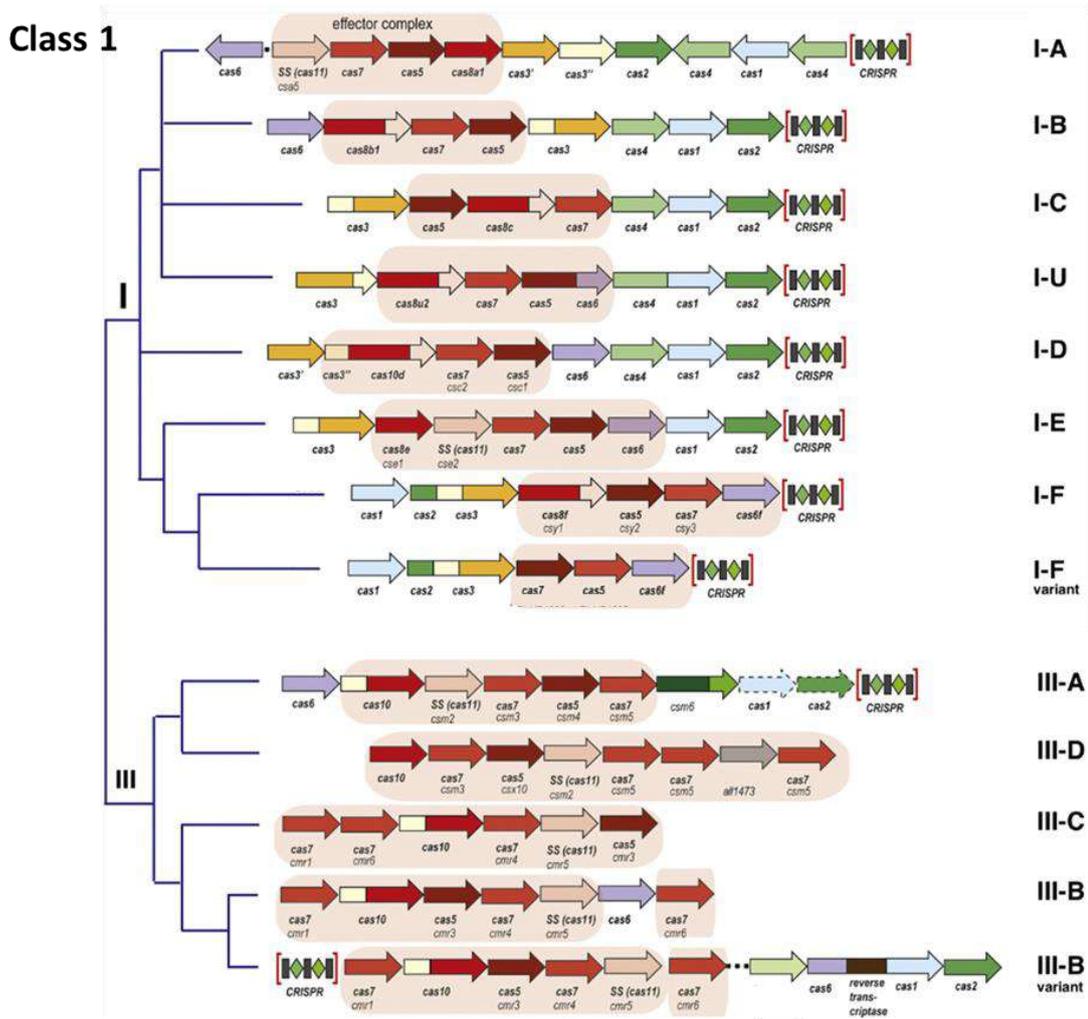


Figure 1.5. Classification of main CRISPR-Cas systems subtypes and the architecture of their loci. There are two classes of CRISPR-Cas systems. Class 1 includes systems with multi-subunit effector modules, Class 2 – with one-protein effector modules. Each Class contains several types named with Roman numerals. These types are divided into subtypes named with Latin alphabet letters. Typical *cas*-operon organizations are shown with colored arrows. Genes encoding components of the effector complexes are highlighted with a light-red background. Adapted from (Koonin et al., 2017) with permission.

Type I CRISPR-Cas systems are the most diversified systems that are classified in 7 subtypes (I-A, I-B, I-C, I-U, I-D, I-E, I-F) (Makarova et al., 2015) (Figure 1.5). Interestingly, subtypes I-C, I-D, I-E, I-F are encoded by a single operon in CRISPR loci, whereas subtype I-A and I-B are often encoded by several operons. At the same time, I-C, I-E and I-F subtypes are largely present in bacterial species, while I-A, I-B and I-D are mostly present in archaea (Makarova et al., 2011b). Subtype I-B, characterized by the presence of a specific Cas8b protein, has been discovered in methanogenic and halophilic archaea and in clostridia species. The studies of I-B CRISPR-Cas systems in haloarchaea revealed several highly interesting features, such as the presence of multiple protospacer adjacent motifs (PAMs) and including a 9-nucleotide noncontiguous seed region (for the detailed description of PAMs and seed regions see 1.2.3 and 1.2.4 parts of this Chapter) (Maier et al., 2015). Subtype I-B has been found in clostridia species, but it has not been studied thoroughly yet. It is suggested that I-B CRISPR-Cas system could have been acquired by clostridia from archaea by a horizontal gene transfer (HGT) and after that underwent an independent evolution (Peng et al., 2014). Other CRISPR-Cas systems subtypes, including I-A, I-C, III-A, III-B and II-C, are also present in some clostridial species (Table 1.1).

Table 1.1. Main subtypes of CRISPR-Cas systems and examples of system-harboring microorganisms and clostridial species. CRISPR-Cas systems subtypes and the composition of *cas* operons are shown according to the classification by (Koonin et al., 2017). Fused *cas* genes in operons are marked with a dash.

Adapted from (Maikova et al., 2018b) with permission.

Class	Subtype	<i>cas</i> operon composition	Example	Examples of clostridial species and strains
Class 1	I-A	<i>cas6, cas11(csa5), cas7, cas5, cas8a1, cas3', cas3'', cas2, cas4, cas1, cas4</i>	<i>Listeria monocytogenes</i> L99 (Sesto et al., 2014)	<i>C. stercorarium</i> subsp. <i>stercorarium</i> DSM 8532 (Pohlein et al., 2013); <i>C. tetani</i> ATCC 9441 (Cohen et al., 2017)
	I-B	<i>cas6, cas8b1, cas7, cas5, cas3, cas4, cas1, cas2</i>	<i>Haloferax volcanii</i> H119 (Maier et al., 2013)	<i>C. difficile</i> 630, <i>C. difficile</i> R20291 (Boudry et al., 2015); <i>C. pasteurianum</i> BC1 (Pyne et al., 2016); <i>C. acetobutylicum</i> GXAS18-1 (Peng et al., 2014); <i>C. tetani</i> ATCC 9441 (Cohen et al., 2017)
	I-C	<i>cas3, cas5, cas8c, cas7, cas4, cas1, cas2</i>	<i>Desulfovibrio vulgaris</i> str. <i>Hildenboroug</i> (Hochstrasser et al., 2016)	<i>C. cellulolyticum</i> H10 (Brown et al., 2014)
	I-U	<i>cas3, cas8u2, cas7, cas5-cas6, cas4-cas1, cas2</i>	<i>Geobacter sulfurreducens</i> (Koonin et al., 2017)	–
	I-D	<i>cas3', cas3'', cas10d, cas7(csc2), cas5(csc1), cas6, cas4, cas1, cas2</i>	<i>Cyanothece</i> sp. 8802 (Koonin et al., 2017)	–
	I-E	<i>cas3, cas8e(cse1), cas11(cse2), cas7, cas5, cas6, cas1, cas2</i>	<i>Escherichia coli</i> K12 (Koonin et al., 2017)	–
	I-F	<i>cas1, cas2-cas3, cas8f(csy1), cas5(csy2), cas7(csy3), cas6f</i>	<i>Pseudomonas aeruginosa</i> PA14 (Wiedenheft et al., 2011)	–
	III-A	<i>cas6, cas10, cas11(csm2), cas7(csm3), cas5(csm4), cas7(csm5), csm6, cas1, cas2</i>	<i>Staphylococcus epidermidis</i> (Koonin et al., 2017)	<i>C. tetani</i> ATCC 453 (Cohen et al., 2017)
	III-B	<i>cas7(cmr1), cas10, cas5(cmr3), cas7(cmr4), cas11(cmr5), cas6, cas7(cmr6)</i>	<i>Pyrococcus furiosus</i> (Koonin et al., 2017)	<i>C. botulinum</i> ATCC 3502 (Negahdaripour et al., 2017)
	III-C	<i>cas7(cmr1), cas7(cmr6), cas10, cas7(cmr4), cas11(cmr5), cas5(cmr3)</i>	<i>Methanothermobacter thermautotrophicus</i> (Koonin et al., 2017)	–
III-D	<i>cas10, cas7(csm3), cas5(csx10), cas11(csm2), cas7(csm7), cas7(csm5), all1473, cas7(csm5)</i>	<i>Synechocystis</i> sp. 6803 (Makarova et al., 2015)	–	
Class 2	II-A	<i>cas9, cas1, cas2, csn2</i>	<i>Enterococcus faecalis</i> OG1RF (Bourgogne et al., 2008)	–
	II-B	<i>cas9, cas1, cas2, cas4</i>	<i>Legionella pneumophila</i> str. <i>Paris</i> (Koonin et al., 2017)	–
	II-C	<i>cas9, cas1, cas2</i>	<i>Neisseria lactamica</i> 020-06 (Koonin et al., 2017)	<i>C. perfringens</i> JGS1495 (Pearson et al., 2015)
	V-A	<i>cas12a(cpfl), cas4, cas1, cas2</i>	<i>Francisella</i> cf. <i>novicida</i> Fx1 (Koonin et al., 2017)	–
	V-B	<i>cas12b(c2c1), cas4, cas1, cas2</i>	<i>Alicyclobacillus acidoterrestris</i> (Koonin et al., 2017)	–
	V-C	<i>cas1, cas12c(c2c3)</i>	<i>Oleiphilus</i> sp. (Koonin et al., 2017)	–
	V-D	<i>cas1, cas12d(casY)</i>	<i>Bacterium</i> CG09_39_24 (Koonin et al., 2017)	–
	V-E	<i>cas12e(casX), cas4, cas1, cas2</i>	<i>Deltaproteobacteria bacterium</i> (Koonin et al., 2017)	–

1.2.3 General description of CRISPR-Cas systems defense mechanisms

CRISPR-based defensive functions include two major processes: immunity (interference) and immunization (adaptation) (Marraffini, 2015). CRISPR interference itself can be divided into two phases: biogenesis of CRISPR RNAs and the targeting phase. During the first phase a CRISPR array is transcribed to a long RNA transcript (pre-crRNA). Then the pre-crRNA is processed into small CRISPR RNAs (crRNAs), each of which consists of one spacer and flanking repeat sequences. Mature crRNAs form a nucleoprotein complex with Cas proteins (crRNP complex, effector complex), which is necessary for the second targeting phase (Figure 1.6). crRNAs serve as guides for recognizing foreign nucleic acids using the complementary base pairing. Through this process, crRNAs direct the cleavage of genetic invaders by Cas nucleases during the process known as “interference” (Figures 1.6 and 1.7) (Garneau et al., 2010). The majority of spacers are incorporated into CRISPR arrays in the process of adaptation. During CRISPR adaptation new spacers are generated and acquired from foreign genetic elements by the complex of Cas proteins (Marraffini, 2015). Cas1 and Cas2 proteins, found in almost all investigated CRISPR-Cas systems, are essential for this process (Figures 1.5, 1.6 and 1.8) (Koonin et al., 2017). Cas4 protein is present in some CRISPR-Cas systems (Koonin et al., 2017) and it is an important part of the adaptation complex (Amitai and Sorek, 2016; Kieper et al., 2018; Lee et al., 2018, 2019; Zhang et al., 2019). A crucial aspect of the CRISPR-based immunity is the ability to distinguish the host DNA from the foreign one. Protospacer-adjacent motifs (PAMs) are short sequences situated at the 3' or 5'-end of the protospacer (i.e., the region of foreign DNA corresponding to the spacer in the CRISPR array). PAMs are essential for interference and adaptation in the majority of CRISPR-Cas systems; at the same time, they are absent in CRISPR arrays. Therefore PAMs prevent the autoimmunity, avoiding self-targeting of CRISPR-array in the most cases (Sorek et al., 2013). Nevertheless, the autoimmunity cases have been reported for all CRISPR-Cas systems types (Heussler and O'Toole, 2016).

In the next parts of this Chapter the processes of interference and adaption in CRISPR-Cas type I systems will be reviewed in detail.

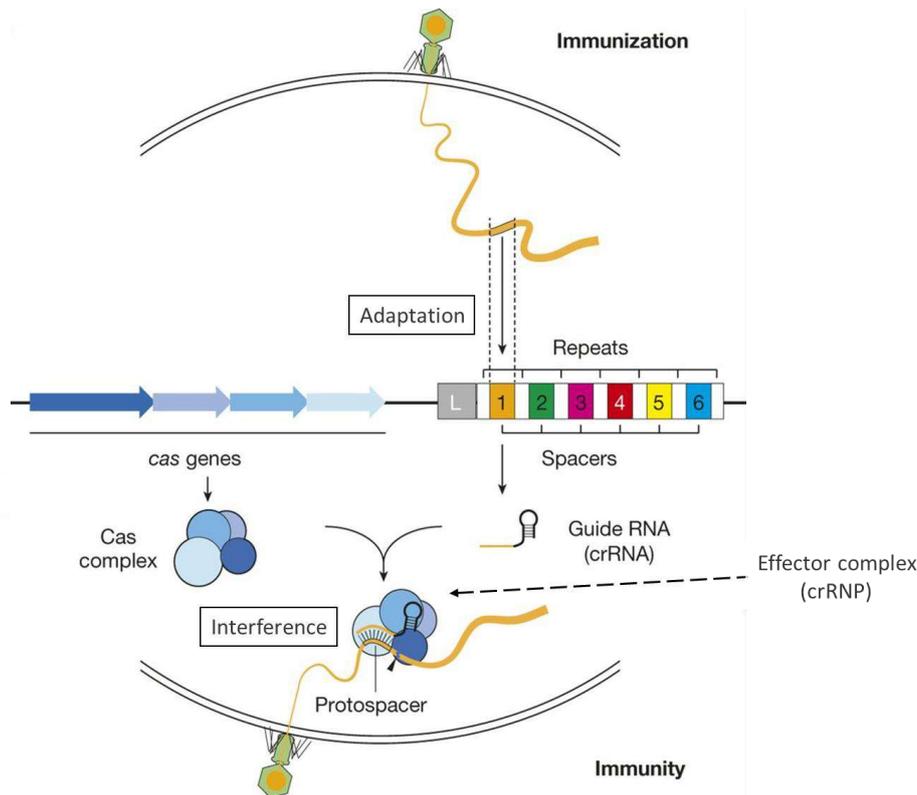


Figure 1.6. CRISPR-based defensive functions: immunity (interference) and immunization (adaptation). During the immunity stage CRISPR-Cas systems recognize foreign genetic elements and destruct them. In the first “crRNA biogenesis” phase, mature small crRNAs are generated from the long pre-crRNA transcripts. These small crRNAs form effector complexes with Cas nucleases and guide them to locate and cleave the target (protospacer), during the second, “targeting phase”. In the immunization stage, new spacer sequences are captured from the foreign DNA and then integrated into the first position of the CRISPR array. Spacers in the CRISPR array are indicated with numbers and colors, “L” designates a leader sequence of the array.

Adapted from (Marraffini, 2015) with permission.

1.2.4 Interference in type I CRISPR-Cas systems

In type I CRISPR-Cas systems, a multi-protein effector complex forms to recognize and cleave the target. The CRISPR interference begins with the transcription of the CRISPR array into a long pre-crRNA transcript. Subsequently, pre-crRNA is processed by Cas6 endoribonuclease to produce mature small crRNAs (Figure 1.7A) (Niewoehner et al., 2014). The exception is the I-C subtype systems, where Cas5 protein performs this function (Nam et al., 2012). In type I Cas6 is a part of the effector complex, called Cascade (CRISPR-associated complex for antiviral defense). A single mature crRNA contains a single spacer flanked by a repeat (Brouns et al., 2008). After the processing,

crRNAs are loaded onto the Cascade complex, composed mainly of Cas5 and Cas7 subunits, and a large Cas8 subunit (Jore et al., 2011). Then, complete effector complexes use crRNAs as guides to locate a complementary sequence (protospacer) in the target DNA (Figure 1.7A) (Wiedenheft et al., 2011). Cas8 protein (also known as CasA or Cse1) recognizes a PAM sequence, and this recognition promotes binding of the Cascade to the protospacer (Westra et al., 2012). During this process, a DNA-RNA hybrid structure, called R-loop, is formed between crRNA and the target double-stranded DNA (Figure 1.7A, B) (Rutkauskas et al., 2015). The first 8-10 nucleotides of the 5'-end of the crRNA ("seed region") are important for CRISPR interference, and mutations in this region can block the immunity (Semenova et al., 2011). After the R-loop formation, the effector complex recruits the Cas3 nuclease, which introduces single-strand breaks into virus or plasmid DNA, triggering its degradation (Brouns et al., 2008).

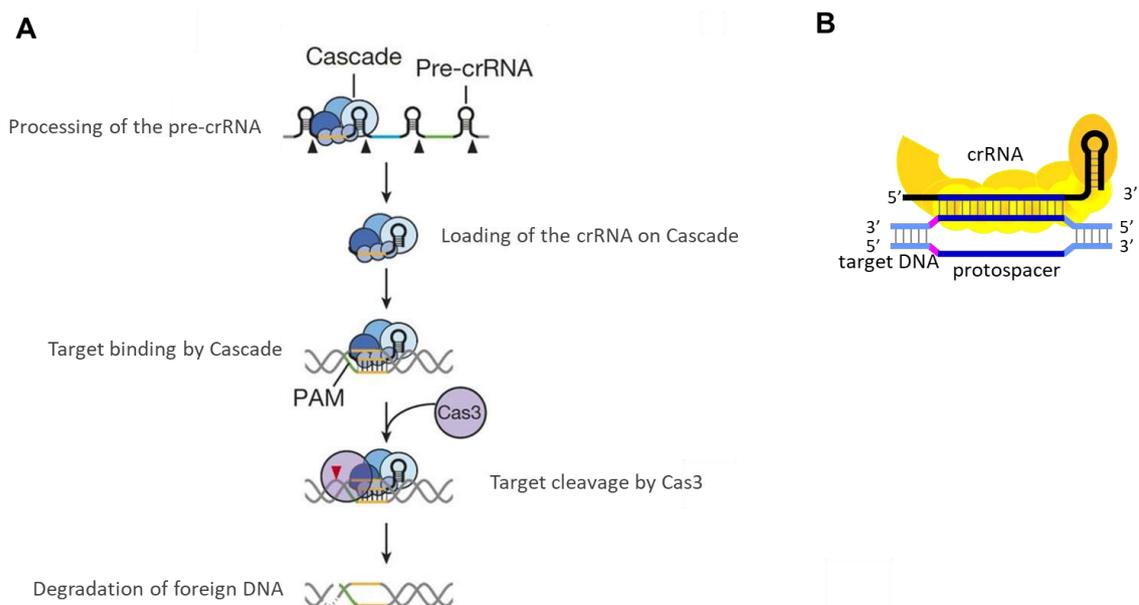


Figure 1.7. Type I CRISPR-Cas systems molecular mechanism of interference. **A** – Cas6 protein within the Cascade complex cleaves at the base of the stem-loop structure of each repeat in the pre-crRNA (black arrowheads), generating short crRNAs. Mature crRNA and Cascade forms the complete effector complex, which scans the target DNA for a matching sequence (protospacer). The protospacer should be flanked by a protospacer adjacent motif (PAM, in green). After the target binding, an R-loop forms between the crRNA and foreign DNA. Subsequently, the Cas3 nuclease is recruited and cleaves the target downstream of the PAM (red arrowhead) and also degrades the opposite strand. **B** – the structure of an R-loop. Adapted from (Marraffini, 2015) with permission.

1.2.5 Adaptation in type I CRISPR-Cas systems

There are two different mechanisms of adaptation that can occur in type I CRISPR-Cas systems: naïve adaptation and primed adaptation. During the naïve adaptation, new spacers are acquired from bacteriophages and other MGE for the first time, and the CRISPR array did not contain spacers against this genetic invader before (Yosef et al., 2012). Primed adaptation occurs when the CRISPR array already possesses a spacer against the invading agent, but the corresponding protospacer or/and PAM is mutated, hindering the recognition of the target by the effector complex (Datsenko et al., 2012).

The process of adaptation includes two steps: production of prespacers, small DNA fragments, which will be processed to spacers; and integration of new spacers into the CRISPR array (Figure 1.8A). Cas1 and Cas2 proteins play a crucial role at each step of the adaptation. These proteins alone are efficient for naïve adaptation in the I-E subtype CRISPR-Cas systems (Yosef et al., 2012). Nevertheless, primed adaptation requires the presence of the interference machinery components, such as Cas3 and the effector complex (Datsenko et al., 2012). Cas4 exonuclease is present in some type I CRISPR-Cas systems, and it contains a RecB-like domain (Zhang et al., 2012). This protein forms a complex with Cas1 and Cas 2 proteins and has an important role in recognizing the PAM sequences and in determining the spacer length and its orientation (Amitai and Sorek, 2016; Kieper et al., 2018; Lee et al., 2018, 2019; Zhang et al., 2019).

To date, only one way of prespacer generation has been identified for naïve adaptation in *E. coli* I-E subtype system (Levy et al., 2015). During DNA replication, a significant number of double-strand DNA breaks occur. An RecBCD exonuclease complex (a DNA repair complex) recognizes ends of these breaks and degrades the DNA till a specific Chi site (an 8 bp sequence motif) (Dillingham and Kowalczykowski, 2008), and Cas1-Cas2 complex may use the degradation products as a source for new spacers (Figure 1.8B). The high density of Chi sites on the bacterial chromosome protects it from the spacer acquisition and further autoimmunity (Levy et al., 2015). Moreover, other ways of prespacers generation may be related to other cellular machinery. For example, fragments produced by restriction-modification systems could be used as a source of prespacers during the CRISPR adaptation process (Dupuis et al., 2013).

Primed adaptation is based on imperfect target binding, and it requires both interference and adaptation protein complexes (Figure 1.8C) (Vorontsova et al., 2015). The mechanism of this spacer acquisition type is still not clear. The partial pairing

between the crRNA and the target initiates primed adaptation, resulting in an alternative mode of Cascade binding to the target (Blosser et al., 2015). It is suggested, that this mode of binding leads to non-sufficient target cleavage, similar to the DNA breaks, occurring at the replication origins (Levy et al., 2015), and it may produce substrates for spacer acquisition.

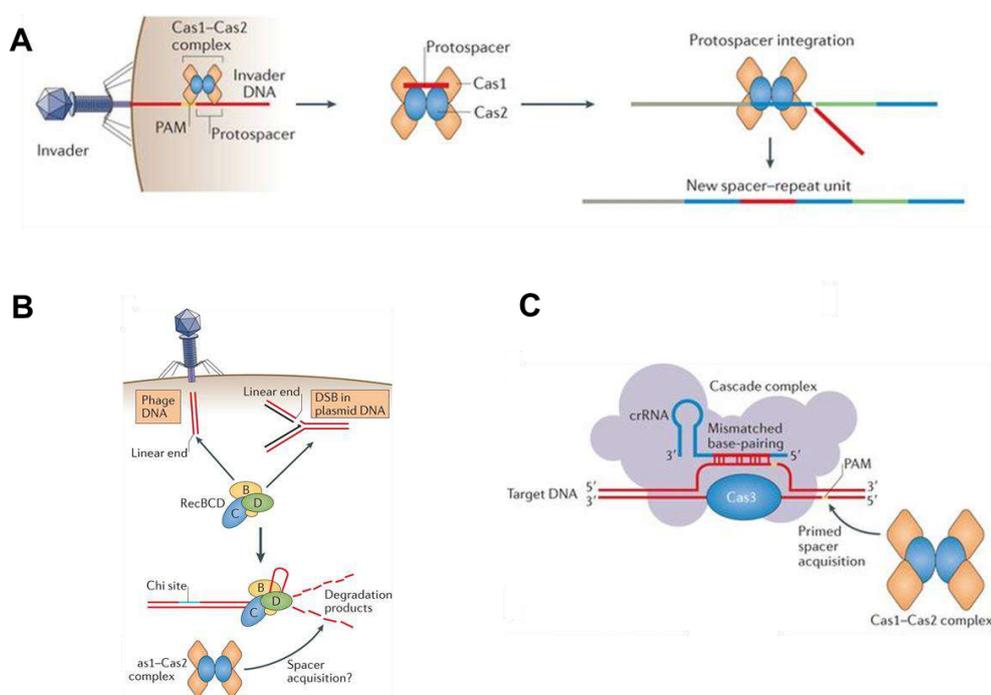


Figure 1.8. CRISPR adaptation in type I CRISPR-Cas systems. **A** – general scheme of the adaptation process. The Cas1–Cas2 complex, consisting of two Cas1 dimers and a single Cas2 dimer, acquires a protospacer from the invader DNA and integrates it as a new spacer into the CRISPR array. The integration is accompanied by a duplication of the first repeat. **B** – the source material for new spacers is suggested to be derived from the processing of linear double-stranded DNA ends, which are found in phage DNA or are formed following a double-strand breaks (DSB). The RecBCD nuclease complex processes these ends, creating single-stranded DNA intermediates. DNA processing by RecBCD proceeds until the complex reaches a specific Chi site. **C** – imperfect matching between crRNA and the target DNA initiates primed adaptation in type I CRISPR-Cas systems.

Adapted from (Amitai and Sorek, 2016) with permission.

The next step of the adaptation process is the spacer integration into the CRISPR array. The mechanism is similar to retroviral integration and it has been well studied for *E. coli* I-E CRISPR-Cas system (Marraffini, 2015; Nuñez et al., 2015a, 2016). First, the integration host factor (IHF) binds the leader sequence and promotes a sharp DNA bend

at its binding site (Figure 1.9). This DNA bent allows the Cas1-Cas2 complex, carrying the protospacer, to recognize a leader-repeat bond in the CRISPR-array (Nuñez et al., 2016). Subsequently, Cas2 produces two 3'OH ends on the protospacer, that is necessary for the nucleophilic attack on each strand of the leader-proximal repeat (Figure 1.9) (Nuñez et al., 2015a). Then two nucleophilic attacks create the full site integration product. When this product is generated, uncharacterized enzymes fill the gaps, introducing new repeats (Amitai and Sorek, 2016).

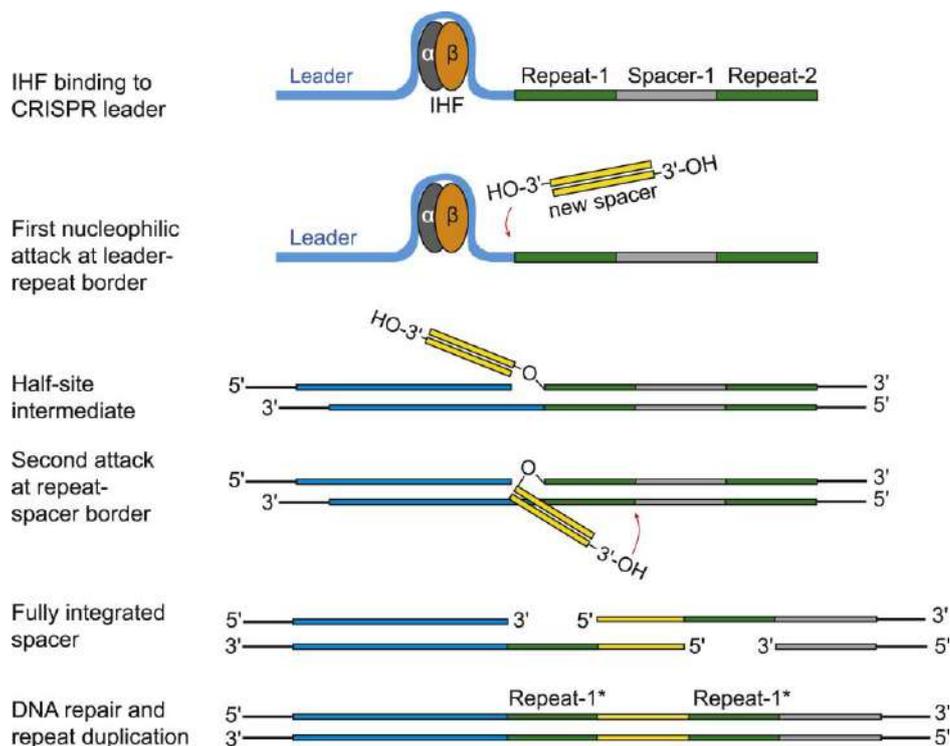


Figure 1.9. A model for protospacer integration into the CRISPR array. IHF (gray and orange) binds to the specific site of the leader sequence, inducing a sharp DNA bend. This provides the access to the leader-repeat border for the Cas1-Cas2 complex with loaded protospacer. The protospacer 3'-OH group performs a nucleophilic attack on the 5' end of the first repeat. This initiates spacer acquisition by forming a half-site intermediate, in which a single strand of the protospacer is connected to a single strand of the CRISPR array. The 3'-OH group of another protospacer strand carries out the second nucleophilic attack on the 5' end of the opposite DNA strand of the repeat, adjacent to the leader sequence. This reaction results in an expanded CRISPR array with a new spacer and a duplicated repeat. Then intracellular enzymes fill and repair the single-strand DNA gaps, formed by these reactions. Adapted from (Nuñez et al., 2016) with permission.

1.3 *C. difficile* CRISPR-Cas system

1.3.1 Characterization of *C. difficile* CRISPR-Cas system

Clostridium difficile is an anaerobic spore-forming bacterium, one of the major clostridial pathogens and the main causes for nosocomial infections associated with antibiotic therapy (Abt et al., 2016). During its infection cycle this enteropathogen has to cope with foreign DNA invaders, including bacteriophages; to do so it relies on efficient defense systems such as CRISPR-Cas to control genetic exchanges favored within these complex environments.

The first evidence suggesting the presence of active CRISPR-Cas system in *C. difficile* was provided by a deep-sequencing analysis of regulatory RNAs in *C. difficile* (Soutourina et al., 2013). This study revealed the abundance of crRNAs in the pathogen. Moreover, the active expression of crRNAs, as well as processing of CRISPR loci, were further confirmed by RNA-seq and Northern blotting (Soutourina et al., 2013; Boudry et al., 2015). These studies and bioinformatic analysis of up to 217 *C. difficile* genomes (Hargreaves et al., 2014; Andersen et al., 2016) demonstrated that *C. difficile* CRISPR-Cas system belongs to the I-B subtype of the system according to the recent classification (Figure 1.5, Table 1.1) (Koonin et al., 2017).

As compared to other class I CRISPR-Cas systems, *C. difficile* CRISPR-Cas system has several original and even unique features. First of all, CRISPR-Cas system of this enteropathogen is characterized by an unusually large set of actively expressed arrays. Genome sequencing and an RNA-seq analysis of the reference 630 strain and hypervirulent R20291 *C. difficile* strain identified 12 and 9 CRISPR arrays, respectively (Figure 1.10) (Soutourina et al., 2013; Boudry et al., 2015). First, *in silico* analysis of nine available *C. difficile* genomes revealed the presence of 6-12 CRISPR arrays (Boudry et al., 2015). Later, a more detailed bioinformatics analysis of 217 *C. difficile* genomes identified 8.5 CRISPR arrays per genome on average (Andersen et al., 2016). Interestingly, likewise it is observed for other highly expressed bacterial genes, in a well-studied reference 630 strain and hypervirulent R20291 strain CRISPR arrays are oriented towards chromosome replication to ensure their optimal transcription (Arakawa and Tomita, 2007; Boudry et al., 2015) (Figure 1.10). Furthermore, in a large amount of sequenced *C. difficile* strains several CRISPR arrays are located in prophages (Hargreaves et al., 2014; Boudry et al., 2015). Prophage-related crRNAs appeared to be the most expressed in 630 and R20291 strains. This prophage localization of the actively expressed

CRISPR arrays can play an important role in preventing infections caused by other phages and may act as an essential element in the horizontal transfer of CRISPR arrays between strains.

Another specific feature of *C. difficile* CRISPR-Cas system is the presence of two or three (in 027 ribotype strains) *cas* gene sets in the majority of sequenced strains (Figure 1.10) (Boudry et al., 2015). The full *cas* operon encodes all necessary genes for CRISPR interference (*cas6*, *cas8b*, *cas7*, *cas5*, *cas3*), as well as *cas1*, *cas2*, *cas4* genes, which form an adaptive module essential for a new spacer acquisition (Amitai and Sorek, 2016; Kieper et al., 2018; Lee et al., 2018, 2019; Zhang et al., 2019). The additional *cas* operons lack the adaptation module of *cas1*, *cas2* and *cas4* genes. Interestingly, the full *cas* gene sets were found in 90% of sequenced *C. difficile* strains, whereas the additional partial *cas* gene sets are conserved in almost all strains (Boudry et al., 2015). This fact can indicate that some *C. difficile* strains may lose their ability to use their CRISPR-Cas system to adapt to new genetic invaders. The full *cas* gene sets are often associated with the longest CRISPR arrays suggesting an active acquisition of a new spacer in these arrays. The conserved structure of CRISPR array and sequences of all CRISPR repeats in *C. difficile* genome suggests that other CRISPR arrays located *in trans* to *cas* operons could use the same set of Cas proteins for their functioning. Notably, the occurrence of the *cas* operon turned out to correlate with evolutionary relationships of *C. difficile* strains reflecting their epidemiological context and the intensity of interaction with foreign DNA invaders (Boudry et al., 2015).

The CRISPR-Cas system, functioning as a bacterial immune system, aims to provide defense against viruses and other MGE. The recent bioinformatic analysis of *C. difficile* CRISPR spacer targeting in different strains has shown that most spacers target *Clostridium* phages and prophage regions within the chromosome (Hargreaves et al., 2014; Boudry et al., 2015). It proposes the idea that this enteropathogen actively interacts with phages, so that functional CRISPR-Cas appears to be necessary for *C. difficile* survival in the phage-rich environment. PAM sequences were also determined in the course of the studies. Using spacer homology analysis and the method of alignment of the regions adjacent to the protospacer, PAMs were detected to take place mostly as 3-nucleotide 5'-motifs CCA or CCT, though alternative sequences CCC, CCG and TCA also appeared to be found frequently. This situation with multiple PAMs was also observed in other type I-B systems (Shah et al., 2013). Moreover, experimental evidence for the defensive function of *C. difficile* CRISPR-Cas was provided by (Boudry et al.,

2015). Conjugation efficiency experiments with plasmid vectors containing CCA and CCT PAMs and protospacers corresponding to the first spacers from actively expressed *C. difficile* 630 CRISPR arrays showed active CRISPR interference to *C. difficile* cells. Phage infection assays in 630 and R20291 strains revealed the correlation between the presence of CRISPR spacer-targeting phage sequences and the corresponding phage susceptibility phenotype. Moreover, experiments in the heterologous *E. coli* system detected a defensive function typical of both *cas* operons of *C. difficile* 630 strain and confirmed the functionality of CRISPR interference in this strain.

C. difficile CRISPR-Cas system actively provides the defense against genetic invaders such as prophages, which are regarded as important elements providing adaptation inside bacteriophage-rich gut community of the host. *C. difficile* is characterized by a highly mobile and mosaic genome (Sebahia et al., 2006). The unique properties of this CRISPR-Cas system reflect the evolutionary balance between the acquisition of new genetic advances through a HGT and efficient defense against foreign genetic elements.

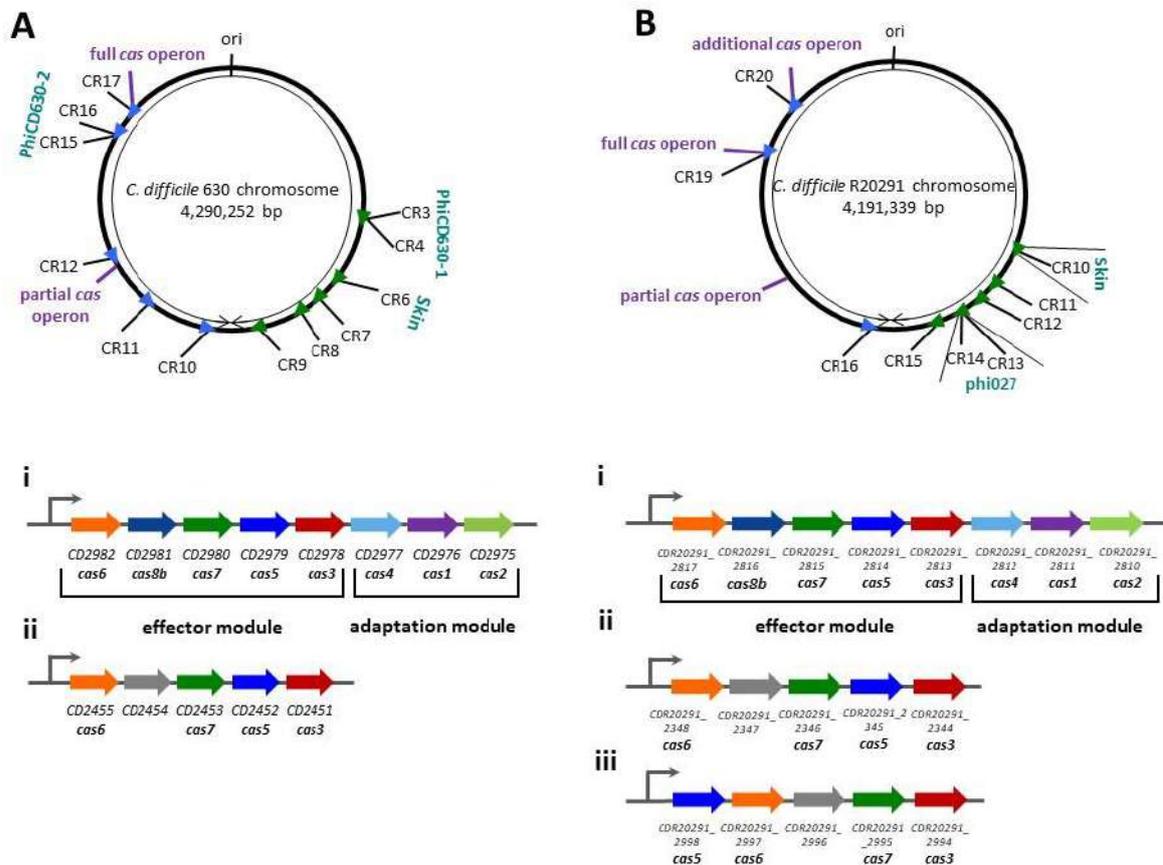


Figure 1.10. Schematic view of the chromosomal location of CRISPR arrays and the organization of *cas* operons in *C. difficile* strains 630 (A) and R20291 (B). CRISPR arrays (CR) are numbered according to the CRISPRdb database (Grissa et al., 2007). Arrow heads signify arrays' position and transcriptional orientation. The locations of associated *cas* operons, prophage regions, and replication origin (ori) are indicated. The organization of the *cas* operons in strain 630 (left) and R20291 (right) are indicated with Roman numerals, where “i” stands for full operons; “ii” – partial operons, “iii” – an additional operon. Functional modules are marked off with braces. The same color was used for homologous *cas* genes (Boudry et al., 2015). Adapted from (Maikova et al., 2018b) with permission.

1.3.2 Regulation and potential alternative functions of *C. difficile* CRISPR-Cas system

During its infection cycle *C. difficile* faces different stress conditions and changing environments inside the host. The phage-rich gut community implies an active interaction with phages and other mobile genetic elements. To survive under such changing and stressful conditions *C. difficile* most possibly relies on the CRISPR-Cas system, which should be regulated in response to different environmental signals. The recent study (Boudry et al., 2015) revealed that all CRISPR arrays and *cas* genes are constantly

expressed under standard laboratory conditions, which may indicate, that *C. difficile* cells are under a continuous pressure of phage infection and other stress factors.

Bacterial pathogens often form biofilms, which help them to resist different threats inside the host. It was shown that during its infection cycle *C. difficile* actively multiplies and forms biofilms (Dapa et al., 2013; Nale et al., 2016; Soavelomandroso et al., 2017). Biofilm conditions are characterized by a high cell density, which increases the possibility of getting phage infection and being subjected to a HGT (Babic et al., 2011; Abedon, 2012). Bacteria have different mechanisms to react to the changes in population density. Quorum sensing is a special chemical signal system for ensuring the communication between bacterial cells; it regulates gene expression depending on the density of population (Miller and Bassler, 2001). Recent studies in the field have shown that *cas* gene expression is induced by quorum sensing signals in *Serratia sp.* (I-E, I-F and III-A subtypes) (Patterson et al., 2016) and in *Pseudomonas aeruginosa* (I-F subtype) (Høyland-Kroghsbo et al., 2016). Moreover, CRISPR-Cas systems may have a significant role in biofilm formation and colonization of the host. For instance, CRISPR-Cas (II-A subtype) harboring *Enterococcus faecalis* strain has shown increased biofilm formation (Bourgogne et al., 2008). Furthermore, CRISPR-Cas-mediated gene regulation of the ability to swarm and form biofilms was revealed in *P. aeruginosa* (Zegans et al., 2009).

Bacteria often have to cope with different stressful conditions of the environment, which may lead to the changes in the expression and functioning of CRISPR-Cas system. One of the most stressful conditions is infecting CRISPR-Cas system by phages. It is often accompanied by the envelope stress response occurring when phages attach to the cell surface (Ratner et al., 2015). Inducing the CRISPR-Cas system expression in response to this type of stress has been found in different bacterial species (Westra et al., 2014). Bacterial pathogens and commensals always combat with the host's immune response, which results in a wide range of stress effects. Several studies report on the changes of *cas* gene transcription in *Desulfovibrio vulgaris* (Mukhopadhyay et al., 2007), *Streptococcus sanguinis* (Rodriguez et al., 2011), *Pasteurella multocida* (Melnikow et al., 2008) and *Lactobacillus rhamnosis* (Koskenniemi et al., 2011) occurring in response to different kinds of stresses, such as changes in growth rate, bile stress, oxidative stress, nitrosative stress and exposure to antibiotics. Virulence is a specific response of pathogens to different stress factors inside the host (Louwen et al., 2014) and the detectable regulation of CRISPR-Cas systems may indicate an important role of CRISPR-Cas systems in the infection cycle. Recently, the role of the alternative Sigma B factor in

response to different stresses has been investigated in *C. difficile* (Kint et al., 2017). Interestingly, Sigma-B-dependent promoters were found upstream the full and partial *cas* operons in *C. difficile* strain 630 (Maikova et al., 2018a). This presupposes the regulation of *C. difficile* CRISPR-Cas system through stress-related signals and a potential role of this system in the adaptation to changing environments inside the host.

Apart from the adaptive immunity, many works revealed other alternative functions of CRISPR-Cas systems (Louwen et al., 2014; Westra et al., 2014). The putative gene regulation of bacterial genes by partial or full CRISPR targeting could be one of them. For instance, *Listeria monocytogenes* (I-A type) possesses a specific long CRISPR-array transcript *rliB*, which cannot be processed in mature crRNA by Cas proteins and controls the expression of *feoAB* genes, which are important for virulence (Mandin et al., 2007). This study also revealed that *rliB* mutant colonizes its host more effectively than a wild type strain. Bioinformatic analysis of *C. difficile* CRISPR spacers has shown that all investigated strains include genome-targeting spacers (Boudry et al., 2015). It may be suggested that *C. difficile* CRISPR-Cas system might take place in regulating the endogenous gene expression. Furthermore, a role of CRISPR-Cas systems in the genome evolution via self-targeting is actively discussed nowadays (Westra et al., 2014). Although, self-targeting is generally cytotoxic, bacterial cells avoid this negative effect by acquiring mutations in CRISPR-Cas system components, thus inactivating self-targeting. In some cases, these mutations can involve large-scale genome rearrangements presumably by the recombinational DNA repair following CRISPR-Cas-mediated genome cleavage (Vercoe et al., 2013). CRISPR-Cas system components encoded by prophages and other mobile genetic elements can participate in the competition between them (Minot et al., 2013). Most likely, *C. difficile* prophage-located CRISPR-arrays perform the same function (Boudry et al., 2015).

Thus, *C. difficile* possesses an unusual, vividly expressed CRISPR-Cas system that may indicate some alternative functions. Further investigations on the subject are needed to shed light on these aspects.

1.3.3 Potential applications of *C. difficile* CRISPR-Cas system

During the last decade the major discoveries concerning CRISPR machinery have led to rapid development of revolutionary biotechnological applications, especially in genome editing with the help of CRISPR-Cas9 technology (Hsu et al., 2014). Different

CRISPR-based tools have proved to be effective both in prokaryotes and eukaryotes (Hsu et al., 2014; Barrangou and Horvath, 2017).

In CRISPR loci, spacers are acquired according to the time course, i.e., the newer the spacer is the closer it is located to the leader sequence in the array (Jackson et al., 2017). This time-depending order of spacers reflects phage invasions to different populations of the same bacterial species that can reveal phylogenetic relations between the strains. Therefore, this interesting feature of CRISPR arrays allows us to use them in genotyping techniques (Louwen et al., 2014; Andersen et al., 2016). CRISPR-typing has been already applied to the outbreak tracking of *Yersinia pestis* (Cui et al., 2008; Barros et al., 2014) and *Salmonella enterica* (Timme et al., 2013; Pettengill et al., 2014). Moreover, CRISPR-Cas typing is able to reveal antibiotic-resistant phenotypes (Palmer and Gilmore, 2010) or prophages (Nozawa et al., 2011) associated with certain CRISPR-Cas system components. These correlations can account for the influence of active CRISPR-Cas systems on HGT, which plays one of the key roles in acquiring new genes and operons, essential for bacterial pathogenesis and adaptation (Louwen et al., 2014). The study into CRISPR-Cas diversity in *C. difficile* has shown that CRISPR-typing approach can be successfully applied to this enteropathogen (Andersen et al., 2016). Since *C. difficile* CRISPR-Cas system possesses a great diversity of arrays and spacers, widely conserved and unique variable arrays were found in almost all the analyzed strains; they can be applied to PCR-amplification, sequencing and typing. Interestingly, a correlation between CRISPR-groups and toxin groups has also been reported in this study.

It is worth mentioning that CRISPR-Cas arrays often contain the spacers targeting bacterial chromosome. Such self-targeting spacers can probably be acquired by mistake. Bacterial cells utilize mutations in *cas* genes, PAMs or seed sequences to prevent the autoimmunity (Horvath et al., 2008; Semenova et al., 2011). At the same time, auto-targeting spacers may have alternative functions in the bacterial cells (Westra et al., 2014). Despite this fact, CRISPR-auto-targeting almost always leads to cell death (Gomaa et al., 2014). Moreover, both Class 1 and 2 CRISPR-Cas systems are efficient to cure plasmids carrying virulence and antibiotic resistance genes (Bikard et al., 2012). Thus, these aspects of CRISPR-Cas systems can be applied to devising new antimicrobial agents, which can be done through several strategies. The most general way among them is harnessing the phage particles and phagemids as vectors to deliver all necessary auto-targeting CRISPR-Cas components inside the pathogen's cell (Bikard and Barrangou,

2017). Many pathogens possess the endogenous active CRISPR-Cas systems and their Cas-machineries can be used for self-targeting. Since *C. difficile* contains an active CRISPR-Cas system (Boudry et al., 2015), this approach seems to be promising when applied to CDI treatment, in particular in the context of the recent emergence of antibiotic-resistant *C. difficile* strains all over the world (Banawas, 2018). Phage therapy of CDI has been proved to be a promising alternative, though it faces certain difficulties (Hargreaves and Clokie, 2014), including anti-phage CRISPR-Cas immunity and the current possibility to identify temperate *C. difficile* phages exclusively (Sekulovic et al., 2014). Thus, CRISPR-based antimicrobials using endogenous Cas machinery may be regarded as an alternative to antibiotic treatment of the CDI.

The most recognized biotechnological application of CRISPR-Cas systems is genome editing. Despite the fact that the majority of works are concentrated on using CRISPR-Cas systems for gene engineering in eukaryotes, CRISPR-Cas-based genome editing in prokaryotes has revealed itself as a useful tool (Barrangou and Horvath, 2017). The most interesting application is, perhaps, using endogenous CRISPR-Cas system in genome editing and engineering, since it simplifies the construction of necessary vectors and the process of editing. During the last three years, several works brought to the limelight the use of endogenous I-B subtype systems. One of them (Pyne et al., 2016) describes the application of the approach to *Clostridium pasteurianum*. In this study, a plasmid vector containing an artificial CRISPR array with a protospacer targeting the gene of interest and arms for homologous recombination was used to delete *cpaAIR* gene encoding a restriction enzyme. This approach allows for the fast and markless deletion or modification of the genes of interest in bacteria. Later, other studies confirmed the efficiency of this method when applied it to other I-B subtype-carrying organisms: archaeon *Haloarcula hispanica* (Cheng et al., 2017) and solventogenic clostridia, *Clostridium tyrobutyricum* (Zhang et al., 2018) and *Clostridium saccharoperbutylacetonicum* (Atmadjaja et al., 2019). Furthermore, a recent study revealed that *Haloferax volcanii* CRISPR-Cas system with deletions of *cas3* and *cas6* genes can be used in gene repression in this archaeon (Stachler and Marchfelder, 2016). To date many efficient approaches to *C. difficile* genome manipulation exist and are put into practice. ClosTron is a method based on the alteration of type II intron, which is able to be inserted in almost every region of the chromosome (Kuehne et al., 2011). Another method is CodA allele exchange technique based on using a semisuicidal plasmid vector carrying *E. coli codA* gene as a counter selective marker (Cartman et al., 2012). Recently

the successful application of CRISPR-Cas9 and Cpf1 (Cas12) systems to genome editing in *C. difficile* was reported (Hong et al., 2018; Inés et al., 2019; Ingle et al., 2019; McAllister et al., 2017; Wang et al., 2018). Despite of the efficiency of the described methods, utilizing endogenous CRISPR-Cas system in genome editing in *C. difficile* can enhance the possibilities of genetic manipulations within this enteropathogen.

In conclusion, recent insights into the subject have demonstrated that *C. difficile* CRISPR-Cas system provides not only a large number of opportunities for basic research of its function in the *C. difficile* infection cycle but also opens up a way for various kinds of highly promising medical and biotechnological applications.

Chapter 2. Functionality of *C. difficile* CRISPR-Cas system

The work, presented in this Chapter, is still ongoing. The results will be complemented by further studies.

2.1 Introduction

The defensive function of CRISPR-Cas systems is based on two related mechanisms: interference and adaptation (Marraffini, 2015). During the interference process, maturation of crRNAs, followed by a formation of a crRNP complex with Cas proteins, occurs. Then this complex recognizes and destructs the foreign DNA target. Adaptation mechanism provides new spacer acquisition into existing CRISPR array, thus allowing prokaryotic cells to cope with new genetic invaders.

C. difficile CRISPR-Cas system is characterized by several unusual aspects, such as a high number of CRISPR arrays and multiple sets of *cas* genes belonging to the same subtype (I-B). Previous works (Boudry et al., 2015; Soutourina et al., 2013) demonstrated that individual crRNAs corresponding to different CRISPR arrays are expressed at very different levels, raising the question of differential contribution of various CRISPR arrays to the CRISPR-Cas mediated defense. Moreover, the relative role of multiple interference genes in CRISPR-Cas defense is not known. *C. difficile* CRISPR-Cas system should actively take part in the pathogen's adaptation to the complex community inside the host and can contribute to *C. difficile* infection cycle regulation. Therefore, it is important to investigate the functional aspects of this defensive system.

This Chapter describes the functionality of *C. difficile* CRISPR-Cas in interference and adaptation. Additionally, all possible PAM sequences were identified, and the role of the partial *cas* operon in interference was revealed. These results correspond to the first objective of this Thesis: the investigation of the role and the functionality of *C. difficile* CRISPR-Cas system in the interactions with plasmids.

2.2 Materials and methods

2.2.1 Bacterial strains and growth conditions

All bacterial strains used in this study are listed in Table S2.1 in Supplementary materials. *C. difficile* strains were grown in brain heart infusion (BHI) (Difco) medium at 37°C under anaerobic conditions (5% H₂, 5% CO₂, and 90% N₂), within an anaerobic chamber (Jacomex). When needed, thiamphenicol (Tm) at the final concentration of 15 µg/ml was added to *C. difficile* cultures. *E. coli* strains were grown in LB medium (Bertani, 1951), supplemented with ampicillin (Amp) (100 µg/ml) and chloramphenicol (Cm) (15 µg/ml) when it was necessary. The non-antibiotic analog anhydrotetracycline (ATc) was used for induction of the inducible P_{tet} promoter of pRPF185 vector derivatives in *C. difficile* (Fagan and Fairweather, 2011). Growth curves were obtained using a GloMax plate reader (Promega).

2.2.2 Construction of plasmids and conjugation into *C. difficile*

All plasmids used in this work are presented in Table S2.1 in Supplementary materials. To construct plasmid PAM libraries, single-stranded synthetic oligonucleotides, containing four random nucleotides on the 5'-end, selected protospacer sequence, and regions, overlapping with pRPF185Δ*gus*, were synthesized. Subsequently, single-stranded synthetic oligonucleotides were amplified by PCR to generate the double-stranded fragments. Double-stranded libraries fragments were cloned into SacI and BamHI sites of pRPF185Δ*gus* using Gibson assembly reaction (Gibson et al., 2009).

Synthetic complementary (5'-3' and 3'-5') single-stranded oligonucleotides, containing SacI and BamHI restriction sites, different PAM and protospacer sequences were used to construct PAM-protospacer carrying conjugative plasmid vectors. The single-stranded oligonucleotides were annealed to each other and resulting double-stranded fragments were ligated into SacI and BamHI sites of the pRPF185Δ*gus* vector.

To create plasmids overexpressing Cas proteins, *C. difficile* 630Δ*erm cas1-cas2* and *cas4-cas1-cas2* gene regions including ribosome-binding sites (-21 to +1252 relative to translational start site of *cas2* gene and -37 to +1773 relative to translational start site of *cas4* gene, respectively) were amplified by PCR and introduced into SacI and BamHI sites of pRPF185Δ*gus* under the control of ATc-inducible P_{tet} promoter resulting in pCas1-2 and pCas1-2-4 plasmids.

To construct a plasmid for deletion of the full *cas* operon in *C. difficile* 630 Δ *erm*, approximately 1200-bp long regions flanking *CD2975-2982* genes were amplified by PCR and cloned into PmeI restriction site of pMTL-SC7315 using Gibson assembly reaction giving pDIA6495.

DNA sequencing was conducted to confirm the plasmid construction. All resulting plasmids were transformed into *E. coli* HB101 (RP4) strain. Obtained *E. coli* transformants were subsequently mated with *C. difficile* cells on BHI agar plates for 24 h at 37°C. *C. difficile* transconjugants were selected on BHI agar containing Tm (15 µg/ml), D-cycloserine (Cs) (25 µg/ml) and cefoxitin (Cfx) (8 µg/ml).

2.2.3 PAM libraries high throughput sequencing and data analysis

Experimentally obtained “PAM libraries before the conjugation” and “PAM libraries after the conjugation” (see 2.3.1.1 in Results) were sequenced using Illumina NextSeq500 system with 2 million reads coverage. At the initial step of the sequencing data analysis, all obtained reads were aligned with reference sequences (synthetic oligonucleotides, see 2.2.2 in Methods) using BWA software (Li and Durbin, 2009). Reads mapped to the reference sequence were selected by the SAMtools software (Li et al., 2009). Randomized PAM regions in selected reads were determined by the specially developed Python scripts (version 3.4). After this, PAM sequences were tested for the quality and selected for the further depletion analysis (see 2.3.1.2 in Results). Each PAM nucleotide quality was greater than or equal to 20 (Q20).

Selected PAM sequences were analyzed using the Pearson chi-square test. PAM sequences depleted with a p-value less than 10^{-12} were selected as statistically significant. Their consensus sequence was then visualized using the WebLogo tool (Crooks et al., 2004). For the additional PAM sequences visualization, PAM wheels were constructed using the previously described approach, adapted for the depletion test (Leenay et al., 2016). For PAM wheels construction KronaExcelTemplate software was used (<https://github.com/marbl/Krona/wiki/Downloads>) (Leenay et al., 2016).

2.2.4 Plasmid conjugation efficiency assays

To evaluate conjugation efficiency, PAM-protospacer carrying conjugative plasmids were transformed into the *E. coli* HB101 (RP4) strain and transferred to *C. difficile* 630 Δ *erm* or *C. difficile* R202091 strains by conjugation. The ratio of *C. difficile* transconjugants was counted by subculturing conjugation mixtures on BHI agar

supplemented with Tm, Cs, and Cfx and comparing the number of CFU obtained after plating serial dilutions on BHI agar plates containing Cfx only.

2.2.5 High throughput sequencing and data analysis of newly acquired spacers

Newly acquired spacers detected using CRISPR adaptation assays (see 2.3.3.1 and 2.3.3.1 in Results) were sequenced using the Illumina NextSeq500 system with 2 million reads coverage. Obtained reads were analyzed in R using ShortRead and Biostrings packages (Morgan et al., 2009). Reads were filtered for Phred quality scores of ≥ 20 and spacer sequences were extracted from the reads containing two or more CRISPR repeats (CRISPR repeats were searched with two mismatches allowed). Spacers of 10-79 bp in length were mapped to the reference genome of *Clostridium difficile* 630 (NCBI Reference Sequence: NC_009089.1) and plasmids pCD630 (NCBI Reference Sequence: NC_008226.2) and pCas1-2-4 to identify "protospacer" sequences with one mismatch allowed. Three nucleotides-motif upstream of the first position of a protospacer was considered as a PAM sequence. Graphical representation of the distribution of protospacers along the genome and all the histograms were generated using ggplot2 package (Wickham, 2009) and the EasyVisio tool, developed by E. Rubtsova.

2.2.5 Construction of the *C. difficile* 630 Δ erm Δ CD2975-2982 mutant and its verification by qRT-PCR

To delete full *cas* operon (CD2975-2982) in *C. difficile* 630 Δ erm, the *codA* allele exchange method (Cartman et al., 2012) was used. Editing plasmid pDIA6495 was transferred to *C. difficile* 630 Δ erm via conjugation. Subsequently, selected transconjugants were twice restreaked onto BHI agar plates, containing Tm, Cs, and Cfx to detect faster growing single-crossover integrants. Then selected colonies were plated onto *C. difficile* minimal media agar (CDMM) supplemented with fluorocytosine (50 μ g/ml) to identify second cross-over events. Subsequently, fluorocytosine-resistant clones were analyzed by PCR. The resulting PCR fragments have been sequenced to confirm the gene deletion.

Total RNA was isolated from *C. difficile* strains after 7.5 h of growth in BHI medium as previously described (André et al., 2008). The cDNA synthesis by reverse transcription and quantitative real-time PCR analysis was performed as previously described (Saujet et al., 2011).

2.3 Results

2.3.1 Determining of PAM-sequences in *C. difficile*

2.3.1.1 Experimental model and construction of PAM libraries

To determine all possible PAM sequences in *C. difficile* 630 Δ *erm* and R20291 strains, plasmid PAM libraries depletion assays were performed (Figure. 2.1). Plasmid PAM libraries were constructed using pRPF185 Δ *gus* as a vector backbone and were based on effective for interference CRISPR16 array of 630 Δ *erm* strain and CRISPR13 array of R20291 strain. Synthetic oligonucleotides, containing four random nucleotides (NNNN) on the 5'-end and chosen spacers sequences were cloned into pRPF185 Δ *gus* plasmid vector giving PAM libraries (Figure. 2.1). According to the previous *in silico* analysis, 3-nucleotide 5'-motifs CCA, CCT, CCC, CCG, and TCA were determined as PAMs for *C. difficile* CRISPR-Cas system (Boudry et al., 2015). We have chosen a 4-nucleotide 5'-PAM motif for the libraries depletion assays to verify whether an additional nucleotide at the -4 position relative to the first nucleotide of the protospacer could be important for CRISPR-Cas system functioning. Four randomized nucleotide positions give 256 variants of the PAM. Therefore, a pool of ~ 3000 transformed cells would provide 10-fold library coverage. After the transformation of the plasmid libraries into *E. coli* NEB10 beta cells ~ 7880 clones for 630 Δ *erm* plasmid library and ~ 9400 clones for R20291 library were obtained. This provided more than 10-fold library coverage. Further, all the colonies were used for plasmid PAM-libraries extraction. Then a PCR with the extracted plasmid libraries and primers with Illumina adaptors was performed giving “PAM libraries before the conjugation” (Figure. 2.1).

Plasmid PAM libraries were transformed into *E. coli* HB101 RP4 cells for further conjugation into *C. difficile* cells (approximately $4.9 \cdot 10^{10}$ plasmid copies for the 630 Δ *erm* library and $2.8 \cdot 10^{10}$ plasmid copies for the R20291 library). After the conjugation, ~ 4000 transconjugants and ~ 2000 transconjugants were obtained for 630 Δ *erm* and R20291 strains, respectively. All the transconjugants were then transferred to liquid BHI medium supplemented with antibiotics thiamphenicol, cefoxitin and cycloserin (Tm, Cfx, and Cs) to eliminate remaining *E. coli* cells. Tm was used to maintain plasmids within *C. difficile* cells, while Cfx and Cs were used to remove *E. coli* cells, since they do not possess resistance to these antibiotics. Subsequently, cells from obtained liquid cultures were collected and PCR, using their InstaGene (BioRad) extracts and primers with Illumina adaptors, was performed, giving “PAM libraries after the conjugation” (Figure 2.1).

Obtained “PAM libraries before the conjugation” and “PAM libraries after the conjugation” were used for further high-throughput sequencing, data analysis, and PAM sequences determination.

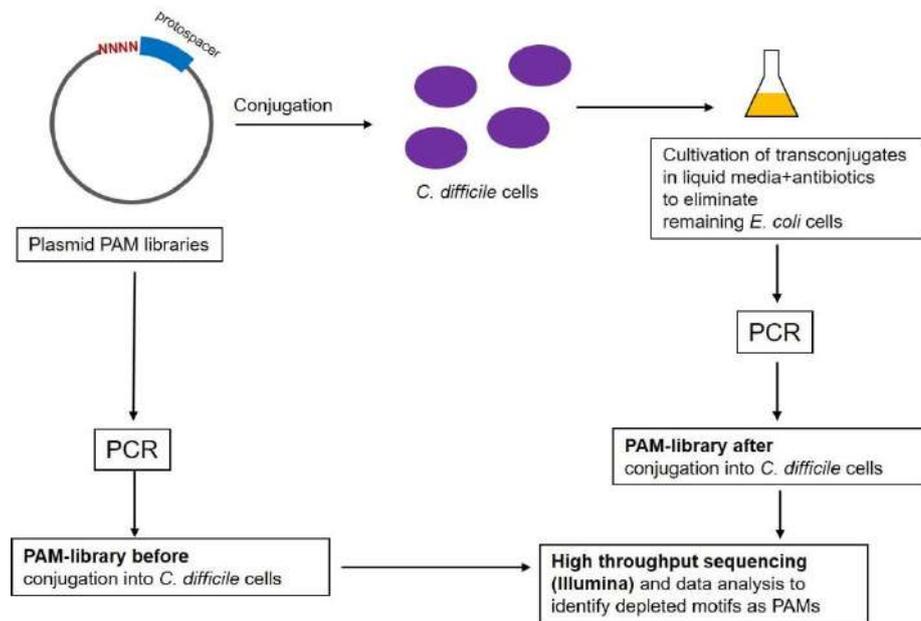


Figure 2.1. PAM libraries experimental strategy for *C. difficile*.

2.3.1.2 PAM sequences determination

After high throughput sequencing, a set of reads was collected for each PAM library and data analysis to identify depleted motifs as PAMs was performed (depletion analysis). The depletion analysis is based on a comparison between the library possessing all PAM sequences (“before the conjugation”) and the library possessing PAM sequences, which were not recognized by *C. difficile* CRISPR-Cas system (“after the conjugation”). Thus, the depletion analysis allows determining all functional PAMs. Using Pearson chi-square test with a p-value less than 10^{-12} statistically significantly depleted PAM sequences were selected. The consensus sequences of selected PAMs were then visualized by the WebLogo tool (Crooks et al., 2004). This analysis suggested that the -4 position of the PAM could not be relevant for *C. difficile* CRISPR-Cas system functioning (data not shown). This is in accordance with previous data obtained *in silico* and during experimental work (Boudry et al., 2015) of functional three-nucleotide PAMs.

WebLogo-based visualization revealed the YCN PAM consensus for both *C. difficile* 630 Δ erm and R20291 strains (Figure 2.2A, B).

For the additional visualization of the results, PAM wheels were constructed for each strain (Figure 2.3A, B). For each individual PAM sequence, a depletion score was counted as the ratio of normalized read count in cleavage reaction and normalized read count in control. The depletion scores were then used as the input for the Krona plot (Leenay et al., 2016). PAM wheels confirmed the general YCN PAM consensus. Minor PAMs were represented as NNN motif that could be due to possible remaining of *E. coli* cells with the plasmids of “before libraries” in “after samples” or due to sequencing errors. These results are in agreement with previously obtained *in silico* PAM prediction and plasmid efficiency assays (Boudry et al., 2015).

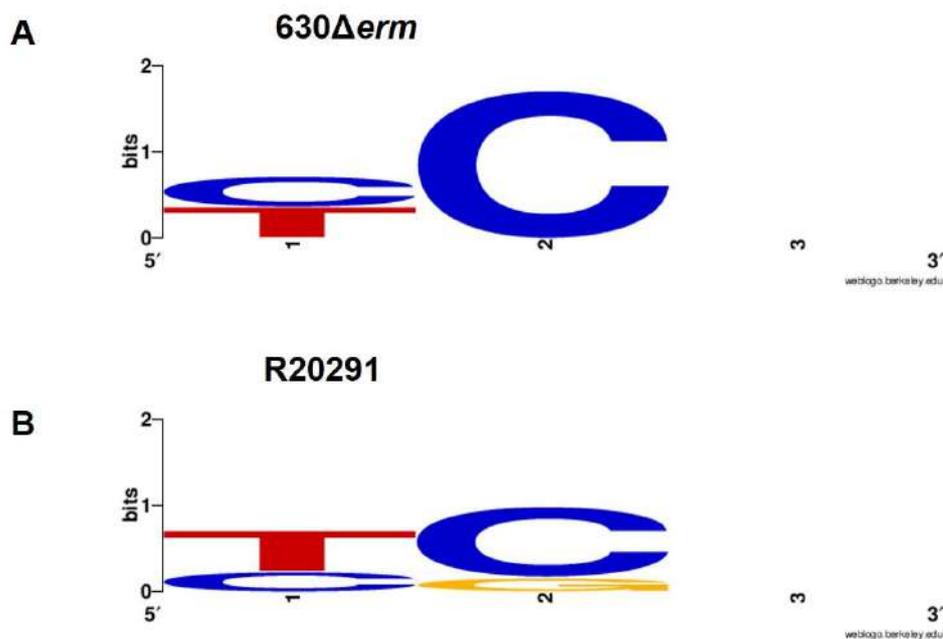


Figure 2.2. PAM-consensus WebLogos for *C. difficile* 630 Δ erm (A) and R20291 (B) strains.

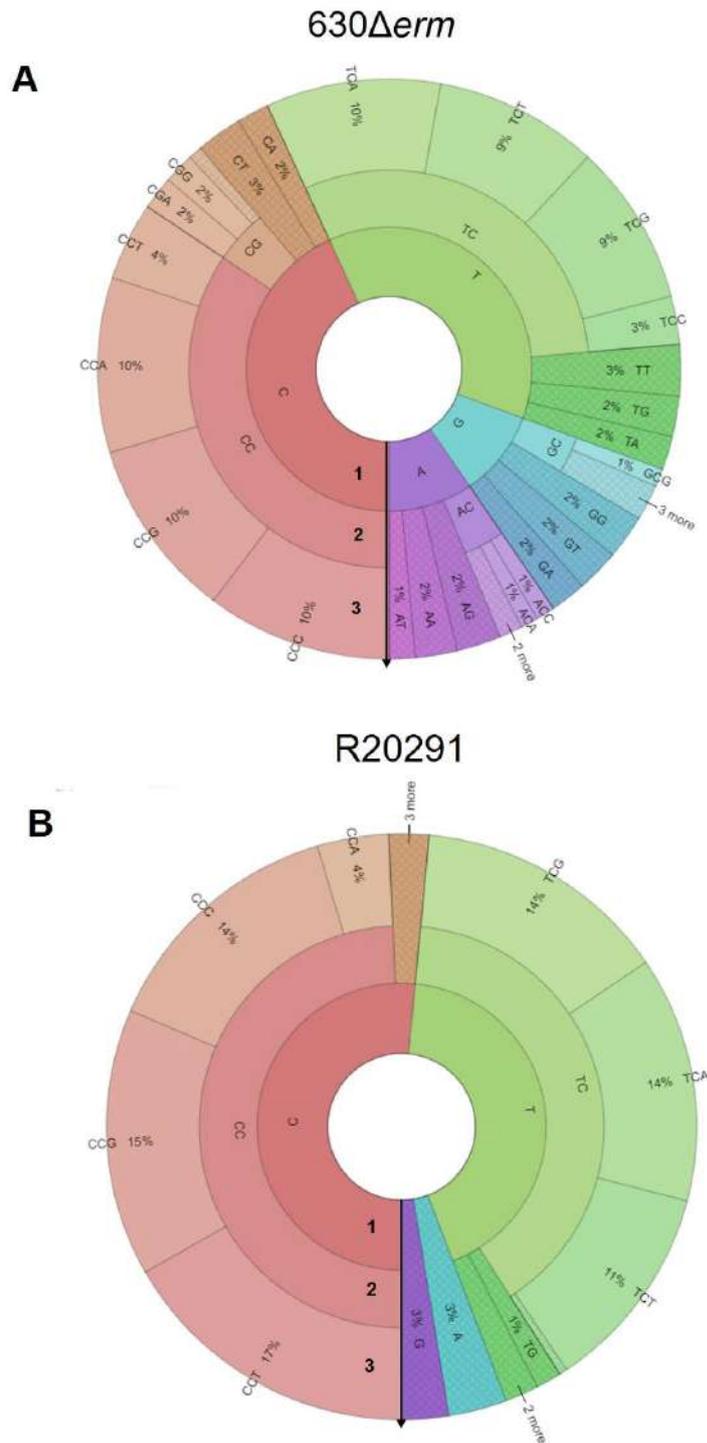


Figure 2.3. PAM wheels for *C. difficile* 630 Δ erm (A) and R202091 (B) strains. Arrows indicate the direction (5' \rightarrow 3') from the 1st nucleotide to the 3^d nucleotide of PAMs. Red sectors correspond to the CCN PAM consensus, and green sectors correspond to the TCN PAM consensus. Different patterns of CCN and TCN PAMs distribution in 630 Δ erm and R202091 strains could be a consequence of the different amount of good-quality selected reads in the libraries “after”.

2.3.2 The functionality of CRISPR interference in *C. difficile*

2.3.2.1 Plasmid interference assays in *C. difficile* 630 Δ erm strain

To study the functionality of CRISPR-Cas system for interference in *C. difficile* 630 Δ erm strain, a set of plasmids was constructed on the basis of the pRPF185 Δ gus vector. These plasmids contained protospacers corresponding to a selected spacer of *C. difficile* 630 Δ erm CRISPR array, flanked by functional CCA PAM on the 5'-end. The PAM sequence was chosen in accordance with analysis from the previous study (Boudry et al., 2015) and PAM libraries results (see 2.3.1). The set of PAM-protospacer carrying plasmids was conjugated into *C. difficile* 630 Δ erm cells, and conjugation efficiency levels were determined. An empty pRPF185 Δ gus vector was used as a conjugation control. The presence of a protospacer with a correct PAM sequence matching a spacer from one of the CRISPR arrays inhibits conjugation efficiency by several orders of magnitude. Hence, higher conjugation efficiencies correspond to lower CRISPR interference levels (Figure 2.4). Conjugation efficiency results are presented in Figure 2.5A. Plasmids carrying protospacers, corresponding to spacers of CRISPR3, 4, 8, and gave no transconjugants. Therefore, CRISPR interference was effective against these plasmids. In contrast, plasmids carrying protospacers, corresponding to spacers of CRISPR6, 7, 9, 10, 11, 12 and 17 arrays, gave transconjugates, and the conjugation efficiency levels in these cases were lower or almost the same as the control. It may mean that these CRISPR-arrays provide lower defense levels and are less active for CRISPR-interference.

Overall, interference levels of *C. difficile* 630 Δ erm CRISPR arrays correlate with their expression levels (Figure 2.5B, C), detected by RNA-seq in the previous work (Boudry et al., 2015). The exceptions are CRISPR10 and 11 arrays, which showed low interference activity (Figure 2.5A). These results demonstrate that almost all CRISPR-arrays of *C. difficile* 630 Δ erm strain are functional for interference.

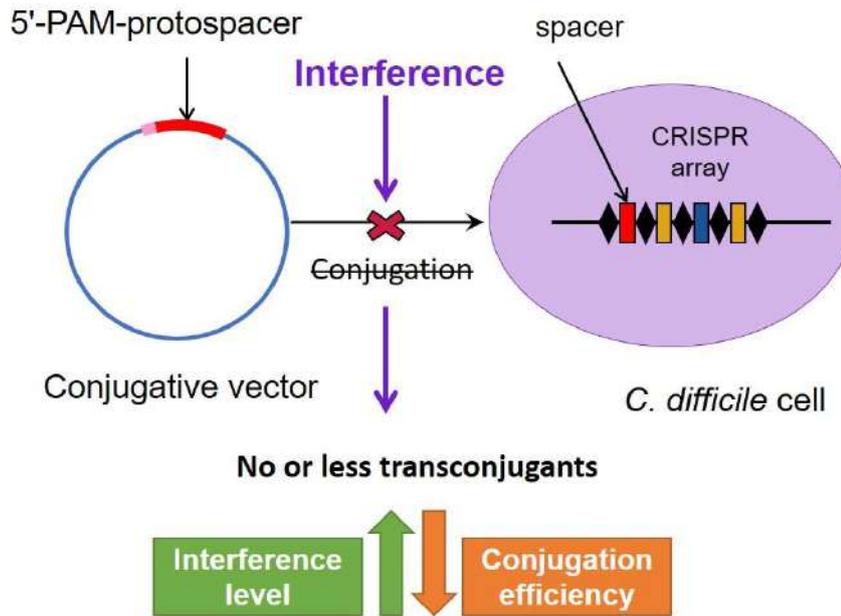


Figure 2.4. Experimental strategy for plasmid interference assays. Conjugative vectors, carrying 5'-PAM-protospacers, corresponding to a selected spacer of *C. difficile* CRISPR arrays were conjugated into *C. difficile* cells, and subsequently, the efficiency of conjugation was determined. Higher conjugation efficiency corresponds to lower interference levels.

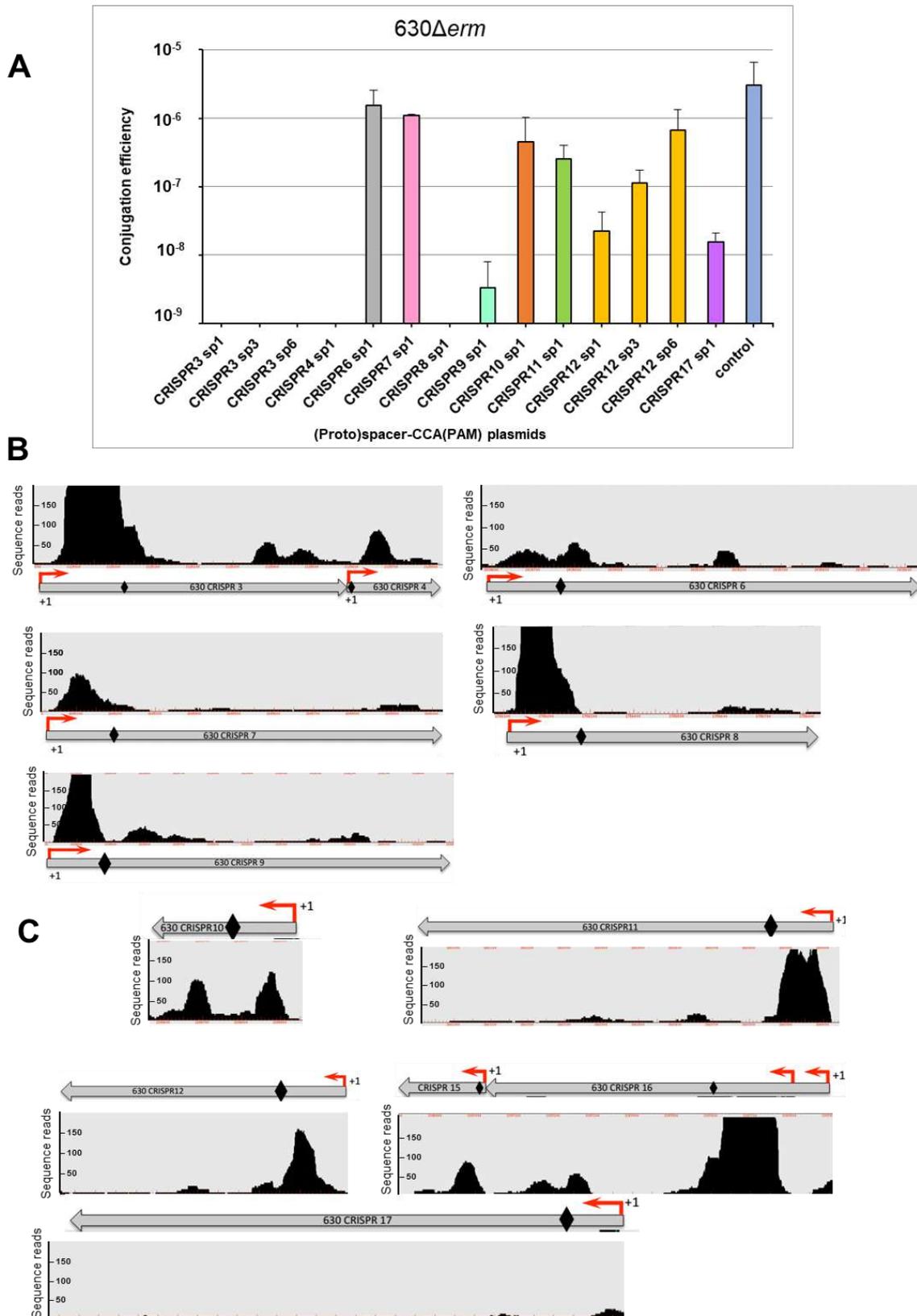


Figure 2.5. Plasmid conjugation efficiency in *C. difficile* 630 Δ erm strain (A) and expression levels of 630 Δ erm strain CRISPR arrays detected by RNaseq (B, C). Expression of CRISPR arrays is presented in their transcriptional order, B – from the + DNA strand, C – from the - DNA strand. Red broken arrows marked with “+1” designate TSS of the arrays and correspond to the

(Figure 2.5. Continue) leader sequences, black diamonds point the beginnings of the first spacers in the arrays. RNAseq figures are adapted from (Boudry et al., 2015) with permission.

2.3.2.2 Plasmid interference assays in *C. difficile* R20291 strain

To investigate the functionality of CRISPR-Cas system for interference in hypervirulent *C. difficile* R20291 strain, a set of conjugative plasmids was created. To construct this set, a protospacer, corresponding to the first spacer of the actively expressed R20291 CRISPR13 array (Figure 2.6B), flanked with several PAM sequences on the 5'-end, was introduced into the pRPF185 Δ *gus* vector. CCA, CCT, CCC, CCG, GAG and AAT trinucleotides were chosen as PAM sequences for these assays. For one plasmid the protospacer contained a mutation of the first nucleotide position, corresponding to the first position of the seed region (the first eight nucleotides of the protospacer, necessary for CRISPR targeting (Semenova et al., 2011), was used. No transconjugants were obtained after the conjugation of CCA, CCT, CCC, and CCG PAM-containing plasmids, which meant a high efficiency of interference (Figure 2.6A). These results confirm *in silico* data, obtained in the previous study (Boudry et al., 2015), and PAM libraries results (see 2.3.1). Plasmids, carrying non-functional GAG and AAT PAMs demonstrated the same conjugation efficiency levels as an empty pRPF185 Δ *gus* vector (Figure 2.6A). The plasmid containing CCA PAM and a mutation in seed region showed significantly lower interference level than non-mutated CCA PAM-protospacer carrying plasmid (Figure 2.6A). This result confirmed the important role of the seed region in protospacer targeting by the CRISPR-Cas system (Semenova et al., 2011).

Altogether, these results indicate that the CRISPR-Cas system of hypervirulent *C. difficile* R20291 is functional for interference and CCA, CCT, CCC, CCG PAM sequences are experimentally confirmed for this strain.

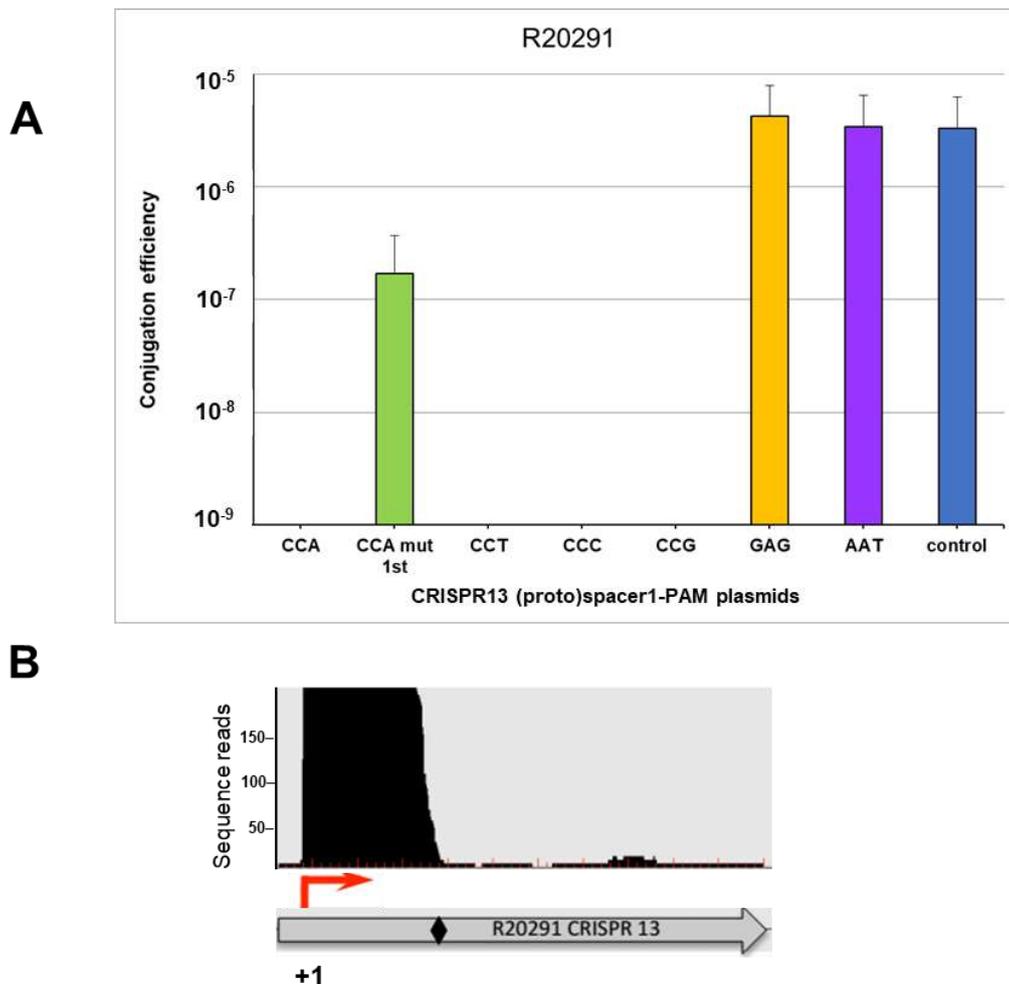


Figure 2.6. Plasmid conjugation efficiency in *C. difficile* R20291 strain (A) and expression level of R20291 strain CRISPR13 array, detected by RNAseq (B). Expression of the CRISPR arrays is presented in its transcriptional order. The red broken arrow, marked with “+1” designates TSS of the array and corresponds to the leader sequences, the black diamond points the beginning of the first spacer in the arrays. The RNAseq figure is adapted from (Boudry et al., 2015) with permission.

2.3.3 The functionality of CRISPR-Cas system for adaptation in *C. difficile*

2.3.3.1 Experimental model of naïve adaptation assays in *C. difficile* 630 Δ erm

Primed adaptation experiments with plasmids, contained PAM(CCA)-protospacers, corresponding to the first spacers of *C. difficile* 630 Δ erm CRISPR12 and 16 arrays and carrying mutations in the seed regions, were unsuccessful. Thus, CRISPR adaptation in native endogenous expression levels of Cas proteins could be non-sufficient for the experimental detection. Therefore, we constructed two variants of plasmids containing *cas1*, *cas2*, and *cas4* genes under the control of the inducible P_{tet} promoter (Table S2.1 in

Supplementary materials) and firstly tested them for naïve adaptation events. The first plasmid variant (pCas1-2) carried *cas1* and *cas2* genes, whereas the second plasmid variant carried the full adaptation module *cas1*, *cas2*, *cas4* (pCas1-2-4). Subsequently, pCas1-2 and pCas1-2-4 were transferred into *C. difficile* 630 Δ *erm* cells by conjugation (Figure 2.7A, B). Transconjugants were then cultivated in BHI medium supplemented with Tm and ATc to maintain plasmids and produce a sufficient amount of Cas proteins. We observed no growth of *C. difficile* cells carrying pCas1-2 caused by a possible toxic effect of overexpression of Cas1 and Cas2 proteins (Figure 2.7A). In contrast, pCas1-2-4 did not give such effect. Therefore, this plasmid was used for further adaptation assays. After overnight growth in BHI medium supplemented with Tm and ATc, pCas1-2-4-containing cells were twice transferred to BHI medium supplemented with ATc without Tm (I and II reseeded) (Figure 2.7B). These additional steps were necessary to gain more bacterial cells with possibly newly acquired spacers. After each reseeded, two rounds of PCR were performed to detect new spacer acquisition. For the PCR amplification, a following set of primers was used: forward primers, which annealed to leader regions of arrays and reverse primers, which annealed to the first or the second spacer (CRISPR10 array) of a native array (Figure 2.8A). *C. difficile* 630 Δ *erm* strain carrying an empty pRPF185 Δ *gus* was used as a control in all naïve adaptation assays.

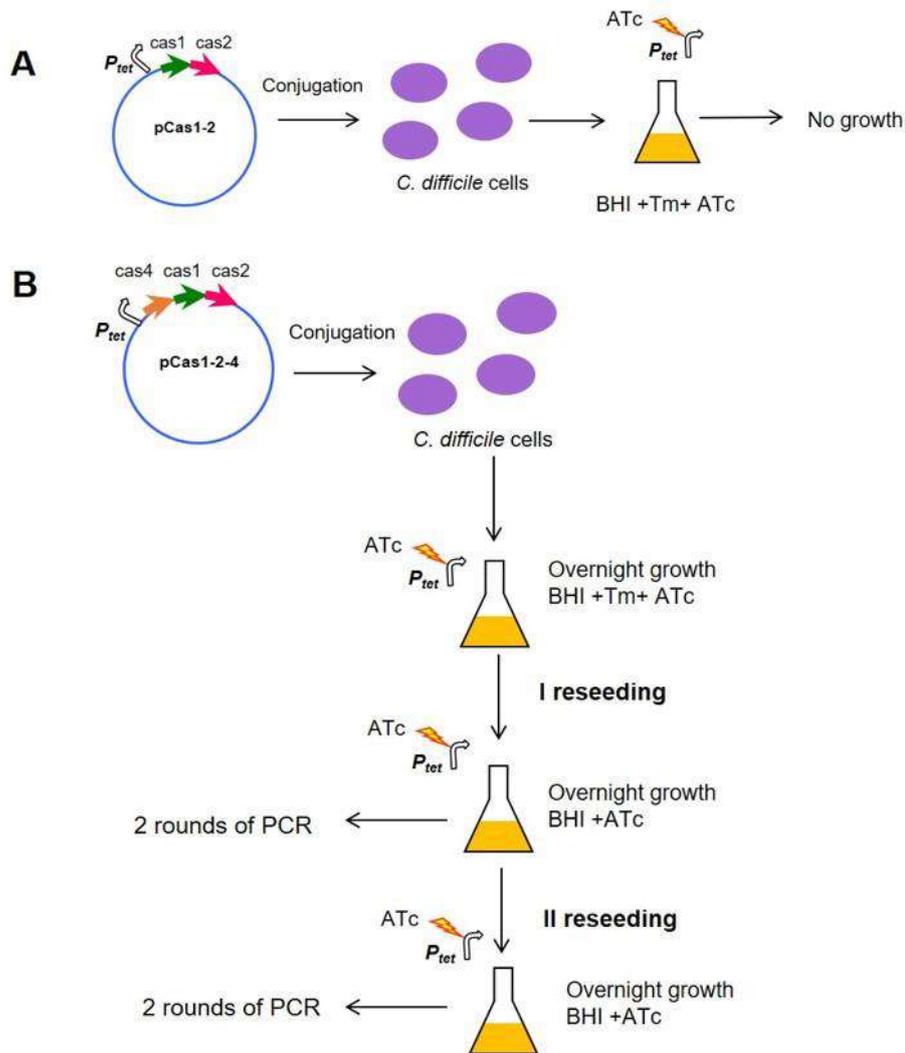


Figure 2.7. Experimental plan of naïve adaptation assays in *C. difficile* 630 Δ erm. **A** – an adaptation assay using pCas1-2, overexpressing Cas1, and Cas2 proteins. No growth was observed after inoculation of transconjugants into BHI+Tm+ATc medium. **B** – an adaptation assay using pCas1-2-4, overexpressing Cas1, Cas2, and Cas4 proteins. Two reseed steps after cultivation of transconjugants in BHI+Tm+ATc medium and following two PCR rounds were performed.

2.3.3.2 Detection of new spacer acquisition

All twelve *C. difficile* 630 Δ erm CRISPR arrays were tested for new spacer acquisition with oligonucleotide primers complementary to the leader region and first spacer of each CRISPR array (Figure 2.8A). After the 1st reseed and two rounds of PCR, no adaptation events were detected (data not shown). New spacer acquisition was observed only in CRISPR8 and CRISPR9 arrays after the 2nd PCR round of the sample

from the second reseedling (Figure 2.8B). These results could indicate that *C. difficile* 630 Δ *erm* CRISPR-Cas system is not highly active for naïve adaptation. Subsequently, DNA bands corresponding to newly acquired spacers were extracted from the gel and used for nested PCR with primers containing Illumina adapters for further high-throughput sequencing and bioinformatic analysis.

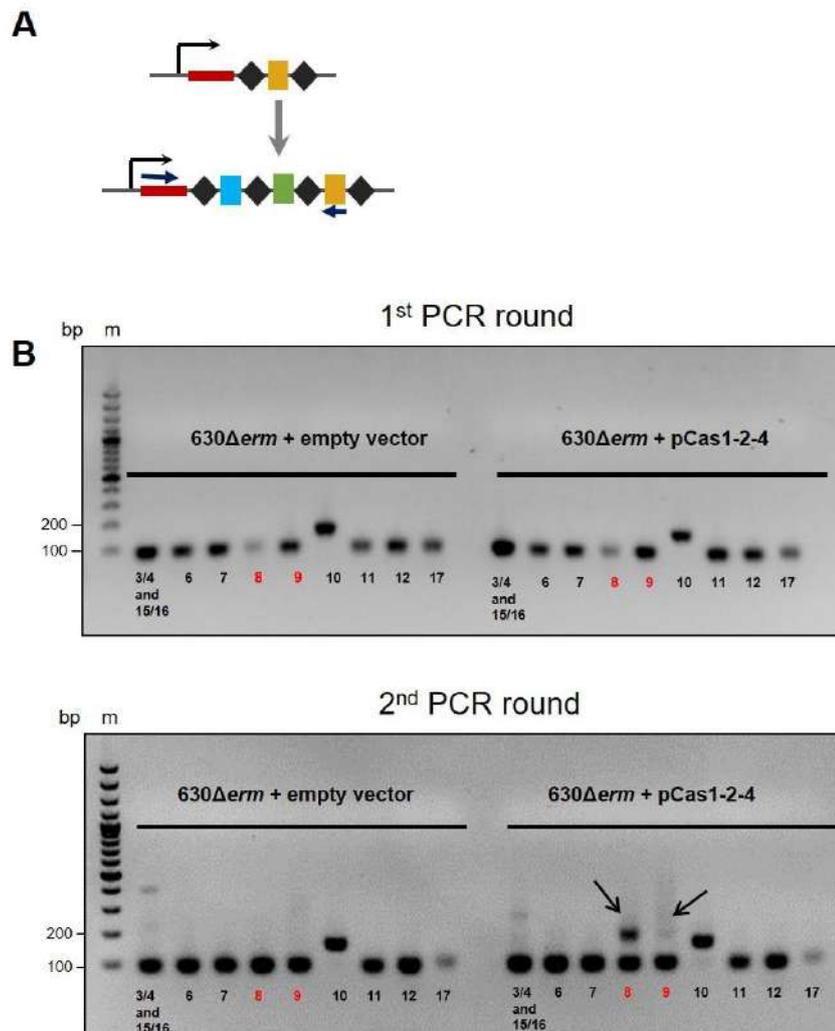


Figure 2.8. PCR analysis of naïve adaptation in *C. difficile* 630 Δ *erm*. **A** – Principal scheme of PCR analysis. PCR amplification was performed using pairs of primers for each CRISPR array of *C. difficile* 630 Δ *erm*. Forward primers annealed to leader regions of arrays and reverse primers annealed to the first or the second spacer (CRISPR10 array) of a native array. **B** – PCR results after the II reseedling step. Numbers bellow PCR bands denote *C. difficile* 630 Δ *erm* CRISPR arrays (CRISPR3/4, CRISPR6, etc.) 89 bp PCR bands correspond to native arrays (155 bp for CRISPR10 array); 155 bp PCR bands correspond to one acquired spacer (221 bp for CRISPR10 array).

2.3.3.3 General analysis of newly acquired spacers

High-throughput sequencing data was analyzed in R using ShortRead and Biostrings packages (Morgan et al., 2009). Spacers of length 10-79 bp were extracted from reads. Then spacers corresponding to the native first spacer and spacers with less than 12 bp were removed from the analysis. Remaining spacers were mapped to the all DNA molecules of *C. difficile* 630 Δ *erm* which could serve as a source for new spacers (chromosome, pCas1-2-4 and native endogenous pCD630 plasmid (Sebahia et al., 2006)) to identify "protospacer" sequences. Additionally, spacers that aligned to multiple positions within the same molecule were removed from the analysis. Spacers, aligned to one DNA molecule were identified as "unique," and spacers, aligned to several molecules (but one position within each molecule) were identified as "non-unique" and analyzed separately (Table 2.1). Some spacers deriving from regions containing a PAM consensus by 1-3 nucleotide upstream of the "protospacer" ("shifters") and spacer, which were inserted into CRISPR arrays in the opposite orientation ("flippers") were removed from the analysis (Shmakov et al., 2014).

General analysis of spacers revealed that 99.77% of all acquired spacers were unique, and 98% of them derived from pCas1-2-4 plasmid (Table 2.1). Only 1.69% of spacers were acquired from the chromosome and 0.07% of spacers derived from pCD630 (Table 2.1). Non-unique spacers constituted 0.23% and were derived from both pCas1-2-4 plasmid and chromosome.

Table 2.1 Statistics of spacers acquired into CRISPR8 (CR8) and CRISPR9 (CR9) arrays.

The table demonstrates the number of aligned spacers and the number of spacers mapped to different DNA molecules (chromosome, pCas1-2-4, and pCD630). All values are shown including or excluding shifters and flippers. Spacers that have multiple alignments but on different molecules (non-unique alignment) were further separately. Percentages of spacers, used in further analysis, are shown in red.

Array	Shifters/Flippers	Unique alignment				Non-unique alignment				Total
		All DNAs	Chromosome	pCas1-2-4	pCD630	All DNAs	Chromosome	pCas1-2-4	pCD630	
CR8	Included	235414	4478	230770	166	431	431	431	0	235845
		99.82%	1.90%	97.85%	0.07%	0.18%	0.18%	0.18%	0.00%	
	Removed	220393	3303	216946	144	329	329	329	0	220722
		99.85%	1.50%	98.29%	0.07%	0.15%	0.15%	0.15%	0.00%	
CR9	Included	91065	4131	86875	59	435	435	435	0	91500
		99.52%	4.51%	94.95%	0.06%	0.48%	0.48%	0.48%	0.00%	
	Removed	78577	1774	76748	55	375	375	375	0	78952
		99.53%	2.25%	97.21%	0.07%	0.47%	0.47%	0.47%	0.00%	
CR8 + CR9	Included	326479	8609	317645	225	866	866	866	0	327345
		99.74%	2.63%	97.04%	0.07%	0.26%	0.26%	0.26%	0.00%	
	Removed	298970	5077	293694	199	704	704	704	0	299674
		99.77%	1.69%	98.00%	0.07%	0.23%	0.23%	0.23%	0.00%	

2.3.3.4 Analysis of the distribution of spacer lengths and frequencies

Next, the detailed analysis of spacer lengths distribution (Figure 2.9A) and adaptation “hot spots” was performed (Figure 2.10A, B). Among unique protospacers on the chromosome, there were several positions highly enriched with spacers. Analysis of the distribution of spacer frequencies showed that the sum of values above the 99th percentile gave 24-34% of all spacers mapped to the chromosome. Presumably, these were PCR artifacts or spacers acquired by bacteria at the early time point and spread among later generations. Therefore, outliers above the 99th percentile were removed from further analysis of total percent of spacers in different regions of the chromosome as well as analysis of the distribution of spacer length and frequencies of different PAM sequences.

Analysis of the distribution of spacer lengths showed that almost all newly acquired spacers had 36-37 bp length on average. This agrees with length distribution of native spacers in CRISPR8, CRISPR9 arrays and all CRISPR arrays of *C. difficile* 630 Δ erm (Fig. 2.9A, B).

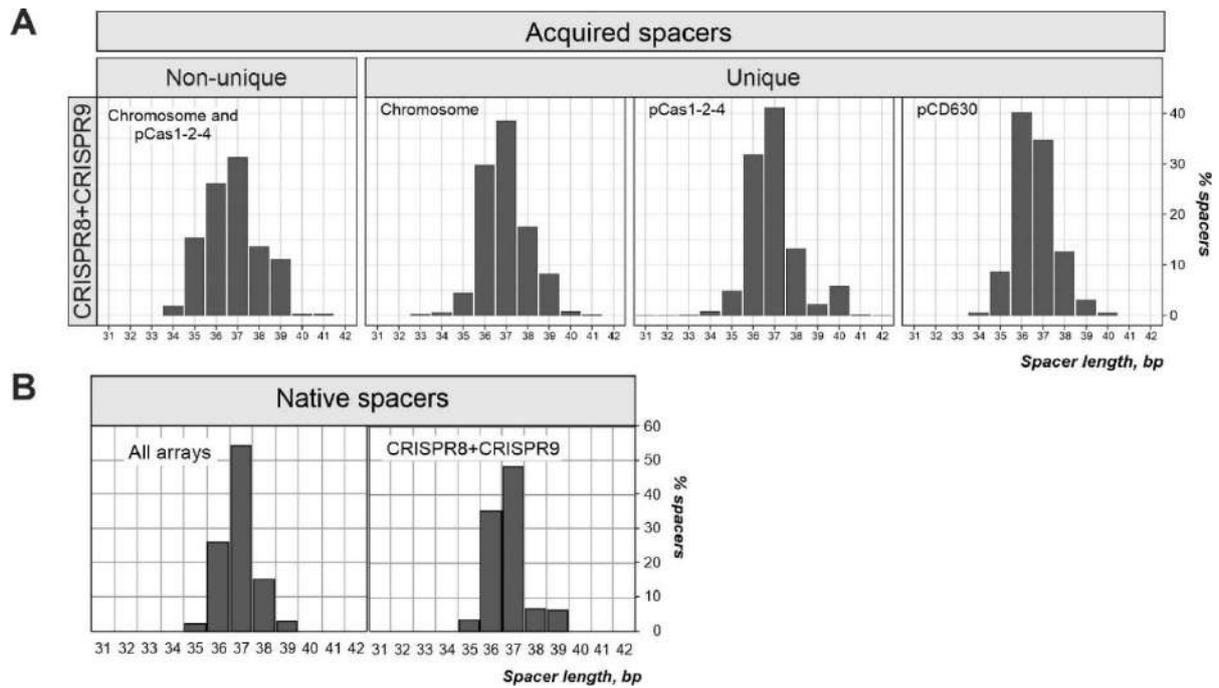


Figure 2.9. The distribution of spacer lengths. **A** – lengths of spacers, acquired into CRISPR8 and CRISPR9 arrays during the adaptation; **B** – lengths of native spacers in *C. difficile* 630 Δ *erm* CRISPR arrays.

Analysis of the distribution of unique spacer frequencies derived from the chromosome revealed several adaptation “hot spots”. Spacers were the most actively acquired from the chromosome in *terC* (replication termination site) region and regions, containing Tn1549-like genes (Figure 2.10A).

As it was mentioned above, the most abundant number of new spacers derived from the pCas1-2-4 plasmid. The most frequent unique spacers were acquired from *traJ* (a regulator for conjugative genes expression), *oriT* (origin of transfer) and *ori* (origin of replication) regions (Figure 2.10B). In contrast, significantly less spacers were acquired from the native pCD630 plasmid (Figure 2.10C). The adaptation “hot spots” for pCD630 were localized at *p70* gene region and close to the *p80* gene region.

Analysis of non-unique spacers distribution revealed that they originated from regions, corresponding to *cas1*, *cas2*, and *cas4* both in chromosome and pCas1-2-4 plasmid (Figure 2.10 A, B).

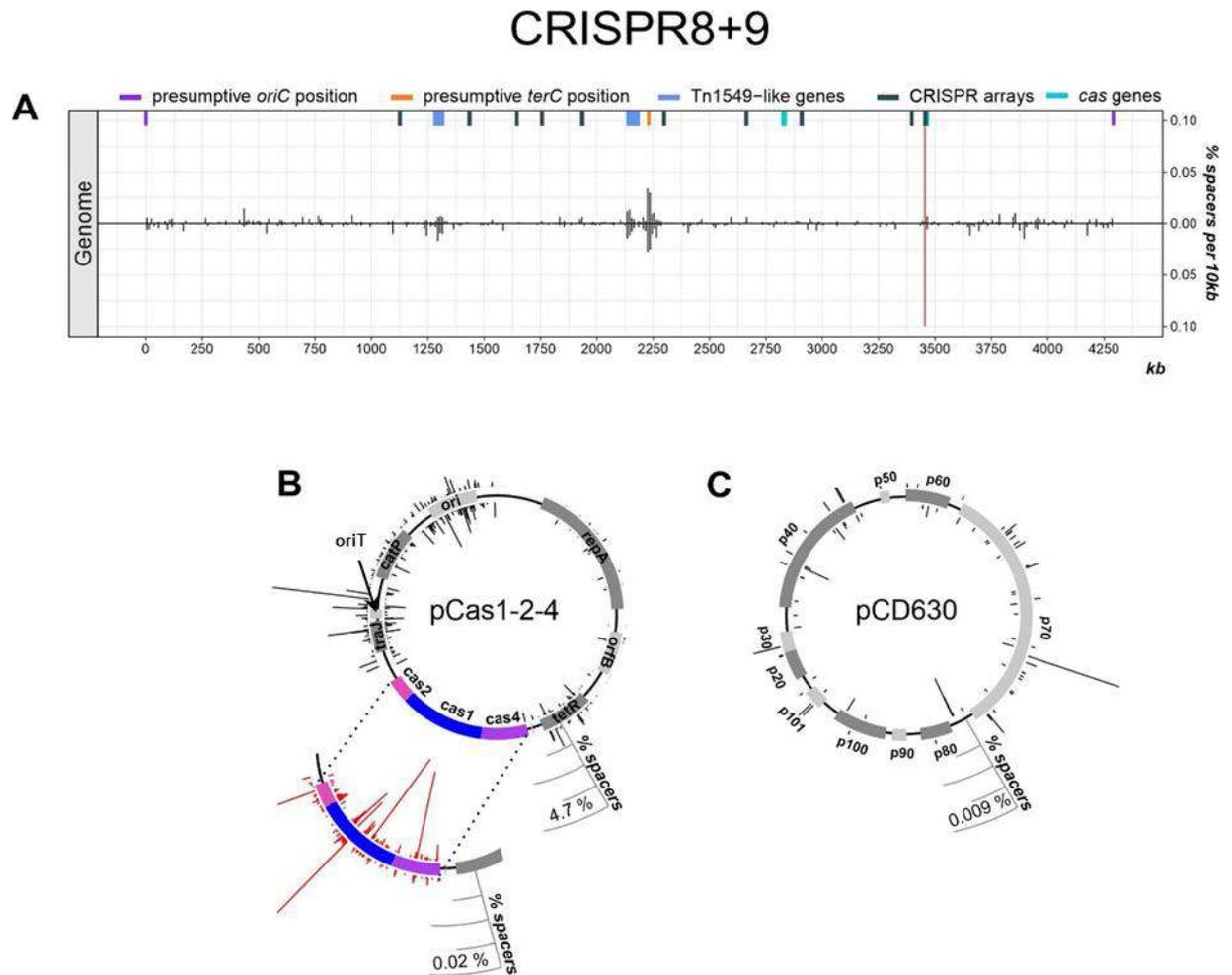


Figure 2.10. The distribution of spacers, aligned to the chromosome (A), pCas1-2-4 plasmid (B) and pCD630 plasmid (C). The height of black bars indicates the percentage of spacers aligned to certain positions on DNA molecules. Bars localized above and below chromosome line (A) and inside and outside of plasmid circles designate spacers derived from different strands of DNA. Red bars indicate non-uniquely aligned spacers.

2.3.3.5 Definition of PAM sequences, corresponding to acquired spacers

PAM sequences were defined as three nucleotides upstream of the first position of a protospacer, which was mapped on DNA molecules during the analysis. For uniquely aligned spacers on the pCas1-2-4 plasmid, the abundant PAMs were CCA, CCG, CCT, and CCC (Figure 2.11). Additionally, CCA motif had the highest percentage in pCD630 case (Figure 2.11). These results confirm the functionality of the CCN PAM consensus for adaptation. In contrast, there were no clear PAM distribution peaks for uniquely aligned spacers on the chromosome (Figure 2.11). For the genome, CCN and TCN consensus PAMs corresponded to less than 10% of spacers.

Moreover, analysis of PAM sequences, corresponding to non-unique aligned spacers revealed that CCT and CCA motifs were the most presented both on pCas1-2-4 and chromosome (Figure 2.11).

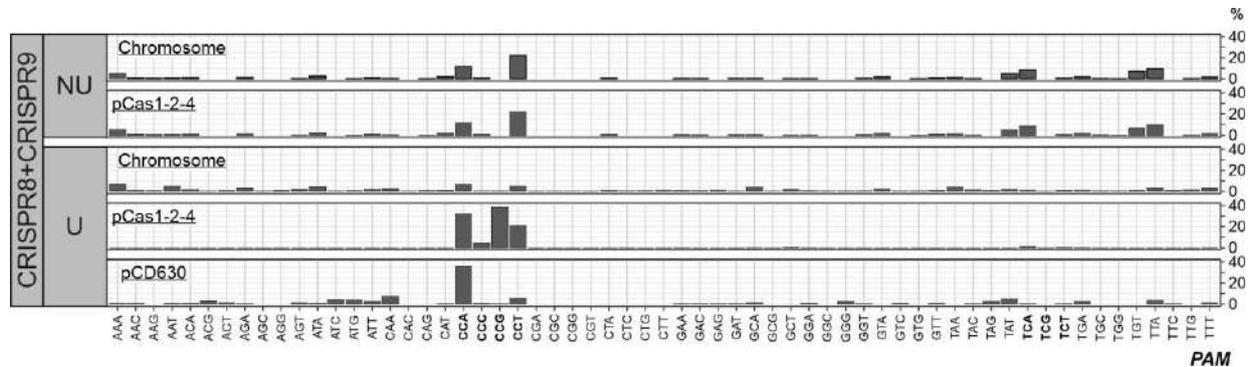


Figure 2.11. The distribution of PAM sequences, corresponding to acquired spacers. NU – non-uniquely aligned spacers; U – uniquely aligned spacers. Functional PAM sequences are indicated in boldface.

2.3.4 Role of multiple *cas* operons

2.3.4.1 Construction of the *C. difficile* 630 Δ *erm* full *cas* operon deletion mutant

As it was mentioned in Chapter 1, the majority of *C. difficile* sequenced strains are characterized by two or three (in 027 ribotype) I-B subtype *cas* gene sets (Boudry et al., 2015). Reference *C. difficile* 630 strain possesses two *cas* operons (full *CD2975-2982* and partial *CD2451-2455*) and hypervirulent *C. difficile* R20291 strain possesses three *cas* operons (full *CDR20291_2810-2817*, partial *CDR20291_2344-2348* and additional *CDR20291_2994-2998*). To study the role of each *cas* operon in *C. difficile* life and infection cycles and in CRISPR-Cas system functionality, we attempted to construct *cas* operon deletion mutants in both 630 and R20291 strains using *codA* allele exchange strategy (Cartman et al., 2012). To date, we have successfully obtained only one full *cas* operon deletion mutant in *C. difficile* 630 Δ *erm* strain (*C. difficile* 630 Δ *erm* Δ *CD2975-2982*). The absence of full *cas* operon was confirmed by PCR (Figure 2.12) as well as by qRT-PCR (data not shown). For the qRT-PCR, the first gene of the full *cas* operon (*CD2982*) was used as a target gene. No expression of *CD2982* gene was observed in *C. difficile* 630 Δ *erm* Δ *CD2975-2982* mutant.

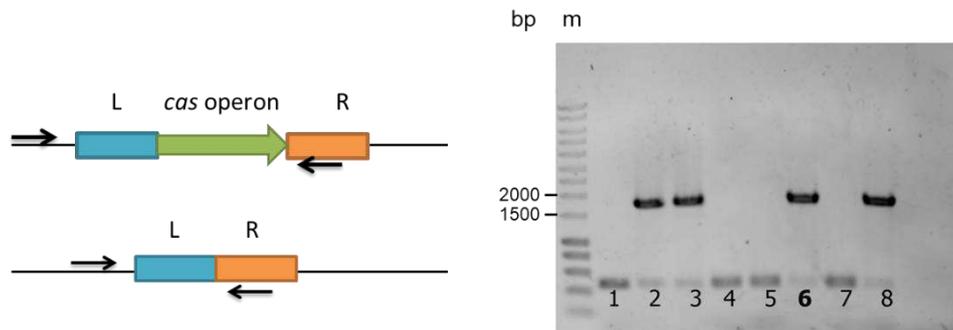


Figure 2.12. PCR analysis of obtained clones after the *codA* allele exchange genome editing to verify *C. difficile* 630 Δ *erm* Δ *CD2975-2982 mutants.* 1738 bp PCR bands correspond to the mutant genotypes; wild-type genotypes did not give PCR bands due to the too large size of the possible PCR products. The 6th clone (in boldface) lost the plasmid after the editing, and it was used for further experiments. L and R – left and right regions, flanking the *cas* operon, which were used as arms for the allele exchange.

2.3.4.2 Growth of the *C. difficile* 630 Δ *erm* Δ *CD2975-2982 mutant*

To investigate the possible role of the full *cas* operon in *C. difficile* 630 Δ *erm* growth in standard culture conditions, we performed growth experiments in liquid BHI medium at 37°C using plate reader (Promega). No growth differences between the wild type the mutant was observed after the 24h of growth (Figure 2.13).

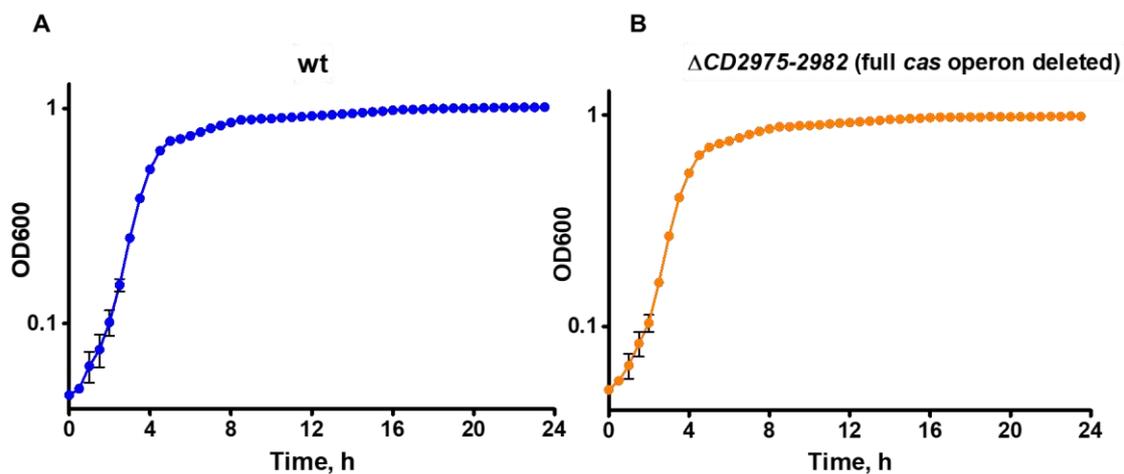


Figure 2.13. Growth curves of *C. difficile* 630 Δ *erm* (wt) (A) and *C. difficile* 630 Δ *erm* Δ *CD2975-2982 (B) strains in BHI medium at 37°C.*

2.3.4.3 Plasmid interference assays in *C. difficile* 630 Δ erm Δ CD2975-2982 mutant

To compare interference activity of the wild-type strain and the mutant lacking the full *cas* operon, plasmid interference experiments were performed. For these assays, we used plasmids containing protospacers corresponding to spacers from *C. difficile* 630 Δ erm CRISPR3, 11, 12 and 17 arrays, and flanked by functional CCA PAM on their 5'-end (see 2.3.2 in Results). An empty pRPF185 Δ gus vector was used as a control. Conjugation efficiency results are presented in Figure 2.14. The full *cas* operon mutant showed lower interference levels than the wild-type strain after conjugation with plasmids, carrying CRISPR12 spacer6 and CRISPR17 spacer1. According to RNA-seq results, these spacers are not actively expressed (Figure 2.5C). Thus, these results demonstrate that *C. difficile* 630 Δ erm Δ CD2975-2982 mutant is less effective for plasmid interference than the wild type, although this effect is observed only in non-actively expressed spacers case.

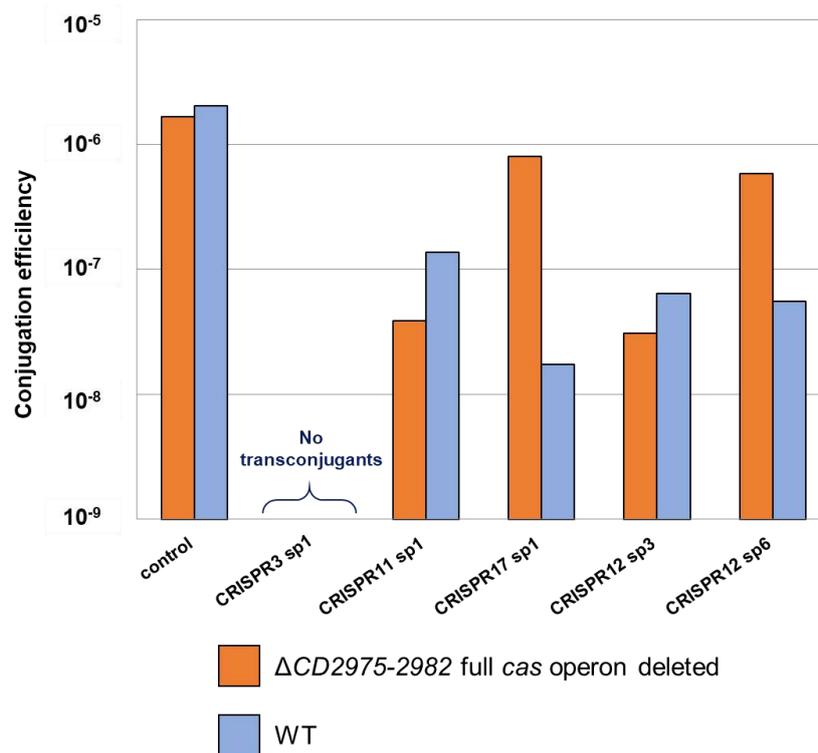


Figure 2.14. Plasmid conjugation efficiency in *C. difficile* 630 Δ erm strain (wt) and *C. difficile* 630 Δ erm Δ CD2975-2982 mutant.

2.4 Discussion

Defensive CRISPR-Cas systems provide adaptive immunity in prokaryotes. They recognize and eliminate foreign DNA agents, such as viruses and plasmids. CRISPR-Cas systems functions may play an important role in the prokaryotic cell physiology. In this Chapter, we present a study of functional aspects of *C. difficile* CRISPR-Cas system.

PAM library experiments allowed us to determine a general PAM consensus sequence (YCN) for *C. difficile* 630 Δ *erm* and R20291 CRISPR-Cas systems. These results confirm previous PAM identification data, obtained *in silico* by the alignment of existing spacers and matching protospacers (Boudry et al., 2015). Interestingly, multiple PAM sequences were found in other type I-B systems (Shah et al., 2013). Moreover, the definition of all possible functional PAMs is necessary for *C. difficile* CRISPR-Cas system biotechnological applications, particularly for genome editing in this pathogen and for developing of new drug types against CDI (see Chapter 4).

CRISPR-Cas systems affect their targets by the interference process. Plasmid interference assays showed the functionality of *C. difficile* CRISPR-Cas system. Experiments with *C. difficile* 630 Δ *erm* strain demonstrated that defense levels of different arrays generally correspond to their expression rates, identified previously (Boudry et al., 2015). Therefore, various CRISPR arrays may differentially contribute to CRISPR-based immunity in 630 Δ *erm* strain. However, an additional detailed quantitative analysis of 630 Δ *erm* CRISPR arrays expression is required for deeper understanding of this correlation. Interference assays, performed in R20291 strain, showed the active defensive function of its CRISPR-Cas system for the first time. Additionally, we experimentally confirmed the functionality of CCA, CCT, CCC, and CCG PAMs in these experiments. Furthermore, obtained results proved the role of the 1st nucleotide of the seed region for protospacer recognition by the effector complex (Semenova et al., 2011). The mutation of this nucleotide decreases the interference efficiency, possibly by reducing the frequency of the target recognition and binding by the effector complex, but at the same time, it does not completely inhibit interference.

In the present Chapter, the functionality of *C. difficile* 630 Δ *erm* CRISPR-Cas system in naïve adaptation was also revealed. We did not detect a new spacer acquisition under the native conditions of Cas proteins expression. Hence, plasmid vectors, overexpressing Cas1, Cas2, and Cas4 proteins were used in the experiments. During overexpressing essential for adaptation Cas1 and Cas2 proteins, we detected no growth of

C. difficile cells. In contrast, overexpression of the full adaptation module (Cas, Cas2, and Cas4) did not give such effect. It could be due to the possible toxic effect of Cas1-Cas2 complex hyperproduction. Several studies showed that Cas2 protein is a derivative of the VapD toxin family (Makarova et al., 2006, 2011a). While Cas1 is the main exonuclease in the Cas1-Cas2 adaptation complex (Nuñez et al., 2015b), nuclease activity of the Cas2 protein is not required for new spacer acquisition (Nuñez et al., 2014). It is suggested that in some CRISPR-Cas systems, Cas2 may have an additional function of an RNase or a toxin (Koonin and Zhang, 2016). Another possible reason for Cas1-Cas2 overexpression toxicity in *C. difficile* cells is the absence of corresponding overexpression of Cas4 protein. Cas4 is a part of the adaptation protein complex of the CRISPR-Cas systems, in which it is present (Lee et al., 2019). This protein participates in the selection and processing of the pre-spacers, defines the correct PAM and provides correct orientation of new spacers during their integration into the CRISPR array (Amitai and Sorek, 2016; Kieper et al., 2018; Lee et al., 2018, 2019; Zhang et al., 2019). The relatively low amount of Cas4, imbalanced with the hyperproduction of Cas1 and Cas2, might lead to the formation of incomplete adaptation complexes, which had some toxic effects on *C. difficile* cells. Further research is needed to better understand the role of Cas4 protein in the *C. difficile* CRISPR adaptation process.

The new spacer acquisition was detected only in *C. difficile* 630 Δ *erm* CRISPR8 and CRISPR9 arrays, which are also active in interference. The majority of new spacers derived from the pCas1-2-4 plasmid, although other spacers were also mapped to the chromosome and the native pCD630 plasmid. In the case of chromosomal loci, the spacers were the most frequently acquired from the *terC* (replication termination site) regions and Tn1549-like regions. Tn1549 is a conjugative transposon, which provides resistance to vancomycin in *Enterococcus faecalis* (Garnier et al., 2000). Subsequently, a significant number of chromosome-derived spacers were obtained from a MGE part of *C. difficile* 630 Δ *erm* genome. The protospacer “hot spots” inside chromosomal transposon regions were also observed in archaea species *Pyrococcus furiosus* (Shiimori et al., 2017). The suggested source for the spacer selection is DNA double-strand breaks (Levy et al., 2015). Therefore, the adaptation “hot spots” are often localized in the regions with the high possibility of the replication-associated DNA double-strand breaks, such as *terC*. Furthermore, the *terC* region of *E. coli* chromosome was shown to be the most active spot for the spacer acquisition (Levy et al., 2015). In the case of acquisition from the pCas1-2-4 plasmid, the largest number of new spacers derived from *oriT* (origin of transfer) and

ori (origin of replication) regions, which the most likely co-localized with double-strand breaks, occurring on the plasmid DNA.

In addition, alignment of newly acquired spacers and *C. difficile* 630 Δ *erm* DNA molecules allowed us to determine PAM sequences of the corresponding protospacers, and it confirmed the functionality of the CCN motif. Notably, the functional PAMs were rather rare in chromosome and pCD630 plasmid cases. The poor distribution of functional PAMs and the small number of the chromosome-deriving spacers could be due to the lethal effect of the acquisition of functional spacers matching the chromosome. Additionally, non-efficient spacer acquisition from the pCD630 suggests, that this plasmid carries crucial genes for *C. difficile* 630 Δ *erm* physiology.

Overall, adaptation assays results demonstrate that *C. difficile* 630 Δ *erm* CRISPR-Cas system is not highly active for naïve adaptation with plasmids since the new spacer acquisition was detected only after overexpression of the adaptation *cas*-module and two rounds of the PCR analysis.

C. difficile CRISPR-Cas system is characterized by two or three I-B subtype *cas* operons, while other prokaryotes usually possess multiple *cas* gene sets, belonging to different types of CRISPR-Cas systems (Li et al., 2015; Patterson et al., 2016; Shiimori et al., 2017; Silas et al., 2017). In this Chapter, a *C. difficile* 630 Δ *erm* full *cas* operon deletion mutant, i.e. containing only partial *cas* operon, was investigated. We observed no differences in growth between the mutant and the wild type strain. Nevertheless, plasmid interference assays revealed that the mutant was less effective for interference than the wild type strain, but only when less-actively expressed spacers were assessed. Possibly, the observed effect could be due to decreased levels of the Cas proteins in the mutant, which influenced the interference levels.

In conclusion, our results reveal the functionality of *C. difficile* CRISPR-Cas system that could protect this bacterium against foreign DNA invaders.

2.5 Supplementary materials

Table S2.1. Bacterial strains and plasmids used in Chapter 2

Strain	Genotype	Source
<i>E. coli</i>		
NEB-10 beta	$\Delta(ara-leu)$ 7697 <i>araD139 fhuA</i> $\Delta lacX74 galK16 galE15 e14- \phi 80dlacZ\Delta M15 recA1 relA1 endA1 nupG rpsL$ (<i>StrR</i>) <i>rph spoT1</i> $\Delta(mrrhsdRMS-mcrBC)$	New England Biolabs
HB101 (RP4)	<i>supE44 aa14 galK2 lacY1</i> $\Delta(gpt-proA)$ 62 <i>rpsL20</i> (<i>StrR</i>) <i>xyl-5 mtl-1 recA13</i> $\Delta(mcrC-mrr)$ <i>hsdSB (rB-mB-)</i> RP4 (<i>Tra+ IncP ApR KmR TcR</i>)	Laboratory stock
<i>C. difficile</i>		
630 Δerm	Sequenced reference strain $\Delta ermB$	Laboratory stock (Hussain et al., 2005)
R20291	PCR-ribotype 027 epidemic strain	Laboratory stock
CNRS_CD059	630 Δerm carrying pRPF Δgus	This work
CNRS_CD001	630 Δerm carrying pCas1-2 plasmid	This work
CNRS_CD002	630 Δerm carrying pCas1-2-4 plasmid	This work
CDIP741	630 Δerm $\Delta CD2975-2982$ (full <i>cas</i> operon)	This work
Plasmid	Description	Reference
pRPF185 Δgus	pRPF185 Δgus vector derivative	(Fagan and Fairweather, 2011; Soutourina et al., 2013)
pDIA6435	pRPF185 Δgus with the 5' CCA-PAM protospacer, corresponding to the spacer1 from 630 Δerm CRISPR3 array	This work
pDIA6436	pRPF185 Δgus with the 5' CCA-PAM protospacer, corresponding to the spacer1 from 630 Δerm CRISPR4 array	This work
pDIA6437	pRPF185 Δgus with the 5' CCA-PAM protospacer, corresponding to the spacer1 from 630 Δerm CRISPR6 array	This work
pDIA6438	pRPF185 Δgus with the 5' CCA-PAM protospacer, corresponding to the spacer1 from 630 Δerm CRISPR7 array	This work
pDIA6439	pRPF185 Δgus with the 5' CCA-PAM protospacer, corresponding to the spacer1 from 630 Δerm CRISPR8 array	This work
pDIA6440	pRPF185 Δgus with the 5' CCA-PAM protospacer, corresponding to the spacer1 from 630 Δerm CRISPR9 array	This work

Table S2.1. Continue.

pDIA6441	pRPF185 Δ <i>gus</i> with the 5' CCA-PAM protospacer, corresponding to the spacer1 from 630 Δ <i>erm</i> CRISPR10 array	This work
pDIA6442	pRPF185 Δ <i>gus</i> with the 5' CCA-PAM protospacer, corresponding to the spacer1 from 630 Δ <i>erm</i> CRISPR11 array	This work
pDIA6443	pRPF185 Δ <i>gus</i> with the 5' CCA-PAM protospacer, corresponding to the spacer1 from 630 Δ <i>erm</i> CRISPR12 array	This work
pDIA6444	pRPF185 Δ <i>gus</i> with the 5' CCA-PAM protospacer, corresponding to the spacer1 from 630 Δ <i>erm</i> CRISPR17 array	This work
pDIA6445	pRPF185 Δ <i>gus</i> with the 5' CCA-PAM protospacer, corresponding to the spacer3 from 630 Δ <i>erm</i> CRISPR3 array	This work
pDIA6446	pRPF185 Δ <i>gus</i> with the 5' CCA-PAM protospacer, corresponding to the spacer6 from 630 Δ <i>erm</i> CRISPR3 array	This work
pDIA6447	pRPF185 Δ <i>gus</i> with the 5' CCA-PAM protospacer, corresponding to the spacer3 from 630 Δ <i>erm</i> CRISPR12 array	This work
pDIA6448	pRPF185 Δ <i>gus</i> with the 5' CCA-PAM protospacer, corresponding to the spacer6 from 630 Δ <i>erm</i> CRISPR12 array	This work
pDIA6475	pRPF185 Δ <i>gus</i> with the 5' CCC-PAM protospacer, corresponding to the spacer1 from R20291 CRISPR13 array	This work
pDIA6476	pRPF185 Δ <i>gus</i> with the 5' GAG-PAM protospacer, corresponding to the spacer1 from R20291 CRISPR13 array	This work
pDIA6477	pRPF185 Δ <i>gus</i> with the 5' CCT-PAM protospacer, corresponding to the spacer1 from R20291 CRISPR13 array	This work
pDIA6478	pRPF185 Δ <i>gus</i> with the 5' AAT-PAM protospacer, corresponding to the spacer1 from R20291 CRISPR13 array	This work
pDIA6479	pRPF185 Δ <i>gus</i> with the 5' CCA-PAM protospacer with a mutation in the 1 st position, corresponding to the spacer1 from R20291 CRISPR13 array	This work
pDIA6480	pRPF185 Δ <i>gus</i> with the 5' CCA-PAM protospacer, corresponding to the spacer1 from R20291 CRISPR13 array	This work
pDIA6493	pRPF185 Δ <i>gus</i> with the 5' CCG-PAM protospacer, corresponding to the spacer1 from R20291 CRISPR13 array	This work
pCas1-2	pRPF185 Δ <i>gus</i> carrying 630 Δ <i>erm cas1</i> and <i>cas2</i> genes under the control of P _{tet} promoter	This work

Table S2.1. Continue.

pCas1-2-4	pRPF185 Δ <i>gus</i> carrying 630 Δ <i>erm cas1, cas2</i> and <i>cas4</i> genes under the control of P _{tet} promoter	This work
pMTL-SC7315	Semi-suicidal vector carrying <i>codA</i>	(Cartman et al., 2012)
pDIA6495	pMTL-SC7315 carrying arms for the recombination in 630 Δ <i>erm</i> strain to delete full <i>cas</i> operon (<i>CD2975-2982</i>)	This work

Chapter 3. Regulation of *C. difficile* CRISPR-Cas system

Several parts of this chapter are published (Maikova et al., 2018a):

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. Oxford University Press.* 2018; 46: 4733-4751.

3.1 Introduction

During the infection cycle, *C. difficile* must cope with changing conditions and various stresses inside the complex colon environment. Additionally, this bacterium interacts with phages and other MGE. The CRISPR-Cas system could play an important role in *C. difficile* adaptation inside the host. Therefore, its functions and activity can be regulated in the response to signals of changing environments.

To survive in different unfavorable and stress conditions inside the host, pathogenic bacteria commonly form biofilms. Biofilms are bacterial communities, developed on surfaces, and they are a “protected mode” of bacterial growth (Hall-Stoodley et al., 2004). Several studies demonstrated that *C. difficile* forms biofilms during its infection cycle (Dapa et al., 2013; Nale et al., 2016; Soavelomandroso et al., 2017). In biofilm conditions, cell populations are highly dense, which favor possible phage infection and HGT (Babic et al., 2011; Abedon, 2012). Thereby, bacterial defense systems (such as CRISPR-Cas) should be regulated by biofilm-related and often stress-associated factors.

c-di-GMP is a bacterial secondary messenger controlling diverse processes in bacterial cells, and it is mostly known to be an important signal molecule for the transition from the planktonic, motile lifestyle to the biofilm lifestyle (Bordeleau et al., 2011). It was shown that *C. difficile* possesses several c-di-GMP-dependent riboswitches, involved in the control of motility and biofilm formation (Soutourina et al., 2013) (see 1.1.3 in Chapter 1). This Chapter demonstrates a possible role of high c-di-GMP levels in *C. difficile* CRISPR-Cas system regulation and the potential link between one c-di-GMP-I riboswitch and CRISPR-Cas system functioning.

TA modules are two-component genomic systems, encoding a stable “toxin” and an unstable “antitoxin” (Page and Peti, 2016). The overexpression of toxin either kills cells

or confers growth stasis. TA systems have been initially discovered on plasmids where they confer stability of maintenance through post-segregation killing (Hayes, 2003). TA systems have also been found on bacterial and archaeal chromosomes, sometimes in great numbers but their function remains largely unclear. Among suggested functions are prophage maintenance, chromosomal region stabilization, prevention of phage infection, stress response and persister formation (Gerdes et al., 2005; Gerdes and Maisonneuve, 2012; Maisonneuve et al., 2013; Page and Peti, 2016; Wang and Wood, 2011; Wen et al., 2014; Yamaguchi and Inouye, 2011). Interestingly, persister cells are often associated with biofilm mode of bacterial growth (Wang and Wood, 2011). TA systems are classified into six types depending on the nature and action of the antitoxin that can be either a protein or a small antisense RNA (Page and Peti, 2016). In type I systems, the antitoxin is a small antisense RNA that forms RNA duplex with the toxin-encoding mRNA (Brantl, 2012b; Brantl and Jahn, 2015). Most studies are devoted to type II TA systems, in which the protein antitoxin sequestering the toxin is more easily defined than the RNA antitoxin of type I TA (Coray et al., 2017). RNA antitoxins belong to the largest and most extensively studied set of small ncRNA regulators that act by modulating the translation and/or stability of their mRNA targets. Most of type I toxins are small hydrophobic proteins of less than 60 amino acids containing a potential transmembrane domain and charged amino acids at the C-terminus (Fozo et al., 2010). In many cases, they seem to act like phage holins by inducing pores into cell membranes and thus impairing ATP synthesis (Brantl and Jahn, 2015). Replication, transcription and translation are consequently inhibited, which leads to cell death. This Chapter describes the identification of new type I TA modules, associated with *C. difficile* CRISPR arrays. The functionality of discovered TA modules and their possible co-regulation with *C. difficile* CRISPR-Cas system by the general stress response Sigma B and biofilm-related factors are also shown.

Results, obtained in this Chapter correspond to the second general objective of the Thesis: to reveal the way *C. difficile* CRISPR-Cas system expression is regulated and functions in different states of bacterial culture, including its response to stresses.

3.2 Materials and methods

3.2.1 Plasmid and bacterial strain construction and growth conditions

C. difficile and *E. coli* strains and plasmids used in this study are presented in Table S3.1 (Supplementary materials). *C. difficile* strains were grown anaerobically (5 % H₂, 5 % CO₂, and 90 % N₂) in tryptone-yeast extract (TY) (Dupuy and Sonenshein, 1998) or BHI (Difco) media in an anaerobic chamber (Jacomex). When necessary, Cfx (25 µg/ml) and thiamphenicol Tm (15 µg/ml) were added to *C. difficile* cultures. *E. coli* strains were grown in LB broth (Bertani, 1951), and when needed, Amp (100 µg/ml) or Cm (15 µg/ml) was added to the culture medium. The non-antibiotic analog ATc was used for induction of the P_{tet} promoter of pRPF185 vector derivatives in the *C. difficile* (Fagan and Fairweather, 2011). Strains carrying pRPF185 derivatives were generally grown in TY medium in the presence of ATc (250 ng/ml) and Tm (7.5 µg/ml) for 7.5 h. Growth curves were obtained using a GloMax plate reader (Promega).

All routine plasmid constructions were carried out using standard procedures (Sambrook, J., Fritsch, E.F. and Maniatis, 1989). For inducible expression of *C. difficile* genes, we used the pRPF185Δ*gus* vector expression system (Fagan and Fairweather, 2011; Soutourina et al., 2013). The *CD2517.1* gene (-89 to +178 relative to the translational start site), the *CD2907.1* gene (-84 to +223 relative to the translational start site), *CD2517.1*-RCd8 TA region with RCd8 promoter (-306 to +504 relative to the translational start site of *CD2517.1*) and *CD2907.1*-RCd9 TA region with RCd9 promoter (-294 to +456 relative to the translational start site of *CD2907.1*) were amplified by PCR and cloned into StuI and BamHI sites of pRPF185Δ*gus* vector under the control of the ATc-inducible P_{tet} promoter giving pDIA6319, pDIA6195, pDIA6202 and pDIA6196, respectively.

For subcellular localization of toxins we used reverse PCR approach to construct *CD2517.1*-HA and *CD2907.1*-HA-expressing plasmids on the basis of corresponding pRPF185Δ*gus*-derivatives with primers designed to introduce the HA-tag sequence at the C-terminal part of coding toxin regions, directly upstream the stop codon. DNA sequencing was performed to verify plasmid constructs. The resulting derivative pRPF185 plasmids were transformed into the *E. coli* HB101 (RP4) and subsequently mated with *C. difficile* 630Δ*erm* (O'Connor et al., 2006) (Table S3.1). *C. difficile* transconjugants were selected by sub-culturing on BHI agar containing Tm (15 µg/ml), Cs (25 µg/ml) and Cfx (8 µg/ml).

3.2.3 Light microscopy

For light microscopy, bacterial cells were observed at 100x magnification on an Axioskop Zeiss Light Microscope. Cell length was estimated for more than 100 cells for each strain using ImageJ software (Collins, 2007).

3.2.4 RNA extraction, quantitative real-time PCR, and 5'/3'RACE

Total RNA was isolated from *C. difficile* strains grown 7.5 h in TY medium containing Tm (7.5 µg/ml) and ATc (250 ng/ml) as previously described (André et al., 2008). For biofilm samples *C. difficile* 630 Δ *erm* strain was grown for 72 h in TY medium using continuous-flow microfermentor culture system (Ghigo, 2001). 24-h planktonic culture in TY medium was used for comparative analysis. To analyze CRISPR-Cas system expression in high c-di-GMP levels conditions, total RNA was isolated from *C. difficile* 630 Δ *erm* and CDIP634 strains, grown 24 h in BHI medium, supplemented with ATc (250 ng/ml). The cDNA synthesis by reverse transcription and quantitative real-time PCR analysis was performed as previously described (Saujet et al., 2011). In each sample, the relative expression for a gene was calculated relatively to the 16S rRNA gene or *dnaF* gene (CD1305) encoding DNA polymerase III or *ccpA* gene encoding catabolite control protein. The relative change in gene expression was recorded as the ratio of normalized target concentrations ($\Delta\Delta$ Ct) (Livak and Schmittgen, 2001). 5'/3'RACE experiments were performed as previously described (Soutourina et al., 2013).

3.2.5 Subcellular localization of HA-tagged toxins by cell fractionation and Western blotting

The *C. difficile* cultures were inoculated from overnight grown cells in 10 ml of TY medium at OD 600 nm of 0.05, allowed to grow for 3 hours before addition of 250 ng/ml ATc and incubation for 90 min followed by centrifugation and protein extraction. Cell lysis, fractionation, and protein analysis were performed as previously described (Peltier et al., 2015). Coomassie staining was performed for loading and fractionation control. Western blotting was performed as previously described (Boudry et al., 2014) with anti-HA antibodies.

3.2.6 *In silico* screening for potential new TA genes and CRISPR arrays co-localization

The raw sequencing read data of 2,584 *C. difficile* strains were downloaded for this genomic analysis (Boudry et al., 2015; Cairns et al., 2017). For each strain, we realized an assembly with Spades (Bankevich et al., 2012) and an automatic annotation using PROKKA (Seemann, 2014). Then we selected small proteins from 40 to 60 amino acids in length, adjacent to CRISPR arrays and performed an orthology analysis using proteinortho5 (Lechner et al., 2011). Multiple alignment was done using ClustalW (Larkin et al., 2007).

3.2.7 Construction of the *C. difficile* CDIP634 strain, overexpressing *dccA* gene

To insert P_{tet} element upstream the *dccA* (*CD1420*) gene in *C. difficile* 630 Δ *erm*, the *codA* allele exchange method (Cartman et al., 2012) was used. Editing plasmid pDIA6404 was transferred to *C. difficile* 630 Δ *erm* via conjugation. Subsequently, selected transconjugants were twice restreaked onto BHI agar plates, containing Tm, Cs, and Cfx to detect faster growing single-crossover integrants. Then selected colonies were plated onto *C. difficile* CDMM agar supplemented with fluorocytosine (50 μ g/ml) to identify second cross-over events. Subsequently, fluorocytosine-resistant clones were analyzed by PCR. The resulting PCR fragments have been sequenced to confirm the P_{tet} element insertion.

3.2.8 Plasmid conjugation efficiency assays

To evaluate conjugation efficiency, PAM-protospacer carrying conjugative plasmids were transformed into the *E. coli* HB101 (RP4) strain and transferred to *C. difficile* 630 Δ *erm* and CDIP634 strains by conjugation. Subsequently, conjugation mixtures were plated on BHI agar, containing ATc (250 ng/ml). The ratio of *C. difficile* transconjugants was counted by subculturing conjugation mixture on BHI agar supplemented with Tm, Cs, and Cfx and comparing the number of CFU obtained after plating serial dilutions on BHI agar plates containing Cfx only.

3.3 Results

3.3.1 Co-regulation of *C. difficile* CRISPR-Cas system and type I toxin-antitoxin systems adjacent to CRISPR arrays

3.3.1.1 Identification of toxin-antitoxin system candidates in *C. difficile* genome

We have revisited the previously reported deep sequencing data (Soutourina et al., 2013) and observed an unusual transcriptional unit organization in the close proximity of CRISPR loci in the genome of *C. difficile* strain 630 Δ *erm* (Figure S3.1 in Supplementary materials). The presence of several overlapping transcripts was detected by comparison of Tobacco Acid Pyrophosphatase treated (TAP+) and non-treated (TAP-) samples for transcriptional start site (TSS) mapping. This analysis combined with the RNA-seq data for whole transcript coverage revealed that the majority of crRNA loci are associated with potential antisense RNAs of genes encoding small proteins of unknown function (Figure 3.1). A more detailed analysis of the nature of the overlapping convergent transcripts allowed to identify candidates for six type I TA systems that co-localized with CRISPR 3/4, CRISPR 6, CRISPR 7, CRISPR 11, CRISPR 12 and CRISPR 16/15 arrays in *C. difficile* (Figure 3.1) (Fozo et al., 2010). An additional pair of antisense RNA and small protein gene near CRISPR 9 array had divergent sequence without common type I TA features.

Interestingly, three potential TA modules together with associated CRISPR arrays are located within prophage regions (in red in Figure 3.1). The pathogenicity-island location of type I TA modules has been reported in *Staphylococcus aureus* (Pinel-Marie et al., 2014; Sayed et al., 2012a). The CRISPR 3/4 array, the associated potential toxin gene *CD0956.2* and the antitoxin gene are located within the phiCD630-1 prophage region while the identical CRISPR 16/15 array and the potential TA module containing *CD2907.1* toxin and antitoxin genes are located within the phiCD630-2 prophage region (Figure 3.1). Similar to the *txpA/RatA* type I TA module in *Bacillus subtilis* (Silvaggi et al., 2005), the *CD1233.1/SQ808* pair is located within the *skin* element of *C. difficile* strain 630 (Serrano et al., 2016), yet there is no sequence homology between the two loci. This *CD1233.1/SQ808* pair is located near the CRISPR 6 array in the *C. difficile skin* element.

We have chosen three representative type I TA modules for further detailed analysis. The RCd8-*CD2517.1* module is located near the CRISPR 12 array, which is associated with a partial *cas* operon. The RCd9-*CD2907.1* and RCd10-*CD0956.2*

modules are located near the CRISPR 16/15 and CRISPR 3/4 arrays. They lie respectively within the phiCD630-1 and phiCD630-2 prophage regions, which have identical sequences and are thus indistinguishable from each other through gene expression analysis. These two highly similar prophages phiCD630-1 (1088001-1143874) and phiCD630-2 (3377033-3434358) of 55.9 and 57.3 kb in length, respectively, are located in an inverted orientation on different replicohores of the *C. difficile* chromosome. The regions encoding TA modules and CRISPR arrays are identical. We assigned the names RCd8 (previously named *SQI781*), RCd9 (previously named CD630_n01000) and RCd10 (previously named CD630_n00370) to the putative antitoxin RNAs (Figure 3.1).

We first mapped by 5'/3'RACE analysis the transcriptional start and termination sites of the genes corresponding to the potential toxin and the antitoxin RNAs for selected loci. Figure 3.2 shows the chromosomal organization of these genes and the position of 5'- and 3'-ends of overlapping transcripts identified by 5'/3'RACE (Figure S3.2 in Supplementary materials). The alignment of the TA genomic regions revealed the presence of conserved sequences upstream of the TSS for both the putative toxin and antitoxin genes and allowed to identify the consensus elements for Sigma A-dependent promoters upstream of their TSS (indicated in blue and red in Figure S4.2 in Supplementary materials). Moreover, the consensus sequence promoters recognized by the alternative Sigma factor of the general stress response, Sigma B, could be identified upstream of the TSS of both the potential antitoxin and toxin genes (indicated in green in Figure S3.2 Supplementary materials).

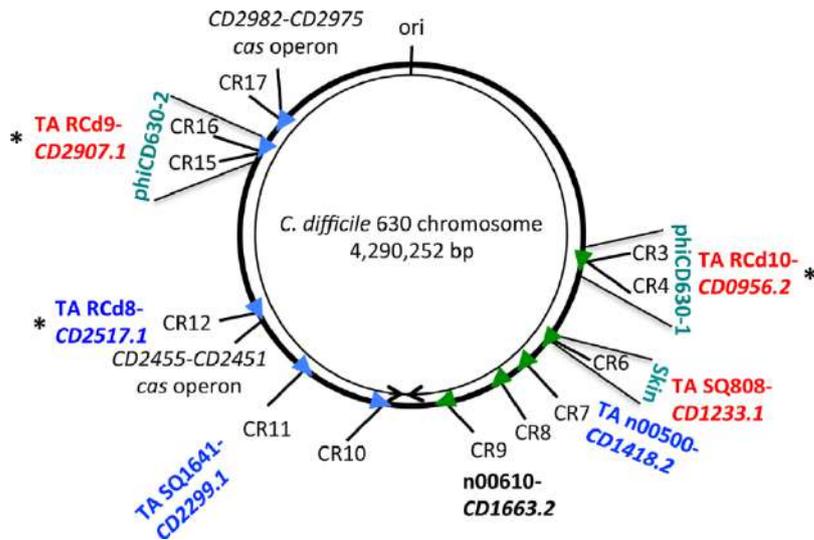


Figure 3.1. Genomic map of potential type I TA loci in association with CRISPR arrays in *Clostridium difficile* strain 630 Δ erm. Schematic view of the genomic location of expressed CRISPR arrays in strain 630 Δ erm. CRISPR arrays are numbered according to CRISPRdb database (Grissa et al., 2007). Arrowheads indicate the array position and the transcriptional orientation. The location of the associated TA modules, the *cas* operons, the prophage regions and the replication origin (ori) are indicated. The right and left replichores are shown by arrows. The n00610 antisense RNA overlaps the *CDI663.2* gene, which encodes a small protein with a divergent sequence associated with CRISPR 9 array. The CRISPR-associated TA modules within prophage regions are RCd9-*CD2907.1*, RCd10-*CD0956.2* and SQ808-*CD1233.1*. “*” indicates the three TA modules that were selected for detailed analysis. Adapted from (Maikova et al., 2018a) with permission.

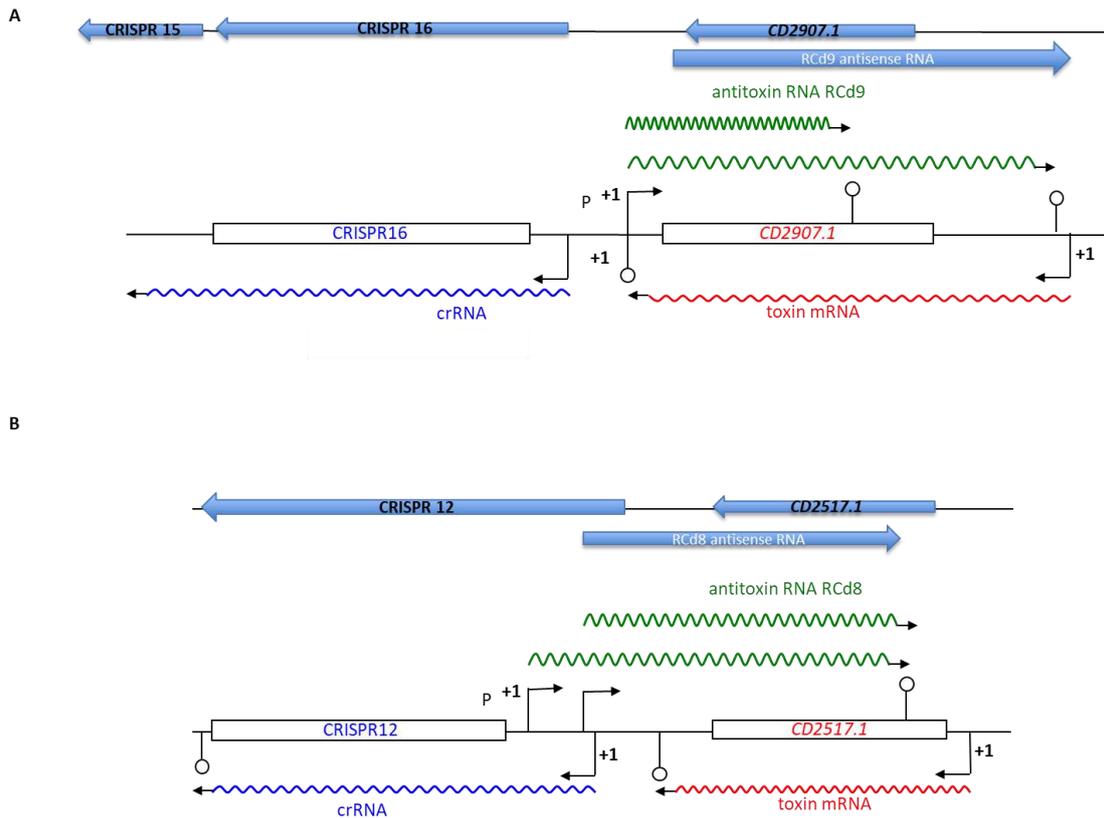


Figure 3.2. Schematic representation of potential type I TA locus in *C. difficile* chromosome. **A** – CRISPR16/15 (CRISPR3/4) regions, **B** – CRISPR12 region. The transcriptional start sites “+1” for sense and antisense transcripts identified by 5'/3'RACE and deep sequencing are indicated by broken arrows. The position of transcriptional terminators is indicated by loops. Overlapping transcripts are drawn in red, green and blue for toxin, antitoxin and crRNAs and the direction of transcription is indicated by arrowheads.

3.3.1.2 Functionality of toxin-antitoxin systems in *C. difficile*

Type I toxins are generally small hydrophobic proteins of less than 60 amino acids containing a potential transmembrane domain and charged amino acids at the C-terminus (Fozo et al., 2010). The alignment of proteins from the potential TA modules encoded in the proximity of CRISPR arrays revealed that these small proteins have all characteristic features of type I toxins. Indeed, as shown in Figure 3.3A, the potential toxic proteins are from 50 to 53 amino acids in length, carry a conserved hydrophobic region at their N-terminal part and a lysine-rich, positively charged region at their C-terminal part in agreement with the hydrophobicity profile predictions by Kyte and Doolittle algorithm (data not shown). Transmembrane domain location in N-terminal moiety was predicted by TMHMM program (data not shown). To experimentally identify the expression and localization of these small proteins in *C. difficile* we constructed plasmids expressing

under inducible P_{tet} promoter either *CD2517.1* or *CD2907.1/CD0956.2* fused with a HA tag at the C-terminus (Table S3.1 Supplementary materials). By Western blotting with anti-HA antibodies, no signal was detected for whole cell extracts from control strains expressing untagged proteins while a specific signal was detected for strains expressing HA-tagged proteins (Figure 3.3B). To precise the subcellular localization of these proteins we then performed cell fractionation and examined supernatant, cell wall, membrane and cytosolic fractions by Western blotting. As shown in Figure 3.3B, HA-tagged *CD2517.1* and *CD2907.1* (*CD0956.2*) were only detected in the membrane fractions of *C. difficile* cell extracts suggesting the association of these small proteins with the cell membrane in *C. difficile*.

To show the toxic nature of these small proteins, we analyzed the effect of their overexpression on the growth of *C. difficile* cells in liquid and solid media. HA-tagged proteins *CD2517.1* and *CD2907.1/CD0956.2* conserved their toxic activity on cell growth when overexpressed from plasmids used for determination of their subcellular localization by Western blotting (Figure S3.3 in Supplementary materials, Figure 3.3B). This result suggests that despite the presence of HA-tag these small proteins remain active for cell growth inhibition.

We then generated plasmids allowing either inducible overexpression of an untagged version of one of the small, potentially toxic proteins or simultaneous expression of both the potential toxin and the antisense RNA for the TA modules near the CRISPR 12 and CRISPR 16/15 (CRISPR 3/4) arrays. For this purpose, we cloned either the small protein-coding region with its ribosome-binding site (RBS) (*CD2517.1* or *CD2907.1/CD0956.2*) under the control of the inducible P_{tet} promoter (pT) or the entire potential TA module (pTA). pTA constructs allow both the inducible overexpression of the putative toxin under the control of the P_{tet} promoter and the expression of the antisense RNA from its own strong promoter (Figure 3.4A). *C. difficile* strain 630 Δ *erm* carrying an empty vector (p) was used as a control. No growth difference was observed for any of the three strains on BHI plates in the absence of ATc inducer for both potential TA modules (Figure 3.4A, 3.5A). By contrast, a dramatic growth defect was observed on BHI plates in the presence of ATc inducer for the strain overexpressing the genes *CD2517.1* or *CD2907.1/CD0956.2* (Figure 3.4 A, 3.5A). Co-expression of these potential toxins with the associated RNA antitoxins led to the full or partial reversion of the growth defect for both TA modules (Figure 3.4A, 3.5A).

The overexpression of toxins from selected TA modules also induced rapid growth arrest in liquid culture. As shown in Figure 3.4B for CD2517.1-RCd8 TA module, the addition of ATc inducer after 3h of exponential growth led to rapid growth arrest for the strain carrying the pT plasmid but allowed near normal growth of the *C. difficile* 630 Δ erm strain carrying pTA. Similar deleterious growth effects were observed for the strain carrying the pT plasmid when strains pre-grown overnight in the absence of inducer and then diluted in an ATc-containing medium were allowed to grow for 24 h in an automatic plate reader (Figure 3.4C). For the CD2907.1-RCd9/CD0956.2-RCd10 TA module, we observed only a partial reversion of the growth defect in liquid culture associated with the toxin gene expression when both toxin and antitoxin were co-expressed on pTA plasmid (Figure 3.5B). This partial restoration of growth could be due to an unbalance in the relative level of toxin and antitoxin expression.

Toxins from TA modules in *B. subtilis* and *Enterococcus faecalis* have been reported to affect cell envelope biosynthesis, nucleoid condensation, cell division and chromosome segregation (Jahn et al., 2015; Patel and Weaver, 2006). To assess whether the changes in cell morphology could be induced by toxin overexpression in *C. difficile*, we analyzed by light microscopy liquid cultures of strain 630 Δ erm carrying the vector, pT or pTA 1 h after ATc addition. For both TA modules (CD2517.1 and CD2907.1/CD0956.2), the overexpression of the toxins in strain 630/pT led to a significant increase in cell length for about 9% and 5.4% of the cells, respectively. The length of these cells was above the value of 630/p mean length with 2 standard deviations (10.5 μ m) (Figure 3.5D and Figures S3.4 and S3.6 in Supplementary materials). For control strain 630/p the length of only 1.7% of cells exceeded this value. Co-expression of the entire TA module (pTA) led to a partial reversion of this phenotype to the control culture morphology.

Altogether this data demonstrate that functional type I TA modules are present in the proximity of CRISPR arrays in *C. difficile*.

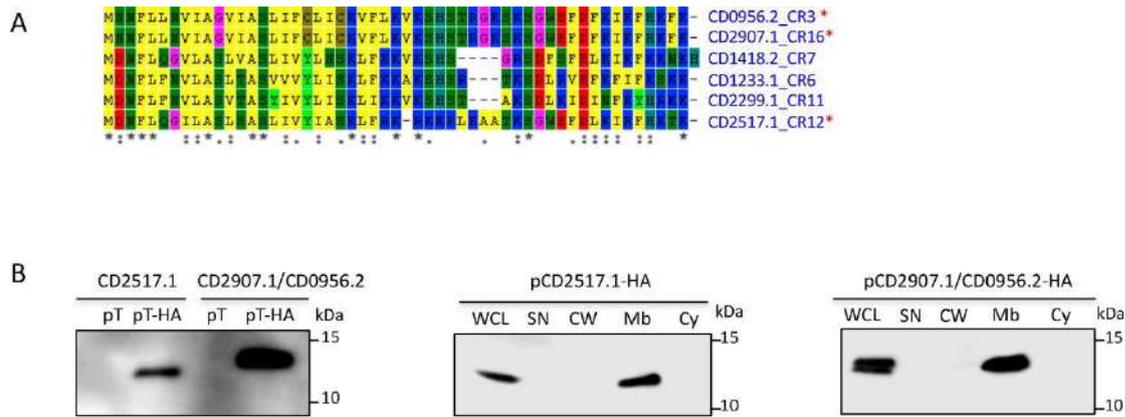


Figure 3.3. Potential type I toxin proteins alignment and analysis. **A** – Proteins alignment using ClustalW. “*” on the right indicates toxins from three TA modules selected for detailed analysis. “*” at the bottom indicates conserved residues. **B** – Western-blot detection and localization of HA-tagged small proteins in the membrane fraction of *C. difficile* cell extracts. WCL: whole cell lysate; SN: supernatant; CW: cell wall; Mb: membrane; Cy: cytosolic fraction. Immunoblotting with anti-HA antibodies detected a major polypeptide of ~10 kDa in whole cell lysates of the strain carrying P_{ter} -T (CD2517.1 or CD2907.1/CD0956.2)-HA (pT-HA) construct grown in the presence of the 250 ng/ml ATc inducer but not in extracts of strains expressing non-tagged toxins (pT) (left panel). The culture of strains carrying P_{ter} -T-HA plasmids induced with 250 ng/ml ATc was fractionated into cell wall (CW), membrane (Mb) and cytosolic (Cy) compartments and immunoblotted with anti-HA antibodies (middle and right panels). Proteins were separated on 12% Bis-Tris polyacrylamide gels in MES buffer. Adapted from (Maikova et al., 2018a) with permission.

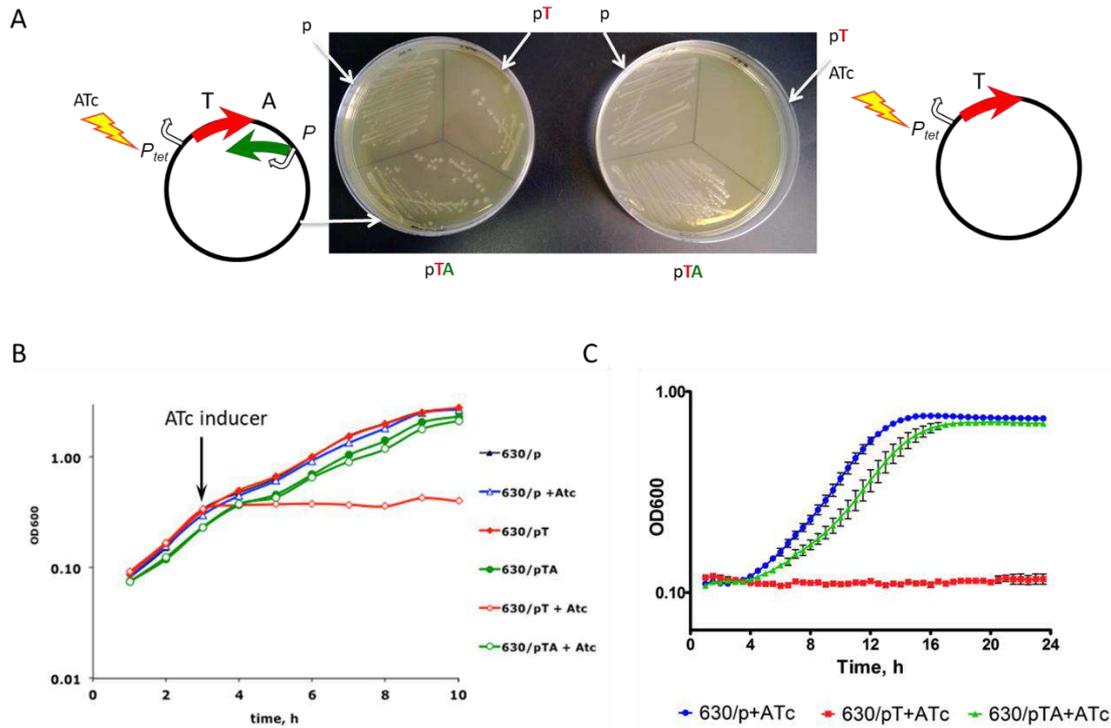


Figure 3.4. Effect of inducible toxin and TA overexpression for CD2517.1-RCd8 TA module near CRISPR 12 on growth in solid (A) and liquid medium (B–C). **A** – growth phenotype of *C. difficile* strains CDIP369 (630/p), CDIP357 (630/pT) and CDIP332 (630/pTA) on BHI agar plates supplemented with Tm alone (on the left) or with the addition of 500 ng/ml of ATc inducer (on the right) after 24 h of incubation at 37°C. Schematic representations of pT and pTA constructs are shown. The 630 Δ *erm* strain carrying an empty vector (p) is used as a control. **B** – growth of 630/p strain (triangles), 630/pT strain (diamond) and 630/pTA strain (circle) in TY medium at 37°C in the presence (open symbols) or absence (closed symbols) of 250 ng/ml ATc. The time point of ATc addition is indicated by an arrow. **C** – growth curves for 630/p strain, 630/pT strain and 630/pTA strain in TY medium at 37°C in the presence of 250 ng/ml ATc using a GloMax plate reader (Promega). The mean values and standard deviations are shown for three independent experiments.

Adapted from (Maikova et al., 2018a) with permission.

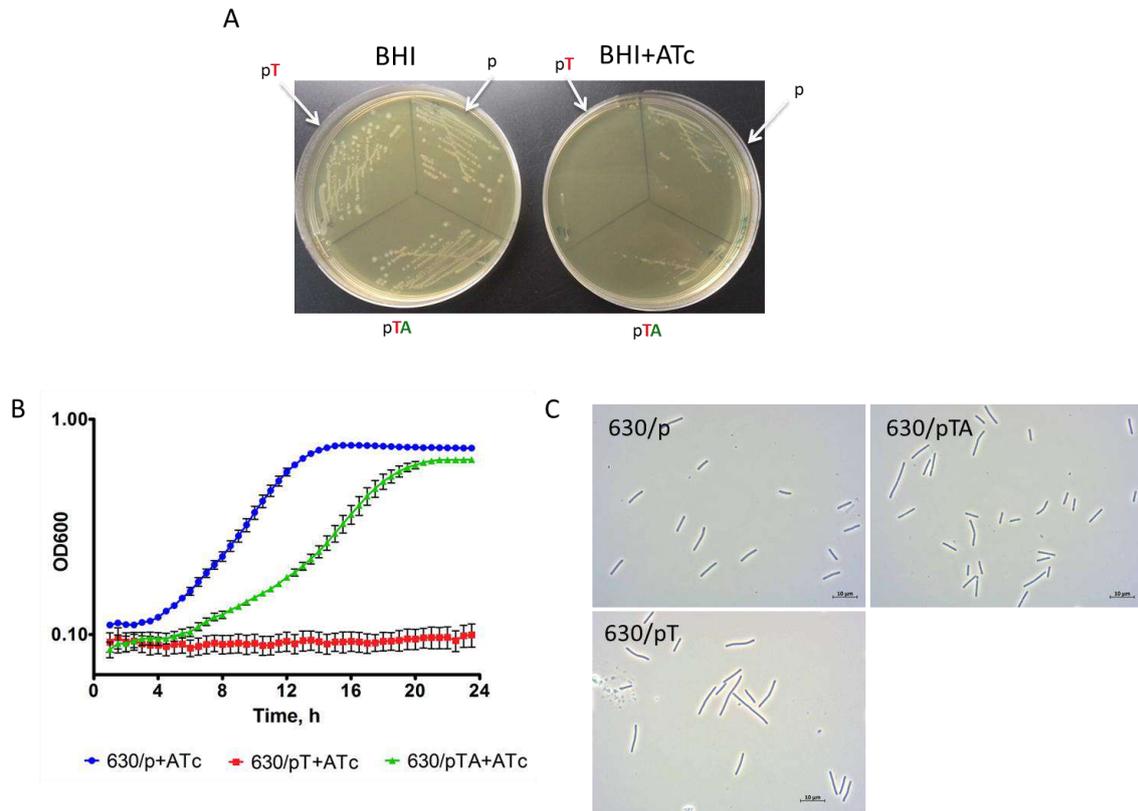


Figure 3.5. Effect of inducible toxin and TA overexpression for CD2907.1-RCd9/CD0956.2-RCd10 TA module near CRISPR 16/15 (CRISPR 3/4) on growth in solid (A) and liquid (B–D) medium. **A** – growth of *C. difficile* strains CDIP369 (630/p), CDIP317 (630/pT) and CDIP319 (630/pTA) on BHI agar plates supplemented with Tm alone (on the left) or with the addition of 500 ng/ml of ATc inducer (on the right) after 24 h of incubation at 37°C. Under inducing conditions 630/pT strain overexpresses CD2907.1 toxin and 630/pTA strain overexpresses entire TA module. The 630 strain carrying an empty vector (p) is used as a control. **B** – growth of 630/p strain, 630/pT strain and 630/pTA strain in TY medium at 37°C in the presence of 250 ng/ml ATc using a GloMax plate reader (Promega). The mean values and standard deviations are shown for three independent experiments. **C** – selected images from light microscopy observation of 630/p, 630/pT and 630/pTA strains grown in TY medium at 37°C after 1 h of 250 ng/ml ATc addition. Adapted from (Maikova et al., 2018a) with permission.

3.3.1.3 Expression analysis of TA and CRISPR-Cas systems

We wondered whether the chromosomal co-localisation of CRISPR arrays and TA modules would imply the possible connection between these systems. As mentioned above, the alignment of CRISPR-associated TA module sequences strongly suggested the presence of both Sigma-A-dependent and Sigma-B-dependent promoters upstream of the TSS of the toxin and antitoxin genes for the 6 TA modules (Figure S3.2 in Supplementary materials). The crucial role of the alternative Sigma B factor has been recently demonstrated for the adaptive strategies of *C. difficile* inside the host (Kint et al., 2017). We re-examined the transcriptome data for the *sigB* mutant as compared to the parental strain and observed up to 5-fold decrease in the expression of the entire gene sets for both the partial and complete *cas* operons (*CD2455* and *CD2982*) of type I-B *C. difficile* CRISPR-Cas system (Table 3.1). qRT-PCR analysis validated these transcriptome data (Table 3.1). In accordance, the search for Sigma-B-dependent promoter sequences revealed the presence of consensus elements GTTTTAA-N12-GGGATTT and TTATAA-N12-GGGTTAA upstream of TSS for *cas* gene operons *CD2455* and *CD2982*, respectively. These promoter sequences are characterized by the presence of a conserved -10 promoter element associated and a less conserved -35 promoter element. Such a promoter structure suggests the possible implication of other regulatory components controlling these operons together with the Sigma B factor. The high sequence conservation among direct repeats within multiple CRISPR arrays suggests that the same set of Cas proteins processes all expressed pre-crRNA in *C. difficile* strains (Boudry et al., 2015). Thus, the induction of *cas* genes under stress conditions would allow the overall activation of CRISPR-Cas defense mechanisms. Transcriptome analysis of the *sigB* mutant also revealed differential expression of several newly identified TA genes and associated CRISPR arrays (Table 3.1). To confirm these data, we performed qRT-PCR analysis for selected TA gene pairs and CRISPR arrays (Table 3.1). In accordance with transcriptome data, we confirmed by qRT-PCR the down-regulation of several CRISPR-associated TA genes in the *sigB* mutant strain as compared to the parental strain even without stress exposure (Table 3.1).

The induction of CRISPR-Cas mediated defense capacities within biofilm community or more generally within the gut microbiota, which includes phages, could be important for bacterial survival under conditions promoting gene transfer. In *E. coli*, type I toxin *ralR* gene expression is induced during growth in biofilms (Domka et al., 2007).

We thus compared the expression of selected CRISPR-associated TA modules and CRISPR-Cas systems within biofilm and planktonic cultures and observed a strong, up to 20-fold, induction of expression of selected genes (Table 3.1). Overall these results suggest that the *cas* operons and the CRISPR arrays could be co-regulated with associated type I TA systems by stress- and biofilm-related factors.

Table 3.1. Differential expression of TA and CRISPR-Cas systems revealed by transcriptome and/or qRT-PCR analysis. Gene names and functions correspond to those indicated in the MaGe database Clostriscope (<https://www.genoscope.cns.fr>).

Gene ID	Function	Ratio <i>sigB</i> /630Δ <i>erm</i> Microarray ^a	Ratio <i>sigB</i> /630Δ <i>erm</i> qRT-PCR	Ratio biofilm/plankton qRT-PCR
CD2982 ^b	CRISPR-associated Cas6 family protein	0.19	0.22	14.7
CD2981	CRISPR-associated protein, CXXC-CXXC	0.21		
CD2980	CRISPR-associated autoregulator DevR family protein	0.26		
CD2979	CRISPR-associated Cas5 family protein	0.25		
CD2978	CRISPR-associated Cas3 family helicase	0.30		
CD2977	CRISPR-associated Cas4 family protein	0.33		
CD2976	CRISPR-associated Cas1 family protein	0.37		
CD2975	CRISPR-associated Cas2 family protein	0.37		
CD2455 ^c	CRISPR-associated protein	0.55	0.61	9.4
CD2454	Conserved hypothetical protein	0.55		
CD2453	CRISPR-associated negative autoregulator	0.47		
CD2452	CRISPR-associated protein	0.53		
CD1233.1	Toxin of TA associated with CRISPR 6	0.51		
CD2517.1	Toxin of TA associated with CRISPR 12	0.25	0.42	26.0
RCd8	Antitoxin of TA associated with CRISPR 12		0.67	7.3
CD630_n00860	CRISPR12			7.7
CD2907.1	Toxin of TA associated with CRISPR 16/15			1.8
RCd9/RCd10	Antitoxin of TA associated with CRISPR 16/15/CRISPR 3/4	0.5	0.42	2.3
CD630_n00990	CRISPR 16/15		0.54	9.8

^aA gene was considered as differentially expressed between the strain 630Δ*erm* and the *sigB* mutant when the P-value is <0.05.

^bFirst gene of the complete *cas* operon CD2982-CD2975.

^cFirst gene of the partial *cas* operon CD2455-CD2452. SQ1781 corresponds to RCd8, CD630_n01000 to RCd9 and CD630_n00370 to RCd10.

Adapted from (Maikova et al., 2018a) with permission.

3.3.1.4 Genomic analysis of TA and CRISPR arrays co-localization

We analyzed the extent of co-localization of potential type I TA with CRISPR arrays in available *C. difficile* sequences. From more than 2,500 *C. difficile* genome sequences assembled and automatically annotated, we first found that 98% contain CRISPR arrays (from 1 to 30). In these CRISPR-containing strains, we then searched for the presence, immediately adjacent to CRISPR loci, of open reading frames from 40 to 60 amino acids, as one of the characteristic features of type I toxins is their small size. This search resulted in about 7,000 hits. The CRISPR-associated small proteins were only absent in 67 genomes of which 58 lacked *cas* gene homologs. Then, an orthology analysis identified 16 proteins present each in more than 25 strains. Figure 3.6 shows an alignment of these 16 representative small proteins adjacent to CRISPR arrays combined in 5 major groups (A-E) according to their homology.

The three small proteins characterized in this study (CD2907.1, CD0956.2 and CD2517.1) belong to group A. This group is largely distributed in *C. difficile* as it is present in two-third of the analyzed strains (Figure 3.6). CRISPR 16/15 and CRISPR 3/4 associated toxins belong to the most represented subgroup, A1, found in 63% of strains, CRISPR 12 associated toxin belongs to subgroup A2, that is present in 20% of the analyzed strains. Other CRISPR-associated toxins of strain 630 are represented in less extent within the same group. Finally, as two of the characterized toxins are located within prophage regions in strain 630, we wondered whether prophage localization could be a common feature of CRISPR-associated small proteins. In 13 from 22 known *C. difficile* phages, we found potential toxins all belonging to the group A that could be part of TA modules. However, the co-localization with CRISPR arrays is detected only in the phi027 prophage of the R20291 strain.

To provide an experimental confirmation of potential TA and CRISPR arrays co-expression in another *C. difficile* strain, we have looked at the RNA-seq data of the epidemic strain R20291 (Maldarelli et al., 2016). In this strain, we detected three co-localized CRISPR and TA pairs (Figure S3.5A in Supplementary materials). One pair was intact (TA and CRISPR), while the two others have a mutation in the toxin genes and only antitoxin was detected (A and CRISPR). We confirmed their co-expression in the published R20291 RNA-seq data using the COV2HTML software for visualization (Monot et al., 2014) (Figure S3.5B in Supplementary materials).

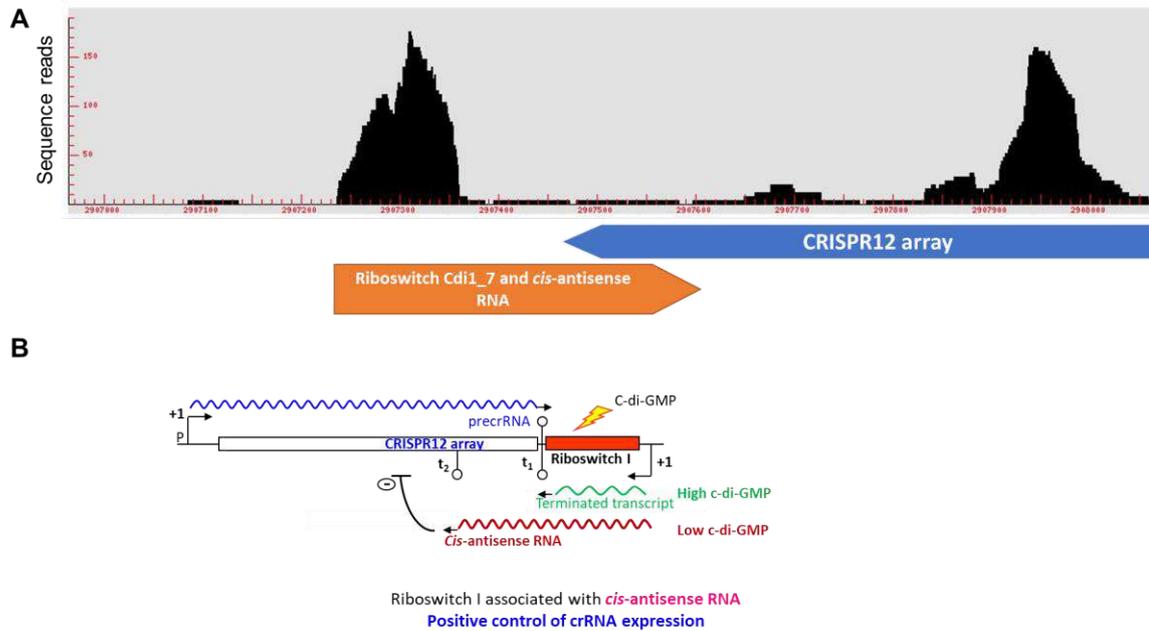


Figure 3.7. c-di-GMP-I riboswitch and a cis-antisense RNA adjacent to *C. difficile* 630 Δ erm CRISPR12 array. **A** – RNAseq of the CRISPR12 and cdi1_7 riboswitch region, **B** – general scheme of possible CRISPR12 array, the cis-antisense RNA, and cdi1_7 riboswitch interaction. “P” indicates the CRISPR12 array promoter, “+1” points designate transcriptional start sites, “t₁” and “t₂” – transcription termination sites of the cis-antisense RNA.

3.3.2.2 Role of high c-di-GMP intracellular levels on *C. difficile* 630 Δ erm CRISPR12 array functionality

A signal molecule c-di-GMP is synthesized from two GTP molecules by diguanylate cyclases (Schmidt et al., 2005). To study the possible role of c-di-GMP levels in *C. difficile* 630 Δ erm CRISPR-Cas system regulation, we constructed a CDIP634 strain with a chromosomal *dccA* (*CD1420*) gene, encoding a *C. difficile* diguanylate cyclase (Purcell et al., 2012), under control of the inducible P_{tet} promoter (Figure 3.8). The *codA* allele exchange strategy was used to construct this strain (Cartman et al., 2012). When the CDIP634 strain grows in medium supplemented with ATc to induce the P_{tet}, intracellular c-di-GMP levels significantly increase due to the abundance of active diguanylate cyclase inside the cells. After the ATc induction the expression of *dccA* gene, detected by qRT-PCR, increased in about 200 times CDIP634 comparing to the wild type strain.

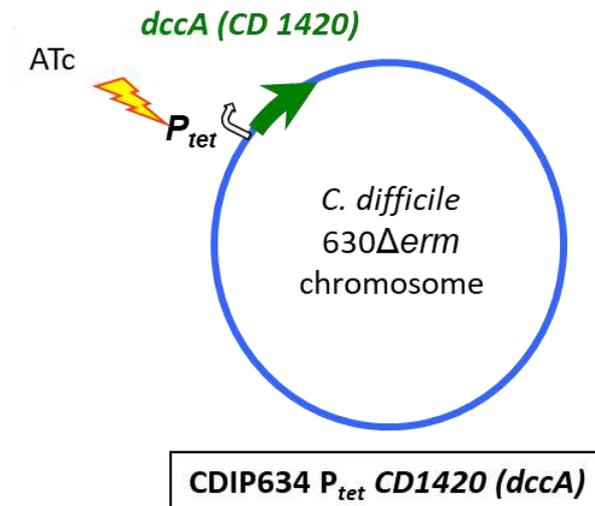


Figure 3.8. *C. difficile* CDIP634 strain, carrying *dccA* gene under the control of the inducible P_{tet} promoter.

To investigate the role of c-di-GMP on the *C. difficile* 630 Δ *erm* CRISPR12 array functionality, we performed plasmid interference assays with CDIP634 and 630 Δ *erm* strains. Accordingly, for these experiments, plasmids containing protospacers corresponding to spacers from *C. difficile* 630 Δ *erm* CRISPR 12 array and flanked by functional CCA PAM on their 5'-end (see 2.3.2 in Results) were used. All plasmid interference assays were carried out using medium supplemented with ATc for the P_{tet} induction. An empty pRPF185 Δ *gus* vector was used as a control. Conjugation efficiency results are presented in Figure 3.9. The difference between CDIP634 and 630 Δ *erm* strains was clear only in the case of the plasmid, carrying protospacer, corresponding to CRISPR12 spacer1. According to the previous work, this spacer is the most actively expressed in this array (Boudry et al., 2015).

Obtained results indicate that the *cdi1_7* c-di-GMP-I riboswitch could regulate *C. difficile* 630 Δ *erm* CRISPR12 array function in the presence of high c-di-GMP intracellular levels.

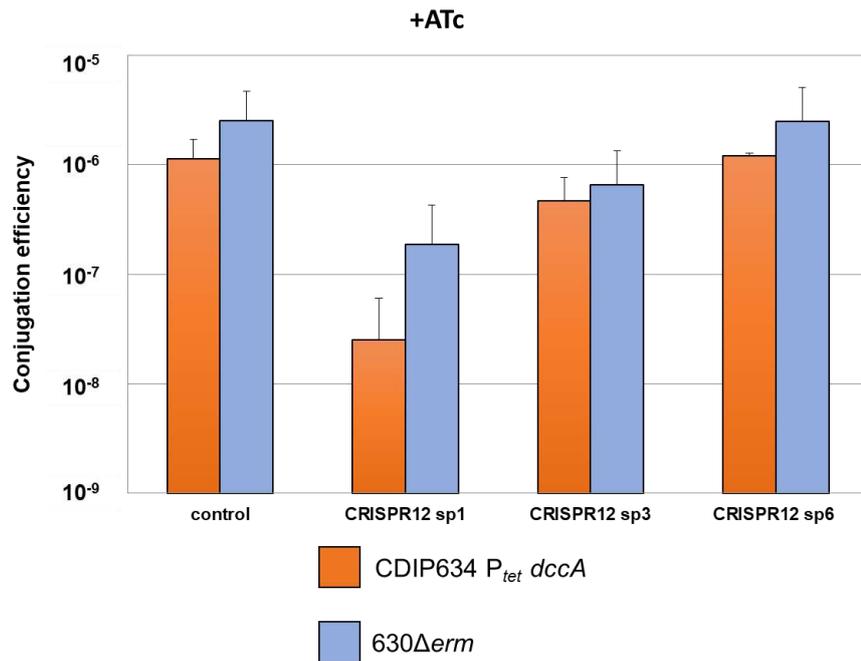


Figure 3.9. Plasmid conjugation efficiency in *C. difficile* 630 Δerm and *C. difficile* CDIP634 strains after their growth in ATc-containing medium. The mean values and standard deviations are shown for two independent experiments.

3.3.2.3 *C. difficile* 630 Δerm CRISPR-Cas system expression under high c-di-GMP intracellular levels

To study the possible effect of high c-di-GMP intracellular levels on *C. difficile* 630 Δerm expression, we compared levels of CRISPR-Cas system components RNAs in *C. difficile* 630 Δerm and CDIP634 strains after their growth in ATc-containing medium. The qRT-PCR analysis revealed that expression of both *cas*-operons and CRISPR6, 12 and 16/15 arrays increased in high c-di-GMP level conditions (Figure 3.10).

Altogether, these results demonstrate the possible positive regulation of *C. difficile* 630 Δerm CRISPR-Cas system by high c-di-GMP levels.

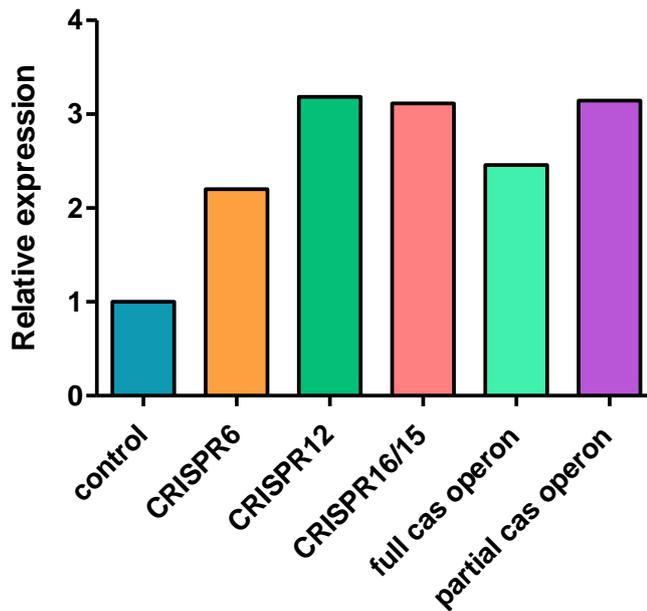


Figure 3.10. qRT-PCR analysis of the *C. difficile* 630 Δ *erm* CRISPR-Cas system expression in high c-di-GMP levels conditions. The expression levels of arrays and operons (in CDIP634 strain) are shown compared to the control expression levels (630 Δ *erm* strain).

3.4 Discussion

3.4.1 Discovery of new type I TA modules, associated and co-regulated with *C. difficile* CRISPR-Cas system

In this Chapter we report the first identification of functional type I TA modules in *C. difficile* 630 chromosome. Deep-sequencing and 5'/3'RACE revealed the presence of overlapping transcripts for type I toxin gene and associated RNA antitoxin in several chromosomal loci. Comparison of the newly identified type I TA systems in *C. difficile* with previously studied TA systems in other bacteria revealed no sequence homology for small toxin proteins. However, we observed a conservation of their membrane association and the presence of charged amino acids in the C-terminal part (Fozo et al., 2010; Jahn et al., 2015; Sayed et al., 2012b). The inducible overexpression of toxin genes strongly impaired the *C. difficile* growth while co-expression of associated antitoxin RNA prevented this growth defect.

The present study demonstrates the unique co-localization of the type I TA modules with CRISPR arrays in the bacterial chromosome. Our large genome analysis revealed that this physical genomic link between TA pairs and CRISPR arrays can be extended to

the majority of sequenced *C. difficile* strains. Initially, TA systems were shown to be important for maintenance of plasmids through a post-segregation killing mechanism (Hayes, 2003; Page and Peti, 2016; Shen et al., 2016). The role of numerous chromosomal TA systems remains largely enigmatic, even though their possible implication in the stabilization of chromosomal regions has been emphasized. For example, a TA module has been shown to promote the maintenance of an integrative conjugative element STX in *V. cholerae* (Wozniak et al., 2009).

The co-localization of functional type I TA systems with CRISPR arrays that we observed on *C. difficile* chromosome has never been reported for any other bacterial genome. Nevertheless, several type I TA systems are located within prophage or prophage-like regions both in *C. difficile* and *B. subtilis* (Brantl and Jahn, 2015; Durand et al., 2012b, 2012a; Jahn et al., 2012; Müller et al., 2016; Silvaggi et al., 2005), even though *B. subtilis* genome lacks CRISPR arrays (Barrangou et al., 2007). The type I TA modules are present within the *skin* element, which is excised from the chromosome during sporulation, in *B. subtilis* and *C. difficile*. Similarly to *B. subtilis* systems, a role in stabilization of these chromosomal regions can be hypothesized for TA systems in *C. difficile*, which carries a high proportion of stable MGE in its genome (Sebahia et al., 2006).

Based on the observations that prophage-located CRISPR arrays are often associated with type I TA modules in *C. difficile*, an interesting evolutionary aspect of the *C. difficile* CRISPR-Cas system can be underlined. Indeed, the TA systems could contribute to the stabilization of the chromosomal regions carrying CRISPR-Cas systems after acquisition of large defense capacities associated with CRISPR arrays. We can hypothesize that TA modules are implicated in maintaining of CRISPR regions, but also in stress response, prophage stability, sporulation control, biofilm formation and other community-associated processes important for this pathogen.

Possible connections between CRISPR and TA systems were highlighted by several recent studies focusing on type II TA (Koonin and Zhang, 2016). Bioinformatics search identified the so-called “defense islands” in bacteria associating immunity and cell death or dormancy functions including CRISPR and type II TA systems (Makarova et al., 2011c, 2013). The original features of *C. difficile* are that type I toxins were not found in “defense islands”. The role of this functional coupling 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 invaders. Thus, our findings are in line with a recently emerged concept on a functional coupling between distinct defense strategies provided by immunity and cell dormancy systems in prokaryotes (Koonin and Zhang, 2016).

The co-regulation of CRISPR-Cas and newly described type I TA systems by the stress-specific factor, Sigma B and the biofilm-related stimuli further suggests the possible connections between these systems in *C. difficile*. Our findings emphasize additional original features of the recently characterized *C. difficile* CRISPR-Cas system including the link with community-behavior control, stress response and type I TA systems. Such control of CRISPR-Cas expression in response to stress-related factors could be relevant for the *C. difficile* infection cycle.

Together with alternative roles of CRISPR-Cas in the control of bacterial physiology and pathogenesis beyond the role in defense against foreign invaders (Richter et al., 2012; Sampson and Weiss, 2013), stimuli and mechanisms controlling CRISPR-Cas system expression just start to be uncovered. However, multiple connections between TA systems in bacteria and stress response have been reported (Gerdes et al., 2005; Wang and Wood, 2011). This Chapter provides new data on the co-regulation of type I TA and CRISPR-Cas systems by the general stress response Sigma B factor in *C. difficile*. Sigma B likely plays a crucial role in the responses to stresses encountered by this pathogen inside the host. Interestingly, the MazEF type II TA module is encoded within the *sigB* operon in *S. aureus* with possible regulatory connections (Donegan and Cheung, 2009). Various environmental stimuli including metabolic and genotoxic stresses induce toxin-antitoxin gene expression of type I TA systems in *B. subtilis*, *E. coli* and *S. aureus* (Brantl and Jahn, 2015, 2016; Jahn et al., 2012; Kawano, 2012; Kawano et al., 2007; Sayed et al., 2012b). In a multistress responsive type I TA system *bsrE/SR5* from *B. subtilis*, the control of antitoxin RNA SR5 by iron limitation stress has been reported to be dependent on the alternative Sigma B factor (Müller et al., 2016).

Key roles of both type II and type I TA systems have been suggested in bacterial pathogens where they can contribute to virulence, fitness inside the host, persistence, intracellular lifestyle, stress response and biofilm formation (Georgiades and Raoult, 2011; Lobato-Márquez et al., 2015; Wen et al., 2014; Yamaguchi and Inouye, 2011). More generally, biofilm formation process has been associated in previous studies with bacterial TA systems (Wen et al., 2014). Recent data suggest that the TxpA type I toxin from the *skin* element acts to eliminate defective cells and preserve symmetry in *B.*

subtilis biofilms (Bloom-Ackermann et al., 2016). We show here that both the expression of the CRISPR-Cas and the associated TA systems are induced in biofilm conditions in *C. difficile*. In general, TA systems including well-documented type II TA exist in surprisingly high numbers in all prokaryotes but clostridial TA modules have been only poorly characterized so far. Before this study, no data were available on TA modules in *C. difficile* with the exception of the recently identified MazEF, a type II TA system member (Rothenbacher et al., 2012). Possible implications of type II TA modules in recurrent *C. difficile* infection, sporulation and biofilm formation were recently discussed (Gil et al., 2015). Among the most challenging aspects of *C. difficile*-associated disease remain the high incidence of recurrent infections and the ability of transition from inert colonization to active infection (Shields et al., 2015; Smits et al., 2016). A comparative genomic study showed that the genomes of most dangerous epidemic bacteria are characterized by the accumulation of TA modules (Georgiades and Raoult, 2011). Promising perspectives for the applications of TA and CRISPR as a basis for the development of new antibacterial strategies could be examined in the future (Lee and Lee, 2016; Yamaguchi and Inouye, 2011).

In conclusion, this study provides the first characterization of type I TA modules in the emergent enteropathogen *C. difficile*. Intriguingly, these chromosomal TA pairs are co-localized with CRISPR array components of bacterial adaptive immunity defense system CRISPR-Cas in the majority of sequenced *C. difficile* strains. Further investigations will help to precise the biological functions of these widespread chromosomal TA loci for *C. difficile* physiology and its successful development inside the host, to uncover the molecular mechanisms involved in their regulation and the possible crosstalk between homologous systems, as well as to evaluate their potential for future therapeutic and biotechnological applications in pathogenic bacteria.

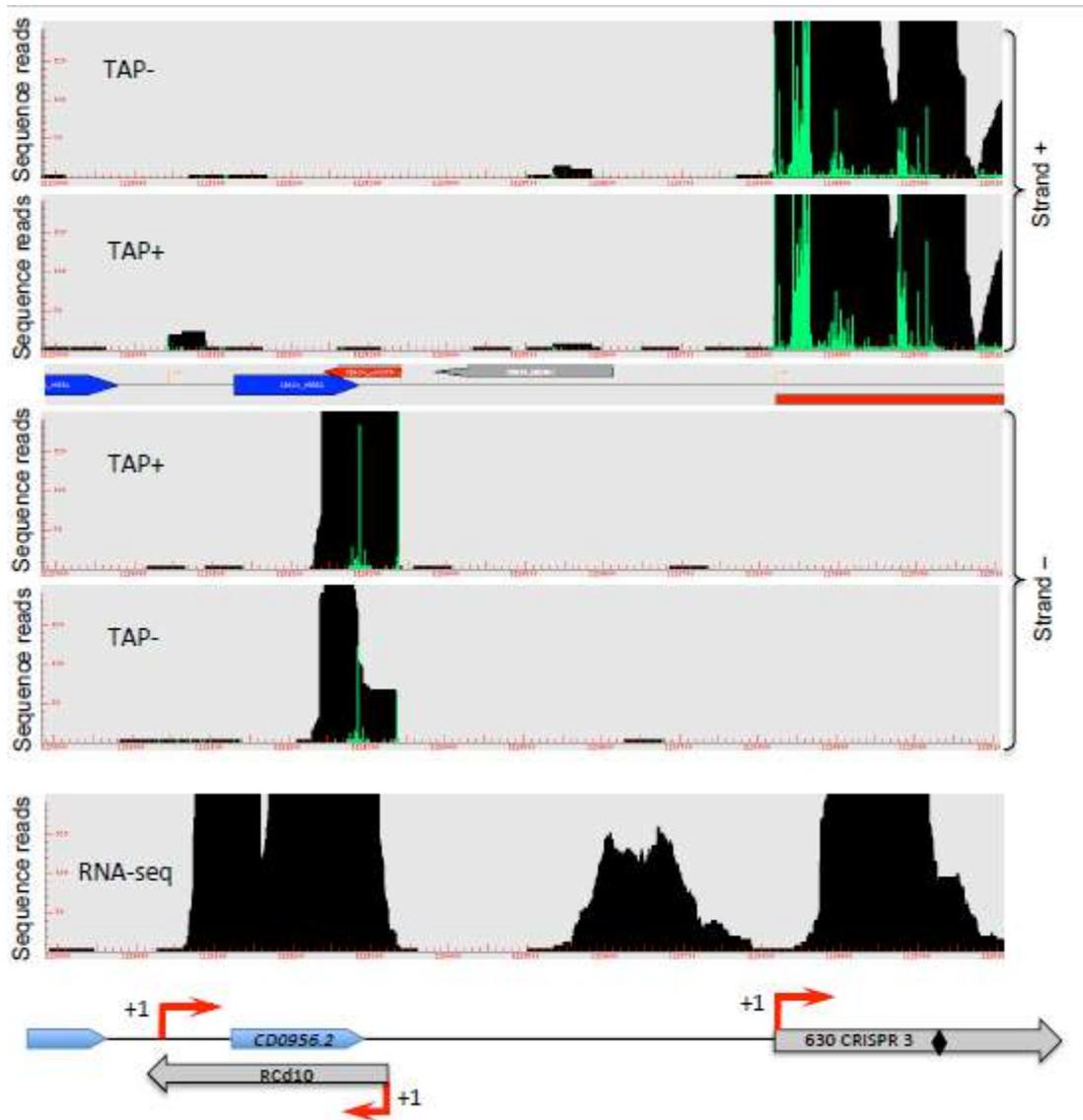
3.4.2 Role of c-di-GMP in *C. difficile* CRISPR-Cas system regulation

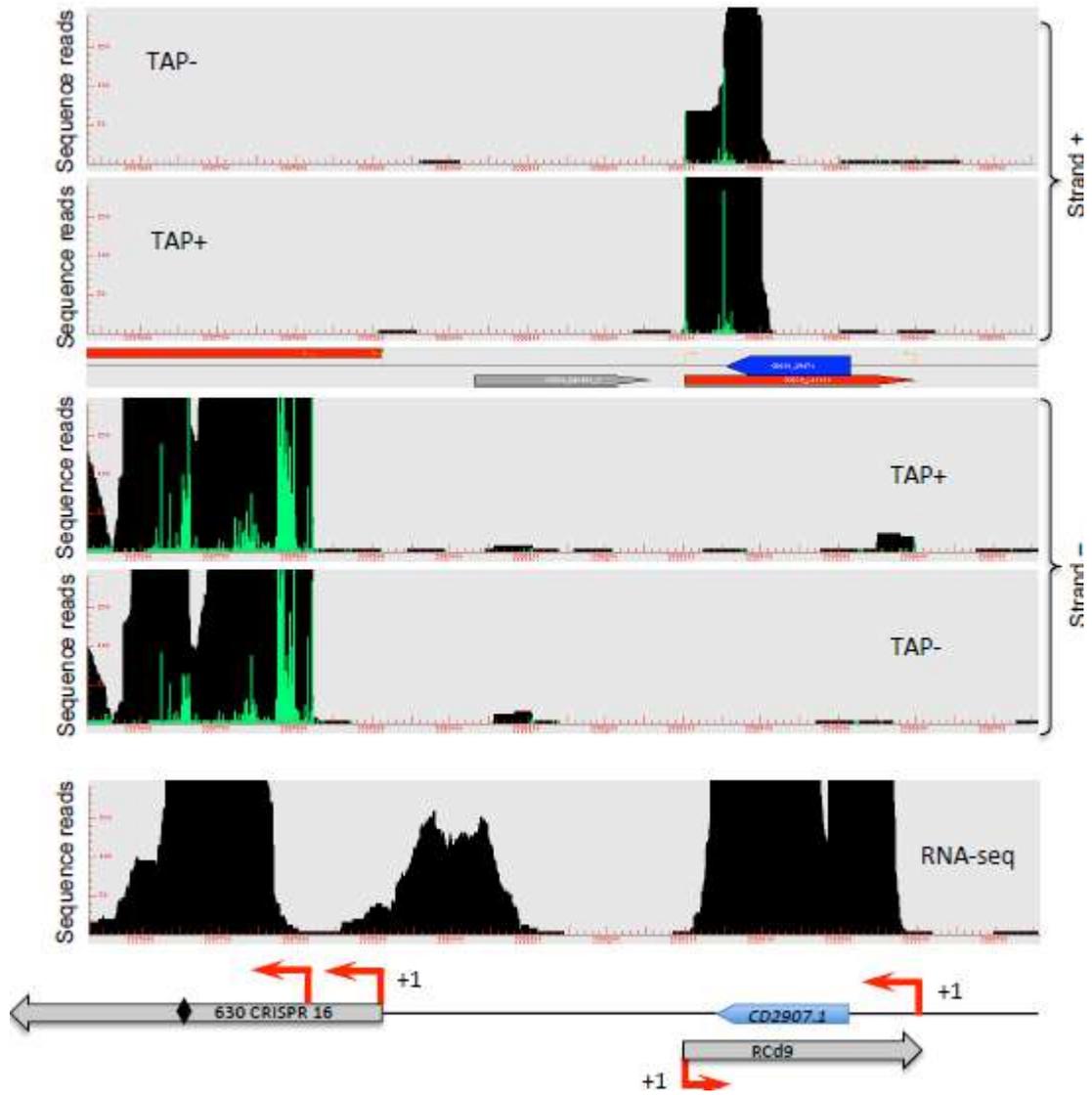
The second messenger c-di-GMP is one of the key components in the regulation of phenotypic shifts in bacteria (Romling et al., 2013). This small molecule often controls the target genes expression through binding with special c-di-GMP riboswitches (Lee et al., 2010; Sudarsan et al., 2008). In this Chapter, we re-examined previously obtained data on *C. difficile* 630 Δ *erm* c-di-GMP-dependent riboswitches (Soutourina et al., 2013) and found, that one of them (*cdi1_7*) is closely related to the CRISPR12 array. Moreover, *cdi1_7* is linked to a *cis*-antisense RNA, overlapping CRISPR12 array on its 3'-end.

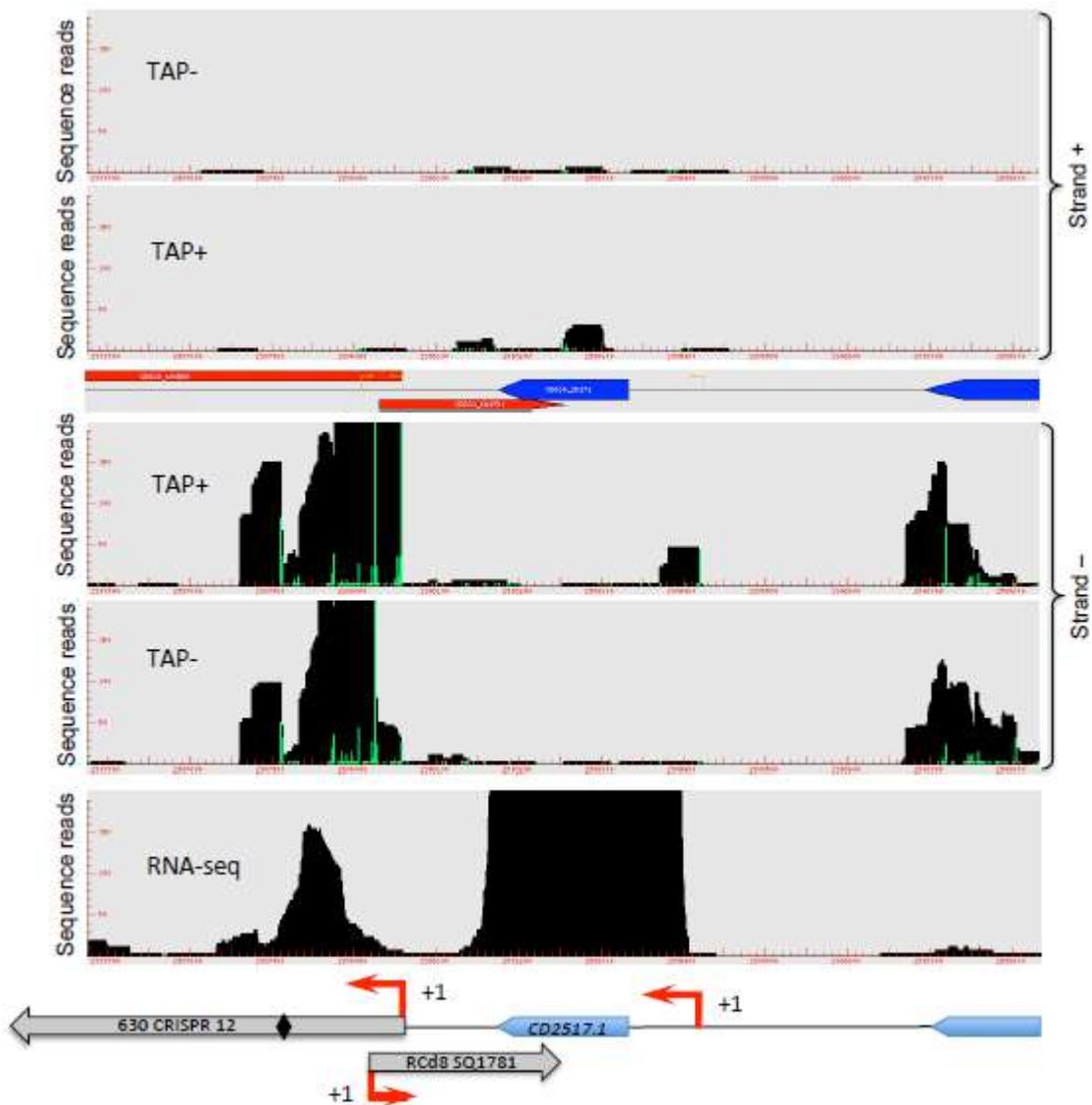
Plasmid efficiency assays with the *C. difficile* CDIP634 strain, overexpressing diguanylate cyclase *dccA*, showed a slight positive effect of high c-di-GMP on CRISPR12 array functionality. Hence, the module *cdi1_7-cis*-antisense RNA may have a moderate impact on the CRISPR12 array expression. Interestingly, this CRISPR array-riboswitch association was identified only in *C. difficile* 630 Δ *erm* CRISPR12 array case. Furthermore, this array is also adjacent to the type I TA module (see above). Apparently, there is a complex genetic system of CRISPR12 array regulation. Further studies would provide more information about the possible role of this genetic system in *C. difficile* CRISPR-Cas regulation.

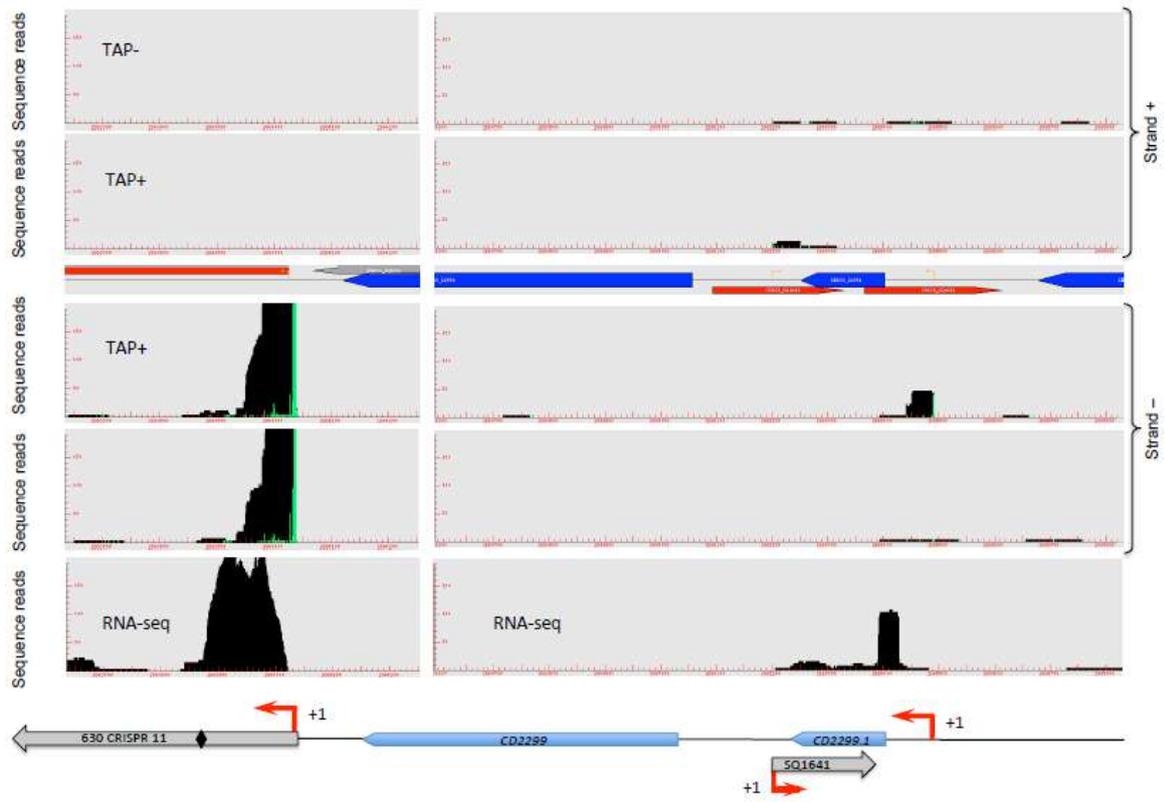
Changes of intracellular c-di-GMP levels widely affect bacterial physiology and gene expression (Romling et al., 2013). Using quantitative PCR, we analyzed expression of all *C. difficile* 630 Δ *erm* CRISPR-Cas system components and revealed the induction of several CRISPR arrays and both *cas* operons in the presence of high c-di-GMP levels. High intracellular concentrations of c-di-GMP are related to switching from the planktonic to the biofilm lifestyle (Bordeleau et al., 2011). Biofilm mode of bacterial growth is characterized by a high density of the cells and the high possibility of phage infection (Babic et al., 2011; Abedon, 2012). Consequently, the positive regulation of the CRISPR-Cas system expression could be an adaptive strategy of *C. difficile* to increased chances of phage infection and HGT. In addition to changing in c-di-GMP levels, transfer to the biofilm state involved great number of different additional factors. Further investigation of *C. difficile* CRISPR-Cas system regulation by biofilm-related signals will lead to a better understanding of this enteropathogenic adaptation to changing environments inside the host.

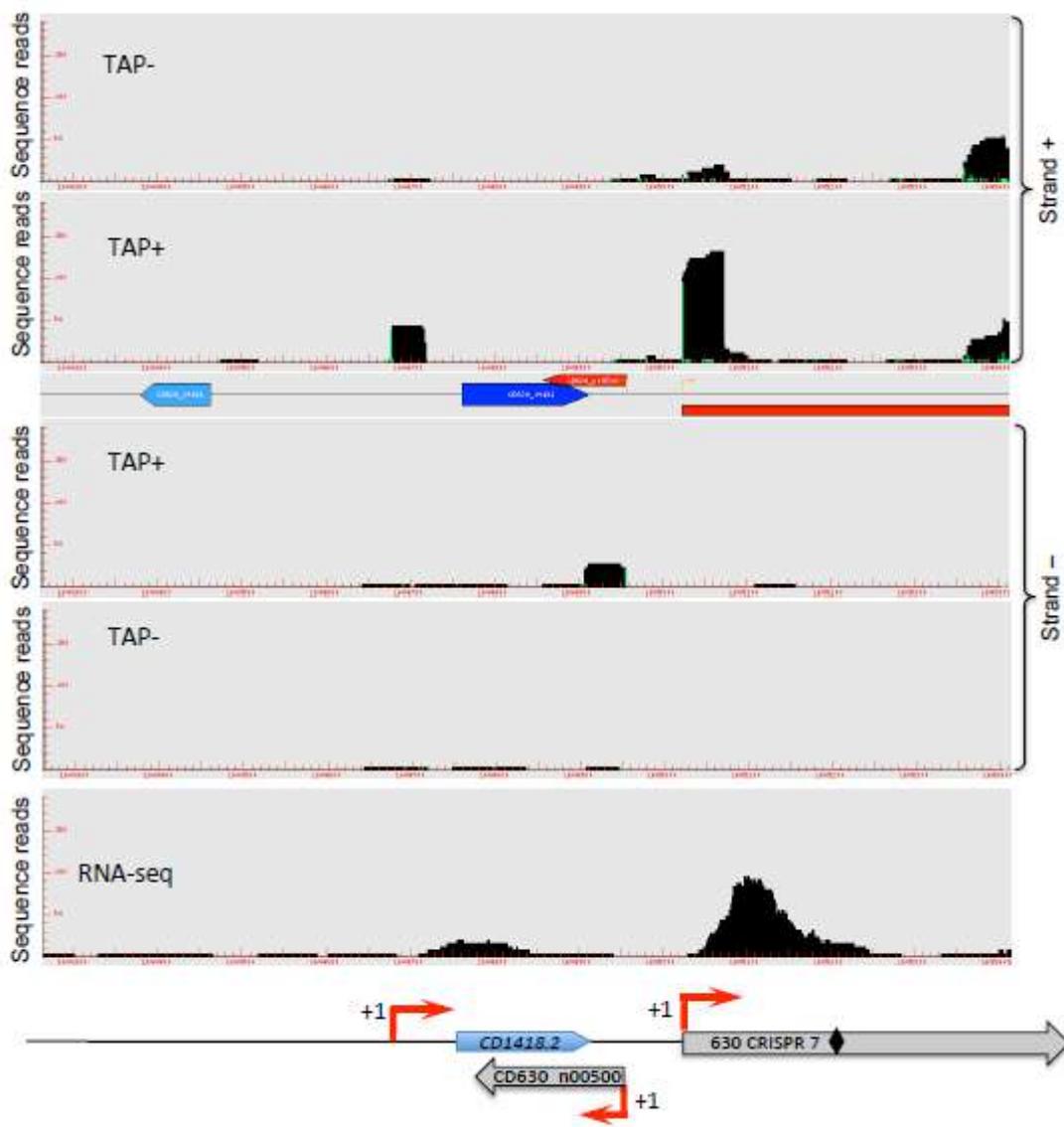
3.5 Supplementary materials











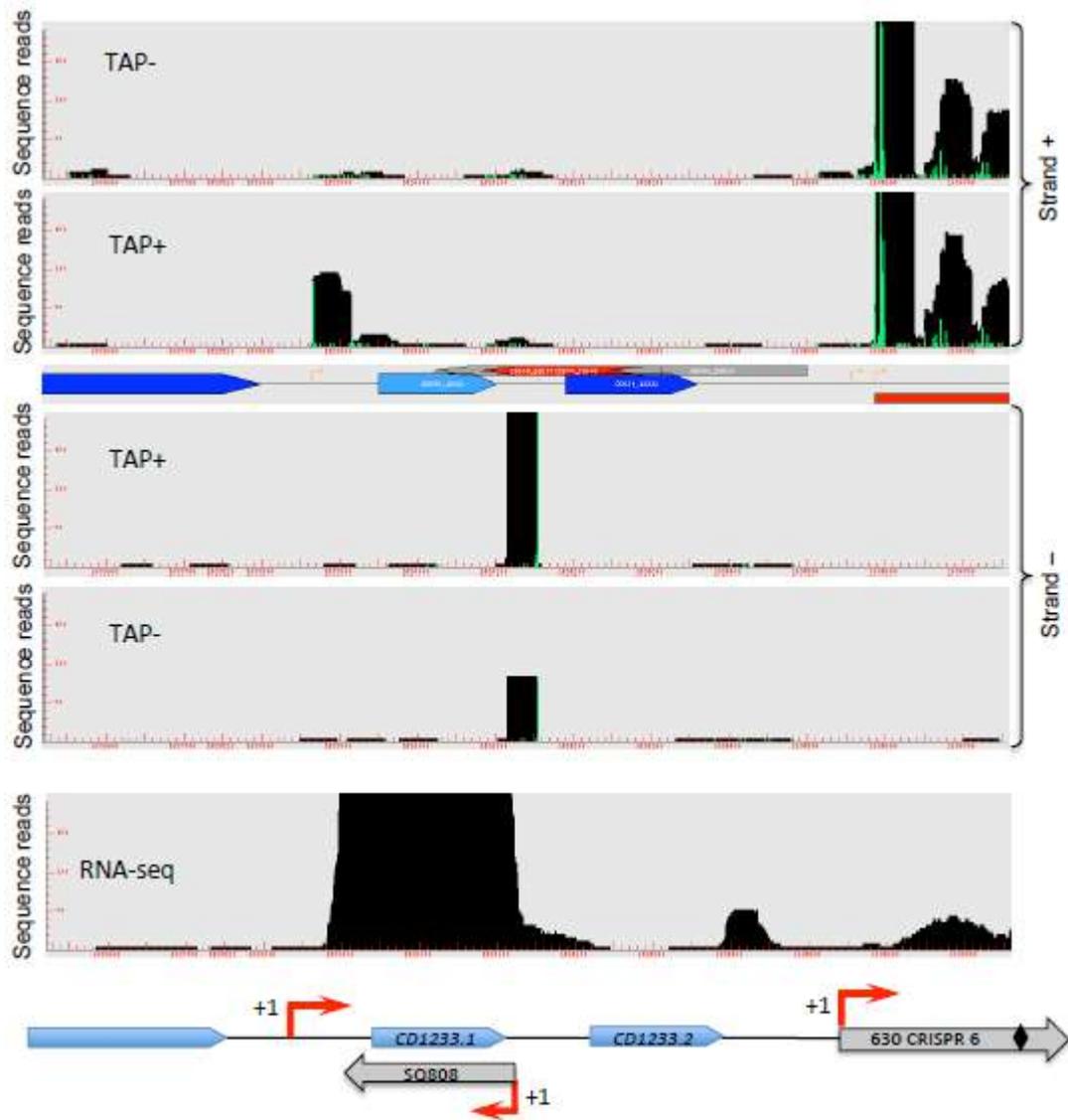


Figure S3.1. RNA-seq and TSS mapping profiles for the TA loci in *C. difficile* strain 630 Δ erm. The TAP-/TAP+ profile comparison for 5'-end RNA-seq data is aligned with RNA-seq data for TA genomic regions. The TSS identified by 5'-end sequencing are indicated by red broken arrows in accordance with the positions of 5'-transcript ends shown by vertical green lines on the sequence read graphs corresponding either to TSS (broken arrows) or to processing sites (vertical arrows). TSS correspond to positions with a significantly greater number of reads in TAP+ sample, potential cleavage sites correspond to positions with a large number of reads in both TAP- and TAP+ samples. 5'-end sequencing data show 51-bp reads matching to the 5'-transcript ends, while RNA-seq data show reads covering the whole transcript. Coding sequences are indicated by blue arrows and the regulatory RNA are indicated by grey arrows. Adapted from (Maikova et al., 2018a) with permission.

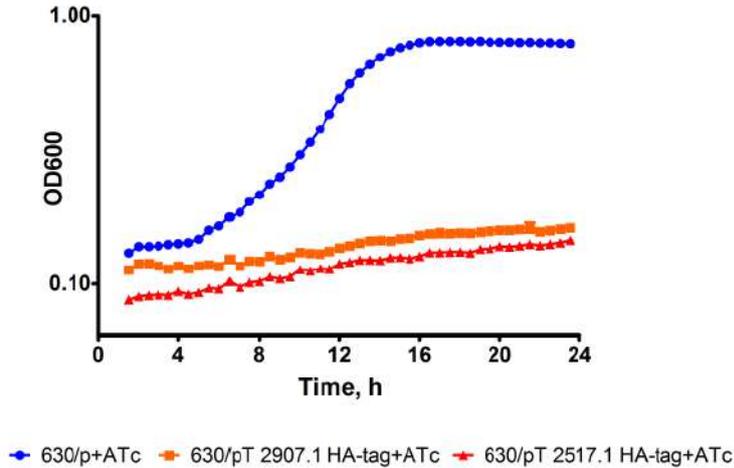


Figure S3.3. Effect of inducible HA-tagged proteins CD2517.1 and CD2907.1/CD0956.2 overexpression on growth in liquid medium.

Growth curves for CDIP369 (630/p) control strain (blue) and 630/p-T-HA strains expressing the HA-tagged CD2517.1 (orange) and CD2907.1/CD0956.2 (red) proteins in TY medium at 37°C in the presence of 250 ng/ml ATc using a GloMax plate reader (Promega). The mean values and standard deviations are shown for three experiments.

Adapted from (Maikova et al., 2018a) with permission.



Figure S3.4. Light microscopy analysis of morphology changes induced by toxin overexpression for CD2517.1 TA module.

Selected images from light microscopy observation of 630/p, 630/pT and 630/pTA strains grown in TY medium at 37°C after 1 h of 250 ng/mL ATc addition.

Adapted from (Maikova et al., 2018a) with permission.

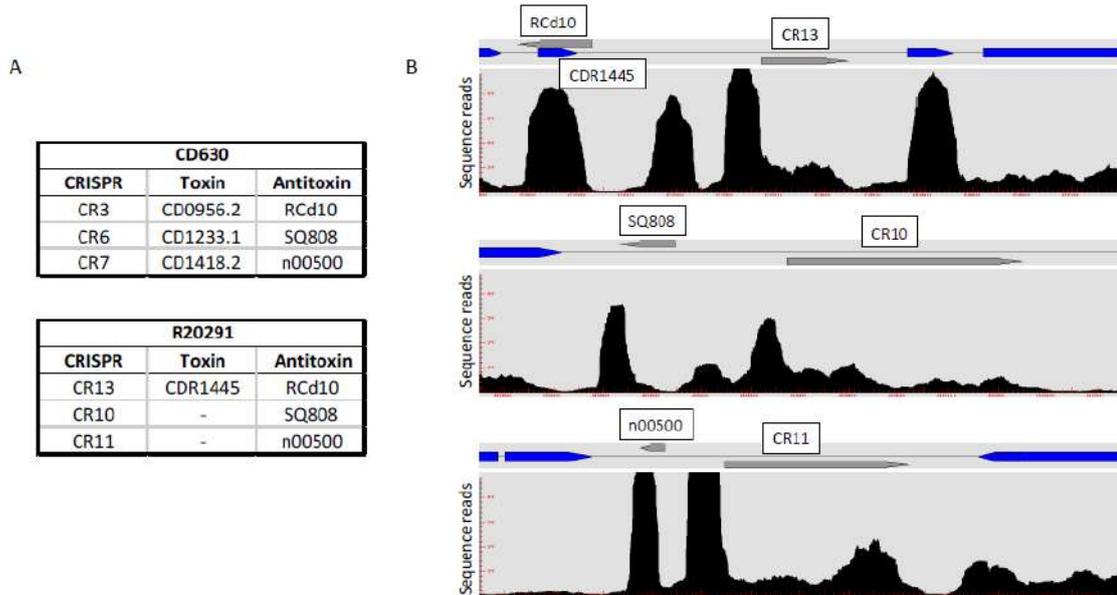


Figure S3.5. RNA-seq profiles for the TA loci in *C. difficile* strain R20291. **A** – potential homologous TA pairs and associated CRISPR arrays are indicated for strains 630 Δ *erm* and R20291. **B** – R20291 RNAseq data (Maldarelli et al., 2016) for corresponding regions were visualized using the COV2HTML software (Monot et al., 2014). Coding sequences are indicated by blue arrows and the regulatory RNA are indicated by grey arrows. Adapted from (Maikova et al., 2018a) with permission.

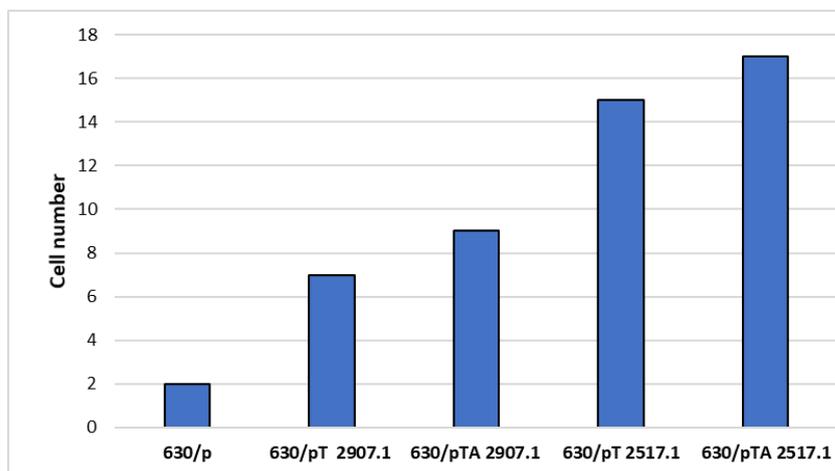


Figure S3.6. Number of the cells with lengths above the value of 630/p mean length with 2 standard deviations (10.5 μ m) in strains, overexpressing CD2907.1 and CD2517.1 TA modules. Cell length was estimated for more than 100 cells for each strain.

Table S3.1. Bacterial strains and plasmids used in Chapter 3.

Adapted from (Maikova et al., 2018a) with permission.

Strain	Genotype	Origin
<i>E. coli</i>		
NEB-10 beta	$\Delta(ara-leu)$ 7697 <i>araD139 fhuA</i> $\Delta lacX74 galK16 galE15 e14-$ $\phi 80\Delta lacZ\Delta M15$ <i>recA1 relA1 endA1 nupG rpsL</i> (Str ^R) <i>rph spoT1</i> $\Delta(mrr-hsdRMS-mcrBC)$	New England Biolabs
DH5 α	F- $\Phi 80\Delta lacZ\Delta M15$ $\Delta(lacZYA-argF)$ U169 <i>recA1 endA1 hsdR17</i> (rK-, mK+) <i>phoA supE44</i> $\lambda-$ <i>thi-1 gyrA96 relA1 supE44 aal14 galK2 lacY1</i> $\Delta(gpt-proA)$ 62 <i>rpsL20</i>	Invitrogen
HB101 (RP4)	(Str ^R) <i>xyt-5 mtl-1 recA13</i> $\Delta(mcrC-mrr)$ <i>hsdSB</i> (r _B -m _B -) RP4 (Tra ⁺ IncP Ap ^R Km ^R Tc ^R)	Laboratory stock
<i>C. difficile</i>		
630 Δerm	630 $\Delta ermB$	Laboratory stock (Hussain et al., 2005)
CDIP51	630 Δerm strain carrying pRPF185 vector	(Boudry et al., 2014)
CDIP369 (630/p)	630 Δerm strain carrying pRPF185 $\Delta gusA$ vector	This work
CDIP317 (630/pT CD2907.1)	630 Δerm strain carrying pDIA6195 plasmid for inducible expression of <i>CD2907.1/CD0956.2</i>	This work
CDIP319 (630/pTA CD2907.1-RCd9)	630 Δerm strain carrying pDIA6196 plasmid for inducible expression of <i>CD2907.1/CD0956.2</i> and co-expression of RCd9/RCd10 from its own promoter	This work
CDIP357 (630/pT CD2517.1)	630 Δerm strain carrying pDIA6319 plasmid for inducible expression of <i>CD2517.1</i>	This work
CDIP332 (630/pTA CD2517.1-RCd8)	630 Δerm strain carrying pDIA6202 plasmid for inducible expression of <i>CD2517.1</i> and co-expression of RCd8 from its own promoter	This work
CDIP998	630 Δerm strain carrying pDIA6623 plasmid for inducible expression of HA-tagged <i>CD2517.1</i>	This work
CDIP999	630 Δerm strain carrying pDIA6624 plasmid for inducible expression of HA-tagged <i>CD2907.1/CD0956.2</i>	This work
CDIP229	630 Δerm <i>sigB::erm</i>	(Kint et al., 2017)
CDIP634	630 Δerm <i>P_{ter}-dcaA</i> (CD1420)	This work
Plasmid		
pRPF185	<i>P_{ter}-gusA</i> Tm ^R expression and cloning vector	(Fagan and Fairweather, 2011)
pDIA6103	pRPF185 Δgus vector derivative	(Soutourina et al., 2013)
pGEM-T easy	TA cloning vector	Promega
pDIA6195	pRPF185 derivative carrying <i>P_{ter}-CD2907.1</i> for inducible <i>CD2907.1/CD0956.2</i> toxin expression	This work
pDIA6196	pRPF185 derivative carrying <i>P_{ter}-CD2907.1-RCd9-P</i> region for inducible <i>CD2907.1/CD0956.2</i> toxin expression and co-expression of RCd9/RCd10	This work
pDIA6319	pRPF185 derivative carrying <i>P_{ter}-CD2517.1</i> for inducible <i>CD2517.1</i> toxin expression	This work
pDIA6202	pRPF185 derivative carrying <i>P_{ter}-CD2517.1-RCd8-P</i> for inducible <i>CD2517.1</i> toxin expression and co-expression of RCd8	This work
pDIA6623	pRPF185 derivative carrying <i>P_{ter}-CD2517.1</i> for inducible <i>CD2517.1</i> -HA-tag toxin expression	This work
pDIA6624	pRPF185 derivative carrying <i>P_{ter}-CD2907.1/CD0956.2</i> for inducible <i>CD2907.1/CD0956.2</i> -HA-tag toxin expression	This work
pMTL-SC7315	Semi-suicidal vector carrying <i>codA</i>	(Cartman et al., 2012)
pDIA6404	pMTL-SC7315 carrying arms for the recombination in 630 Δerm strain to insert <i>P_{tet}</i> element upstream <i>CD1420</i>	This work

Chapter 4. Using endogenous CRISPR-Cas system for genome editing in the human pathogen *C. difficile*

Results from this chapter are in press in Applied and Environmental Microbiology journal (Maikova et al., 2019):

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*. 2019.

4.1 Introduction

C. difficile represents today a real danger for human and animal health. It is the leading cause of diarrhea associated with healthcare 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 remain unanswered on the mechanisms contributing to *C. difficile* success inside the host. Therefore, it is important to develop new genome editing approaches for further investigations of this emerging human pathogen.

During the last years, substantial efforts were concentrated on the development of various CRISPR-based biotechnological tools (Barrangou and Horvath, 2017). In particular, the type II Cas9 and type V Cpf1 (Cas12a) technologies are popular and widespread class 2 systems-based tools are applied for the genome editing in different organisms (Hsu et al., 2014; Safari et al., 2019). Nevertheless, the application of other types of CRISPR-Cas system 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 (Barrangou and Horvath, 2017; Li et al., 2015). This approach is based on the use of a plasmid vector containing artificial CRISPR mini-array with a spacer, targeting a chromosomal gene (Li et al., 2015). crRNAs expressed from plasmid-borne mini-array utilize the endogenous Cas machinery to form a ribonucleoprotein complex, which recognizes the protospacer of choice leading to its cleavage. Destruction of chromosomal DNA leads to killing of wild type cells (Figure 4.1A). Providing the homologous arms in the editing plasmid triggers homologous recombination and allelic exchange with a targeted chromosomal region (Figure 4.1B). This will lead then to the elimination of the resulting plasmid carrying a wild type allele by CRISPR-Cas and the preservation of the chromosomal mutants since

they do not possess the targeted protospacer anymore (Figure 4.1B). Compared to CRISPR-Cas9 and Cpf1 (Cas12a) technologies, this endogenous CRISPR-based method could be easier to set up for editing in prokaryotes. Another advantage of this approach is that there is no need to express heterologous Cas proteins inside bacterial or archaeal cells that may have toxic effects. Until now, the genome editing approach based on endogenous CRISPR-Cas system was successfully applied in several prokaryotic organisms with the examples reported for CRISPR-Cas subtype I-A and III-B or subtype I-B in archaea: *Sulfolobus islandicus* (Li et al., 2015) and *Haloarcula hispanica* (Cheng et al., 2017), respectively; and for subtype I-B in several clostridial species: *C. pasteurianum* (Pyne et al., 2016), *C. tyrobutyricum* (Zhang et al., 2018), *C. saccharoperbutylacetonicum* (Atmadjaja et al., 2019).

In *C. difficile*, various genetic tools for genome manipulation have been established. One of most widely used methods is the ClosTron technology, based on mobile altered type II introns and utilization of retrotransposable activated markers (RAM) (Heap et al., 2010; Kuehne et al., 2011). Though this genome editing technique allows targeting of almost any chromosomal region and RAM markers 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. The additional limitation comes from difficulties in finding an efficient insertion site within a gene of small size. Another popular *C. difficile* genome editing approach is allelic-coupled exchange technique based on a semi-suicidal plasmid vector carrying *Escherichia coli* cytosine deaminase *codA* gene or *C. difficile* orotate phosphoribosyltransferase *pyrE* gene as counter-selection markers (Cartman et al., 2012; Ng et al., 2013). 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 counter-selection procedure is based on the generation of highly toxic compounds from these substrates. Despite the fact that this approach allows creating *C. difficile* mutants carrying point mutations, deletions, and insertions, it can be difficult to apply in some cases. First, mutations that result in growth deficiency phenotype or inactivation of metabolic genes may affect growth on nutrient-poor medium. Secondly, there are some difficulties with losing the editing plasmids in mutant strains after editing, which could lead to spontaneous creation of revertant strains.

Chapter 4. Using endogenous CRISPR-Cas system for genome editing in the human pathogen *C. difficile*.

Recently, the method based on the DNA double-strand breaks in *C. difficile* has been reported (Theophilou et al., 2019). This technology uses the site-specific cleavage by the yeast homing endonuclease I-SceI whose recognition site is introduced to the editing plasmid vector. After the integration of the editing vector into the chromosome, another vector containing the I-SceI endonuclease gene under a control of 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 expression of I-SceI endonuclease that could induce side effects. During last years, successful application of CRISPR-Cas9 and CRISPR-Cpf1 (Cas12a) for genome editing in *C. difficile* was reported (Hong et al., 2018; Inés et al., 2019; Ingle et al., 2019; McAllister et al., 2017). These approaches enhanced the possibilities of genetic manipulation in *C. difficile* and have proven to be efficient. However, Cas9 and Cpf1 technologies require the design of specific single guide RNAs (sgRNAs) plasmids and the editing plasmid is not cured automatically after the editing is complete.

The use of endogenous CRISPR-Cas system can enhance the possibilities of genetic manipulation of *C. difficile*. This Chapter describes the utilization of native *C. difficile* subtype I-B CRISPR-Cas system to generate deletion mutants of the *hfq* gene encoding the RNA chaperone protein Hfq in 630 Δ *erm* and R20291 strains.

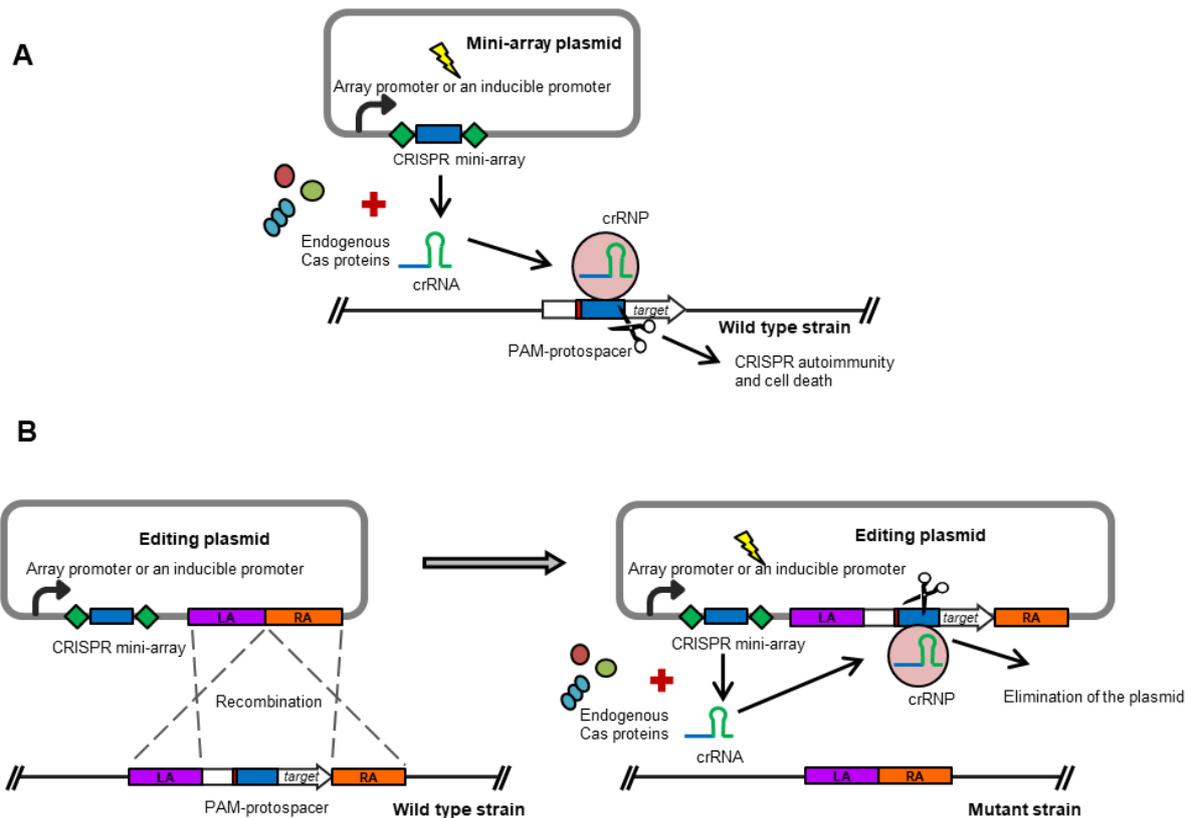


Figure 4.1. General scheme of using endogenous CRISPR-Cas systems for genome editing in bacteria and archaea. **A** – the crRNA is expressed from a vector-borne mini CRISPR array under the control of native or inducible promoters. The crRNA forms ribonucleoprotein (crRNP) complex with endogenous Cas proteins, which recognizes and directs the cleavage of the PAM-associated protospacer, localized at the target chromosome region. This leads to chromosome disruption and cell death. **B** – an editing plasmid, additionally carrying homologous arms (LA and RA), allows the recombination between the plasmid and the chromosome to occur before the CRISPR interference. The crRNP targets the PAM-protospacer on the plasmid, which leads to the elimination of plasmid and preservation of the chromosomal mutants. LA – left arm, RA – right arm.

4.2 Materials and methods

4.2.1 Bacterial strains, plasmids and growth conditions

All the plasmids and bacterial strains used in this study are listed in Table 4.1. *C. difficile* strains were grown BHI (Difco) or TY (Dupuy and Sonenshein, 1998) medium at 37°C under anaerobic conditions (5% H₂, 5% CO₂, and 90% N₂), within an anaerobic chamber (Jacomex). BHI medium supplemented with yeast extract (5 mg/ml), and L-cysteine (0.1%) (BHIS) was used in sporulation experiments. When needed, Tm at final

concentration of 15 µg/ml was added to *C. difficile* cultures. *E. coli* strains (Table 4.1) were grown in LB medium (Bertani, 1951), supplemented with Amp (100 µg/ml) and Cm (15 µg/ml) when it was suitable. The non-antibiotic analog ATc was used for induction of the P_{tet} promoter of pRPF185 vector derivatives in *C. difficile* (Fagan and Fairweather, 2011).

4.2.2 Plasmid construction and conjugation into *C. difficile*

To create artificial CRISPR mini-arrays targeting *C. difficile* *hfq* gene, the full (-403 to -1 relative to the first nucleotide of the first repeat in the array) and partial (-154 to -1 relative to the first nucleotide of the first repeat in the array) leader sequences of *C. difficile* 630 Δ *erm* CRISPR16 array were amplified by PCR on genomic DNA (Figure S4.1A, B in Supplementary materials). The artificial repeat-spacer-repeat motif was amplified by PCR from synthetic oligonucleotides to generate the double-stranded fragment. Full or partial leader sequences and repeat-spacer-repeat motif were assembled and cloned into BamHI and XhoI sites of pRPF185 Δ *gus* plasmid vector (Soutourina et al., 2013) using Gibson assembly reaction (Gibson et al., 2009) giving pECrF_*hfq* and pECrP_*hfq* mini-array plasmids (Figure 4.2B).

To construct editing plasmids, approximately 1200-bp long regions flanking the *hfq* gene of the 630 Δ *erm* and R20291 strains were amplified by PCR and introduced into SmaI restriction site of pECrF_*hfq* or pECrP_*hfq* using Gibson assembly reaction resulting in pECrFA_*hfq*630, pECrPA_*hfq*630 and pECrPA_*hfq*R20291 plasmids (Figure 4.2C).

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

The DNA sequencing was performed to verify plasmid constructs. All resulting plasmids were transformed into *E. coli* HB101 (RP4) strain and further transferred to *C. difficile* cells by conjugation. Heat shock method with incubation for 15 min at 50°C was used to get the highest conjugation efficiency (Kirk and Fagan, 2016). *C. difficile* transconjugants were selected on BHI agar containing Tm (15 µg/ml), Cs (25 µg/ml) and Cfx (8 µg/ml).

4.2.3 Deletion of the *hfq* gene and validation of Δhfq mutants

To induce the expression of the CRISPR mini-arrays under the control of P_{tet} promoter, *C. difficile* transconjugants containing pECrP_hfq, pECrPA_hfq630 or pECrPA_hfqR20291 plasmids 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 μ g/ml) to check for the plasmid loss. Subsequently, selected clones without plasmids were analyzed by PCR to detect the chromosomal deletion of the *hfq* gene. The resulting PCR fragments have been sequenced to confirm the gene deletion.

4.2.4 RNA extraction and qRT-PCR

For the total RNA extraction, *C. difficile* 630 Δerm - and R20291-derived pRPF185 Δgus 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). The total RNA isolation was performed as previously described (André et al., 2008). The cDNA synthesis by reverse transcription and real-time quantitative PCR (qRT-PCR) was performed as previously described (Saujet et al., 2011) using BioRad CFX Connect Real-Time system. The expression level of the *hfq* gene was calculated relative to that of the 16S RNA gene (Metcalf et al., 2010).

4.2.5 Protein extract preparation and Western blotting

To extract total proteins, *C. difficile* 630 Δerm - and R20291-derived pRPF185 Δgus 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 (Boudry et al., 2014).

For each sample, 30 μ g of protein extract was loaded on two 15% SDS polyacrylamide gels in parallel. After the electrophoresis, proteins from the 1st gel were transferred to a polyvinylidene fluoride membrane. The membrane hybridization with primary and secondary antibodies was then performed as described before (Boudry et al., 2014). The bioluminescent signal from the secondary antibodies was detected using the SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific) and the Fusion FX (Vilber Lourmat) digital camera. The 2nd gel was stained with the InstantBlue dye (Expedeon) and used as a loading control (Figure S4.2 in Supplementary materials).

4.2.6 Sporulation assay

C. difficile strains harboring 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 at OD₆₀₀ of 0.1 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 OD₆₀₀ of 1.0 – 1.5, they were diluted to 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 CFUs, the first sample was serially diluted and spotted (10 μ l per spot) onto BHI agar containing 0.1 % of 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 % of taurocholate to estimate the number of spores.

Table 4.1. Bacterial strains and plasmids used in Chapter 4.

Strain	Genotype	Source
<i>E. coli</i>		
NEB-10 beta	$\Delta(ara-leu)$ 7697 <i>araD139 fhuA</i> $\Delta lacX74 galK16 galE15 e14- \phi 80dlacZ\Delta M15 recA1 relA1 endA1 nupG rpsL$ (<i>StrR</i>) <i>rph spoT1</i> $\Delta(mrrhsdRMS-mcrBC)$	New England Biolabs
HB101 (RP4)	<i>supE44 aa14 galK2 lacY1</i> $\Delta(gpt-proA)$ 62 <i>rpsL20</i> (<i>StrR</i>) <i>xyl-5 mtl-1 recA13</i> $\Delta(mcrC-mrr)$ <i>hsdSB</i> (<i>rB-mB-</i>) RP4 (<i>Tra+ IncP ApR KmR TcR</i>)	Laboratory stock
<i>C. difficile</i>		
630 Δerm	Sequenced reference strain $\Delta ermB$	Laboratory stock (Hussain et al., 2005)
R20291	PCR-ribotype 027 epidemic strain	Laboratory stock
wt/p	630 Δerm or R20291 carrying pRPF Δgus plasmid	This work
$\Delta hfq/p$	630 $\Delta erm\Delta hfq$ or R20291 Δhfq carrying pRPF Δgus plasmid	This work
$\Delta hfq/p-hfq$	630 $\Delta erm\Delta hfq$ or R20291 Δhfq carrying p- <i>hfq</i> plasmid	This work
Plasmid	Description	Reference
pRPF185 Δgus	pRPF185 Δgus vector derivative	(Fagan and Fairweather, 2011; Soutourina et al., 2013)
pECrF_hfq	pRPF185 Δgus carrying the <i>hfq</i> gene targeting CRISPR mini-array with the full leader sequence	This work
pECrP_hfq	pRPF185 Δgus carrying the <i>hfq</i> gene targeting CRISPR mini-array with the partial leader sequence under the control of P _{tet} promoter	This work
pECrFA_hfq630	pECrF_hfq carrying arms for the recombination in 630 Δerm strain	This work
pECrPA_hfq630	pECrP_hfq carrying arms for the recombination in 630 Δerm strain	This work
pECrPA_hfqR20291	pECrP_hfq carrying arms for the recombination in R20291 strain	This work
p- <i>hfq</i>	pRPF185 Δgus carrying <i>hfq</i> gene under the control of P _{tet} promoter	This work

4.3 Results

4.3.1 Construction of targeting mini-array plasmids and verification of their functionality

To evaluate the possibility of using endogenous *C. difficile* CRISPR-Cas system for targeting of specific sequences on bacterial chromosome, we have chosen the *hfq* gene. Hfq is a bacterial RNA-binding protein that plays major roles in RNA metabolism and global posttranscriptional network, in particular in Gram-negative bacteria (Sobrero and Valverde, 2012). The study of Hfq depletion in *C. difficile* 630 Δ *erm* (Boudry et al., 2014) suggested a pleiotropic role of this protein in *C. difficile* physiology with the most pronounced effect 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*. The previous attempts to inactivate the *hfq* gene using Clostron gene knockout system were unsuccessful (Boudry et al., 2014). We have also tried to delete *hfq* using the *codA* allelic exchange approach (Cartman et al., 2012; Ng et al., 2013), but also without success (data not shown).

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 Figure 4.2. We first constructed two CRISPR mini-array plasmids targeting the *hfq* gene (pECrF_hfq and pECrP_hfq). The mini-array was based on *C. difficile* 630 Δ *erm* CRISPR 16 array, which is highly expressed and capable of interference (Boudry et al., 2015). Two variants of the leader sequence upstream of the mini-array were used (Figure S4.1A, B in Supplementary materials): the full leader (403 bp sequence upstream of the first direct repeat of CRISPR 16 array) containing all native promoters that should allow autonomous expression of the mini-array and a partial leader (a 154 bp region upstream of the first direct repeat of CRISPR 16 array), which lacked native promoters but should allow inducible expression of the mini-array from the vector-borne ATc-inducible promoter (P_{tet}). The repeat-spacer-repeat motif of the synthetic mini-array was also based on 29-bp repeat sequences of *C. difficile* 630 Δ *erm* CRISPR 16 array (Figure 4.2A and Figure 4.S1A, B in Supplementary materials). For successful recognition of the protospacer by *C. difficile* CRISPR-Cas system, the functional PAM flanking protospacer at the 5'-end is necessary (Boudry et al., 2015). Two functional trinucleotide 5' CCA and CCT PAMs of *C. difficile* CRISPR-Cas system have been experimentally validated and additional alternative motifs such as CCC, CCG and TCA

have been predicted (Boudry et al., 2015) and also confirmed (see Chapter 2). The coding region of the *hfq* gene possesses at least three functional CCW motifs and two alternative TCA motifs. The mean length of *C. difficile* spacers is 37 bp. A 37-bp sequence associated with the 5' CCT PAM was chosen inside the *hfq* gene sequence (Figure 4.2A). The pECrF_hfq and pECrP_hfq plasmids (Figure 4.2B) were conjugated to *C. difficile* 630 Δ *erm* cells using the heat shock method to ensure the highest conjugation efficiency (Kirk and Fagan, 2016). No transconjugants were obtained after conjugation of the pECrF_hfq plasmid in *C. difficile* 630 Δ *erm* suggesting CRISPR autoimmunity due to self-targeting (Figure 4.3A). Conjugation efficiency of 380 transconjugants/ml was observed after conjugation with pECrP_hfq. In contrast, control conjugation with the pRPF185 Δ *gus* vector revealed 5480 transconjugants/ml. The smaller number of transconjugants in pECrF_hfq conjugation reaction could be due to possible P_{tet} promoter leakage leading to partial self-cleavage. To check for the efficiency of self-targeting by crRNA expressed from the pECrP_hfq plasmid, eight transconjugants were restreaked on BHI agar plates supplemented with 500 ng/ml ATc to fully induce the expression of the mini-array. No growth was observed on these plates indicating highly efficient self-targeting by the induced mini-array (Figure 4.3B). The same effects were observed after conjugation of pECrF_hfq and pECrP_hfq plasmids in *C. difficile* R20291 cells suggesting that the synthetic array based on the *C. difficile* 630 Δ *erm* CRISPR 16 leader and repeat sequences mimic well native subtype I-B CRISPR arrays in *C. difficile* for at least both 630 and R20291 strains. Therefore, *C. difficile* endogenous CRISPR-Cas system can recognize and target protospacers on the bacterial chromosome using crRNAs expressed from plasmid-borne artificial mini-array and this feature can be utilized for genome editing.

Chapter 4. Using endogenous CRISPR-Cas system for genome editing in the human pathogen *C. difficile*.

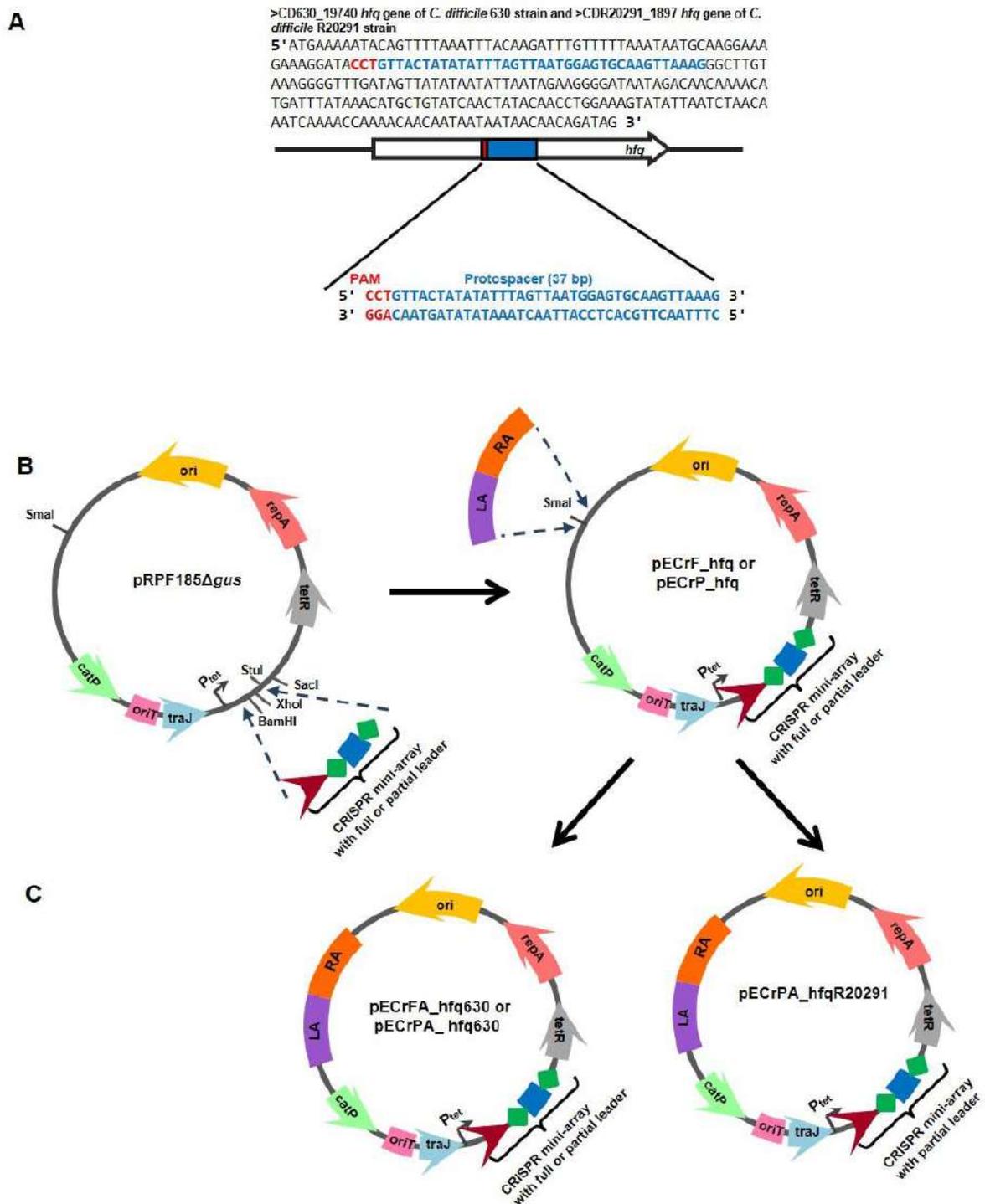


Figure 4.2. Strategy for the design of the editing plasmids to delete the *hfq* gene in *C. difficile* 630 Δ *erm* and R20291 strains. **A** – the coding sequence of *C. difficile* 630 and R20291 *hfq* gene and a 37-bp sequence associated with the 5' CCT PAM, selected as a protospacer for the mini-array. **B** – construction of the pECrF_hfq and pECrP_hfq mini-array plasmids on the basis of pRPF185 Δ *gus* vector. The mini-arrays sequences were cloned into 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 homologous arms (LA and RA) were cloned into SmaI restriction site.

(Figure 4.2. Continue) The “F” in the plasmid names states for the full-length leader region for autonomous expression of mini-array under the control of native promoters while “P” points out the presence of partial leader region without native promoters for mini-array expression under the control of inducible P_{tet} promoter. The presence of homologous arms for recombination within 630 Δerm or R20291 strains is indicated by “A” and a strain name. pECrFA_hfq630 plasmid carrying the mini-array with the full-length leader region was not efficient for gene deletion in 630 Δerm strain, by contrast, pECrPA_hfq630 and pECrPA_hfqR20291 were efficiently used for the hfq gene deletion in 630 Δerm and R20291 strains, respectively.

4.3.2 Construction of the genome editing plasmid and deletion of the *hfq* gene of *C. difficile* 630 Δerm and R20291

We first assessed which mini-array plasmid, pECrF_hfq or pECrP_hfq, is best for *C. difficile* genome manipulation. 1200 bp-long regions flanking the *hfq* gene of the 630 Δerm strain (Figure S4.1C in Supplementary materials) were amplified by PCR and introduced into the SmaI restriction sites of pECrF_hfq or pECrP_hfq using Gibson assembly (Figure 4.2C). No transconjugants were obtained after conjugation of *C. difficile* 630 Δerm with pECrFA_hfq630 carrying the mini-array with the full-length leader region (Figure 4.3C). Presumably, this means that the CRISPR-induced autoimmunity/degradation of DNA around the targeted protospacer is more efficient than homologous recombination between the chromosome and the homologous region of pECrFA_hfq630. Be that as it may, the plasmid with the full-length CRISPR array leader sequence is clearly not suitable for genome editing. After conjugation with pECrPA_hfq630 plasmid carrying the mini-array under the control of inducible P_{tet} promoter, about 460 transconjugants/ml were obtained. To induce expression of the *hfq* targeting mini-array, ten 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 chromosome and plasmid had occurred (Figure 4.3D) or that CRISPR interference was not efficient. One clone from each plate was then restreaked on BHI plates with or without Tm to check for plasmid loss. Three out of ten clones lost the plasmid. When analyzed by PCR, these clones turned out to be Δhfq mutants (Figure 4.4A). Thus, a plasmid containing an inducibly transcribed mini CRISPR array and arms for homologous recombination at the targeted protospacer allows efficient genome editing in *C. difficile*.

Chapter 4. Using endogenous CRISPR-Cas system for genome editing in the human pathogen *C. difficile*.

The coding region of the *hfq* gene of *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 on the basis of pECrP_*hfq* mini-array plasmid the pECrPA_*hfq*R20291 plasmid with R20291 variants of homologous arms of *hfq* flanking sequences (Figure 4.2C and Figure S4.1D in Supplementary materials). Nine out of ten selected transconjugants had lost the plasmid and PCR analysis showed that seven out of nine clones without the plasmid were Δ *hfq* mutants (Figure 4.4.A).

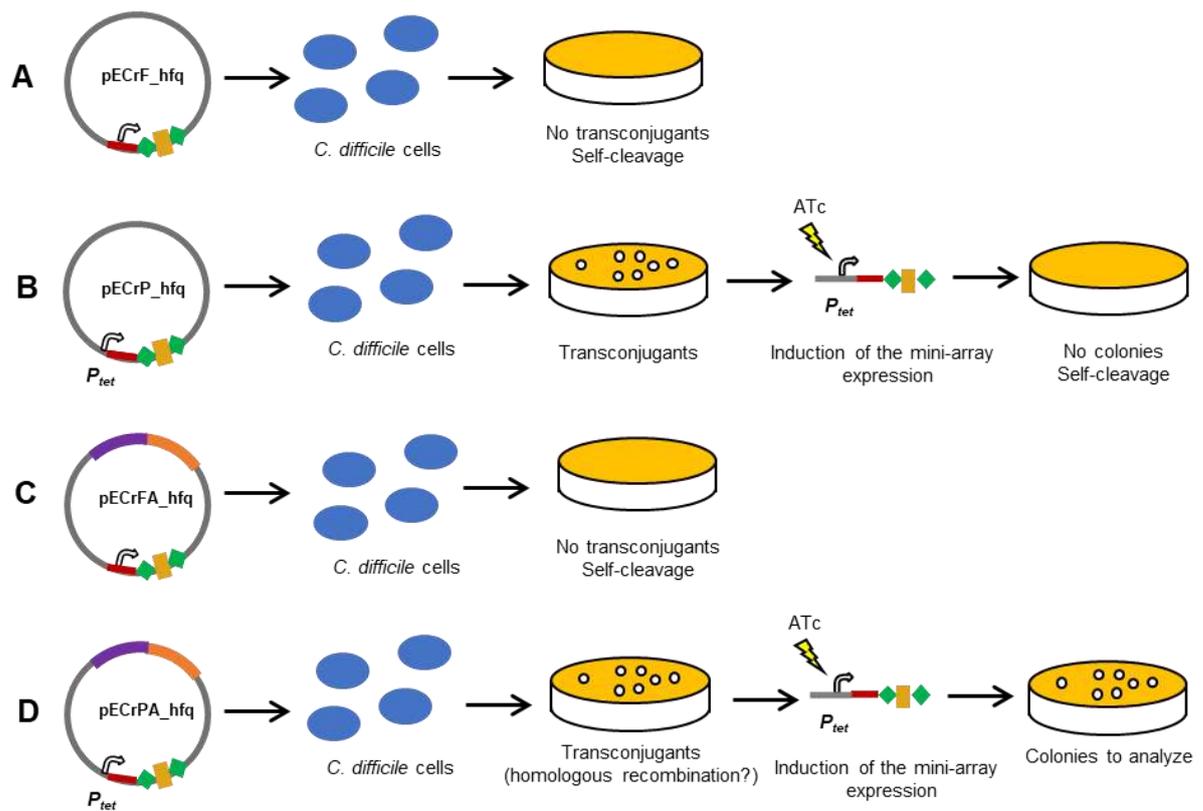


Figure 4.3. Different effects of the conjugation of mini-array and editing plasmids into *C. difficile* cells. **A** – pECrF_*hfq*: CRISPR self-cleavage induced by an immediate expression of the mini-array from the plasmid after conjugation. **B** – pECrP_*hfq*: CRISPR self-cleavage resulted from the ATc-induced expression of the mini-array from the plasmid after second plating of transconjugants. **C** – pECrFA_*hfq*: CRISPR self-cleavage induced by an immediate expression of the mini-array from pECrFA_*hfq* plasmid after conjugation. **D** – pECrPA_*hfq*: homologous recombination between the chromosome and the plasmid and cleavage of the plasmid resulted from the ATc-induced expression of the mini-array from the plasmid after second plating of transconjugants.

4.3.3 Validation and complementation of *hfq* deletion strains

To validate the *hfq* deletion, we have assessed the *hfq* mRNA expression in the wild type (wt/p) and Δhfq mutant strains carrying an empty pRPF185 Δgus vector ($\Delta hfq/p$) as well as in the complemented Δhfq *C. difficile* strains $\Delta hfq/p-hfq$ expressing plasmid-borne *hfq*. Quantitative reverse transcription-PCR (qRT-PCR) analysis confirmed the absence of *hfq* expression in *C. difficile* 630 $\Delta erm \Delta hfq$ and R20291 Δhfq strains and the presence of the transcript in wild-type strains (Figure 4.4B). Furthermore, 400-500-fold increase in *hfq* mRNA abundance was detected in complemented strains after P_{tet} induction in the presence of ATc (Figure 4.4B). Western blotting with polyclonal anti-Hfq antibodies confirmed the lack of the Hfq protein in $\Delta hfq/p$ strains (Figure 4.4C). InstantBlue stained protein gels used as loading controls are shown in Figure S4.2 (Supplementary materials).

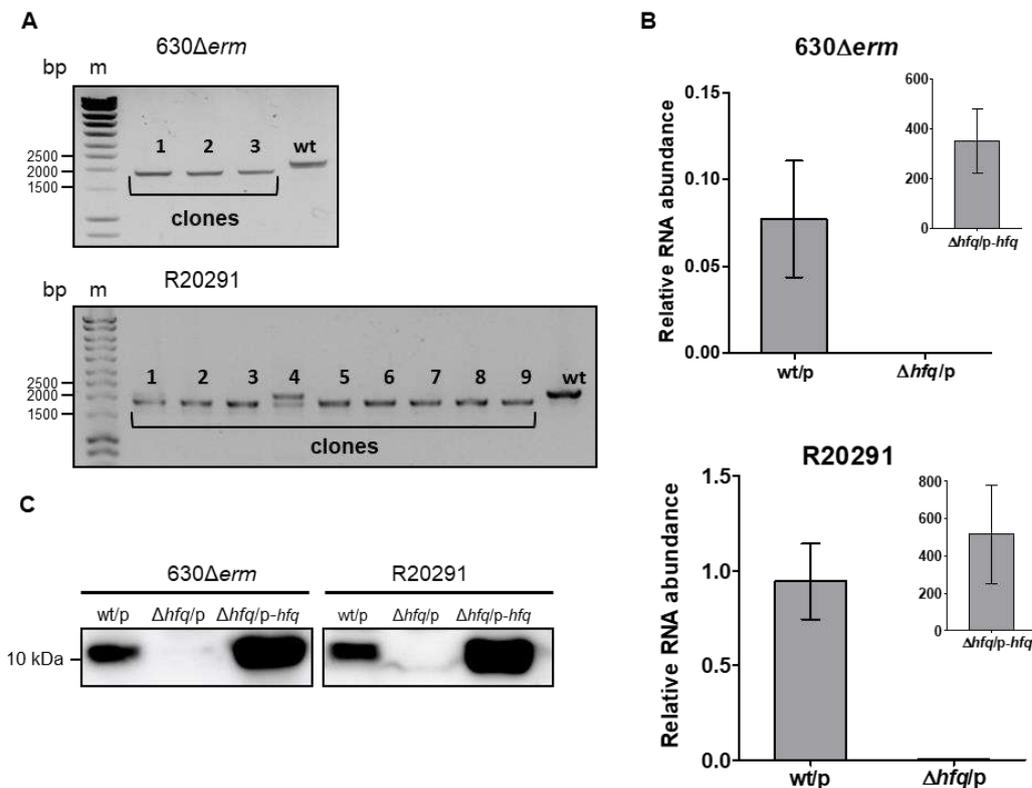


Figure 4.4. Validation of *hfq* deletion mutants. **A** – PCR analysis of the *C. difficile* clones, which have lost the plasmid after the genome editing. 2151 bp PCR bands correspond to the wild type genotype; 1893 bp PCR bands correspond to the mutant genotype. For R20291 strain, both wt and mutant copy has been detected with clone 4 (lane 4), this clone was discarded from further analysis. **B** – qRT-PCR analysis of the wild type (wt/p) and Δhfq mutant strains ($\Delta hfq/p$) carrying an empty pRPF185 Δgus , and complemented Δhfq *C. difficile* strains ($\Delta hfq/p-hfq$). **C** – Western blot analysis of wt/p Δhfq -p, and $\Delta hfq/p-hfq$ *C. difficile* strains. As loading controls, InstantBlue stained protein gels were used (Figure S4.2 in Supplementary materials).

4.3.4 Sporulation assay of *C. difficile* 630 Δ erm Δ hfq mutants

Sporulation represents one of the crucial features of *C. difficile* as a successful pathogen. The previous work revealed that the Hfq protein is likely to control sporulation rates in *C. difficile* 630 Δ erm-derived strains (Boudry et al., 2014). 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 sporulation rates in 630 wt/p, Δ hfq/p, and Δ hfq/p-*hfq* strains. After 24 h and 48 h in BHIS medium supplemented with Tm and ATc, the mutant strain (Δ hfq/p) demonstrated a higher level of sporulation than the wild type (wt/p) (Figure 4.5). In addition, the complemented strain (Δ hfq/p-*hfq*) showed the reversion of sporulation efficiency to the level close to the wild type (Figure 4.5). Thus, these results are consistent with previously obtained data and confirm the potential involvement of Hfq protein in the control of the sporulation process in *C. difficile* (Boudry et al., 2014).

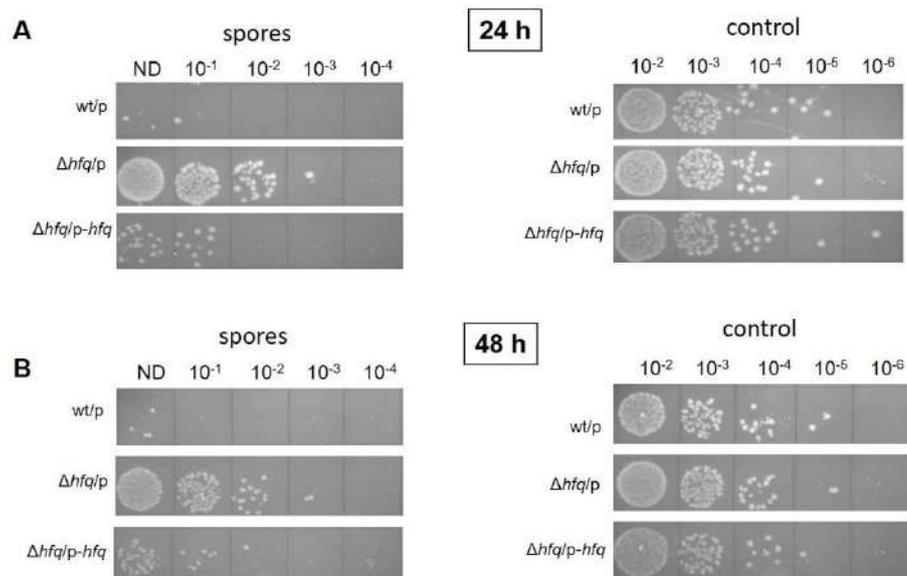


Figure 4.5. Sporulation levels in *C. difficile* 630 Δ erm wt/p, Δ hfq/p, and Δ hfq/p-*hfq* strains (spores) and the total amount of bacteria in CFUs (control). A – After 24 h of growth in BHIS supplemented with Tm and ATc. B – After 48 h of growth in BHIS supplemented with Tm and ATc. The serial dilutions of the cultures spotted on BHI+tautocholate plates are indicated (ND – not diluted). "Spores" samples have been heated to kill all cells other than spores, while the "control" samples have not been heated to estimate the total amount of bacteria.

4.4 Discussion

Over the last decade, the rapid development of various biotechnological tools based on prokaryotic adaptive immune CRISPR-Cas systems has occurred (Barrangou and Horvath, 2017). In addition to the most popular CRISPR tools based on class 2 Cas9 and Cpf1 (Cas12a) proteins (Hsu et al., 2014; Safari et al., 2019), other CRISPR-Cas systems are also being actively explored for the genetic manipulation purposes. One of the most promising applications is the use of endogenous CRISPR-Cas system for genome editing and engineering in bacteria and archaea (Barrangou and Horvath, 2017; Li et al., 2015). In contrast to Cas9- and Cpf1 (Cas12a)-based approaches, this method does not require the expression of heterologous proteins in bacterial or archaeal cells.

In this Chapter 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 (Cartman et al., 2012; Heap et al., 2010; Hong et al., 2018; Kuehne et al., 2011; McAllister et al., 2017; Ng et al., 2013; Theophilou et al., 2019), they could present some limitations in their 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 RNA chaperone Hfq. Attempts to inactivate this gene using other approaches including ClosTron technology (Boudry et al., 2014) and *codA* allelic exchange were not successful (data not shown). Though a strain depleted for Hfq by expression of antisense RNA was available, construction of an *hfq* deletion mutant would have interesting possibilities for future studies of the regulatory role of Hfq and its RNA network in *C. difficile*.

The general workflow for application of native CRISPR-Cas genome editing method in *C. difficile* is presented in Figure 4.6. To repurpose the endogenous CRISPR-Cas system for deletion of the *hfq* gene, we designed plasmid vectors carrying targeting mini-array and full editing plasmids (Figure 4.2B, C). The *C. difficile* 630 Δ *erm* CRISPR 16 array was chosen as a basis for the synthetic mini-array since it was functional for interference (Boudry et al., 2015), (see Chapter 2). Full leader for the autonomous expression of the mini-array and the partial leader sequence completed with a plasmid-borne P_{tet} promoter for the inducible expression were used to construct two versions of mini-array plasmids (Figure S4.1A, B). The repeat-spacer-repeat motif for the artificial mini-array was composed of 29-bp repeat sequences and a 37-bp spacer sequence, associated with the functional 5' CCT PAM inside the *hfq* gene coding region.

Conjugation experiments with pECrF_hfq and pECrP_hfq suggested that both variants of the synthetic array constructions are suitable for genome targeting and efficiently induced genome cleavage in *C. difficile* 630 Δ erm strain (Figure 4.3A, B).

The editing plasmid carrying homologous arms flanking the *hfq* gene and the full leader sequence with native promoters (pECrFA_hfq630) had the same autoimmune effect as pECrF_hfq (Figure 4.3C). To facilitate the genome editing procedure, we, therefore, used pECrPA_hfq630 plasmid containing the mini-array under inducible P_{tet} promoter. This strategy allowed us to successfully generate *hfq* deletion mutants in both *C. difficile* 630 Δ erm and epidemic *C. difficile* R20291 strains. The previous work showed that CRISPR repeats in 630 and R20291 strains have similar consensus sequences (Boudry et al., 2015). Moreover, both strains possess homologous complete and partial subtype I-B *cas* operons conserved in the majority of sequenced *C. difficile* strains (Boudry et al., 2015). Thus, the Cas machineries of the R20291 strain can successfully recognize and utilize crRNAs expressed from a 630-based mini-array. These results demonstrate that the artificial mini-array designed from the *C. difficile* 630 Δ erm CRISPR 16 leader and repeat sequences is suitable for targeting specific protospacer sequences on the bacterial chromosome and can be used for genome editing in at least two *C. difficile* strains. This may help to save time in the design of mini-array constructions. The general conservation of subtype I-B *cas* operons in *C. difficile* could suggest even more large application of the same targeting arrays suitable for the majority of *C. difficile* strains.

The deletion of the *hfq* gene was confirmed on mRNA and protein levels in both *C. difficile* 630 Δ erm and R20291 strains (Figure. 4.4B, C). Moreover, sporulation assay of the *C. difficile* 630 Δ erm Δ hfq strain revealed higher sporulation level than the wild type (Figure 4.5) that could be complemented by expressing *hfq* from the plasmid. These results are consistent with the previous observations for the Hfq-depleted strain (Boudry et al., 2014) and indicate that the harnessing of the endogenous CRISPR-Cas system can be effectively used to create deletion mutants in *C. difficile*.

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, this method does not need to express heterologous proteins inside *C. difficile* cells that may have toxic or other unpredictable effects. Localized on an editing plasmid, a mini-array mimics natural *C. difficile* CRISPR array and should not have undesirable impacts during genome manipulation. Secondly, this approach includes only one conjugation round and less plating steps saving significantly the time needed for the procedure completion

(Figure 4.6). In contrast, the *codA* allelic exchange method needs at least three more colony plating steps that result in three extra days for the experiment. Finally, the plasmid is readily lost after the editing process, preventing the spontaneous emergence of revertant strains.

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. In the study of Boudry et al., two PAMs (CCA and CCT) were experimentally confirmed for *C. difficile* CRISPR-Cas target recognition (Boudry et al., 2015). At the same time, general *in silico* analysis of CRISPR spacer homology to phage protospacers revealed a rather unconstrained PAM consensus YCN for *C. difficile* CRISPR-Cas system (Boudry et al., 2015), which was later confirmed by PAM libraries experiments (see Chapter 2). 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, that expands opportunities of finding functional PAM in the target region (Li et al., 2015).

The applications of endogenous CRISPR-Cas system for genome editing in *C. difficile* could be potentially larger than the generation of deletion mutants. This technique could be readily applied for introducing other types of mutations, i.e. point mutations and insertions (Li et al., 2015). For a point mutation, the homologous arms on the editing plasmid should be designed to introduce changes of functional PAM at the editing region to a non-functional motif. Alternatively, substitutions should be introduced into a seed region, the first eight nucleotides of the protospacer, crucial for CRISPR targeting (Semenova et al., 2011). As a priority choice, a point mutation design could be achieved by introducing changes at the first or second positions of PAM. Combining the changes within PAM and seed region could even increase the efficiency of editing as reported for other endogenous CRISPR-editing tools (Atmadjaja et al., 2019; Li et al., 2015). The previous work showed, that the non-functional PAM and mutation in the first position of protospacer within seed region abolished or considerably impaired the CRISPR interference (Boudry et al., 2015) (see Chapter 2). Genome insertions can be introduced by the homologous arms, designed to make a break in the integrity of chosen protospacer or/and PAM of the targeted genome sequence (Li et al., 2015) or insert a mutation to “knockout” the PAM (Atmadjaja et al., 2019).

Chapter 4. Using endogenous CRISPR-Cas system for genome editing in the human pathogen *C. difficile*.

The role of essential genes cannot be easily investigated since no deletion mutant could be generated. Therefore, a CRISPRi (CRISPR interference), which allows repressing the expression of target genes has been recently developed (Gross et al., 2016). This technology is primarily based on CRISPR-Cas9 systems with the mutated catalytic site of Cas9 protein (“catalytically dead Cas9”, dCas9) (Qi et al., 2013). The dCas9-based method has been already used in *C. difficile* (Müh et al., 2019). In addition, it was shown, that *E. coli* native subtype I-E CRISPR-Cas system lacking *cas3* could be repurposed for programmable transcriptional repression (Luo et al., 2015). Furthermore, a recent study showed that subtype I-B CRISPR-Cas system of *Haloferax volcanii* lacking *cas3* and *cas6* genes could be used for gene repression in this archaeon (Stachler and Marchfelder, 2016). Altogether, this data suggests that *C. difficile* native CRISPR-Cas system may be used for this goal too in a particular context. However, about 90% of sequenced *C. difficile* strains possess two subtype I-B *cas* operons each carrying *cas3* nuclease gene. An additional partial *cas* operon with *cas3* gene is present in the majority of MLST 3 group of *C. difficile* strains including the PCR ribotype 027 strains (Boudry et al., 2015). Thus, depending on the strain, the creation of double or triple *cas3* mutant background would be necessary to consider this CRISPRi method application.

CRISPR self-targeting could lead to bacterial cell death. This feature of CRISPR-Cas system can be applied for the development of new antimicrobial agents (Bikard and Barrangou, 2017). Among suggested strategies reside the use of phage particles and phagemids as vectors to deliver all the necessary auto-targeting CRISPR-Cas components inside the cell of a targeted pathogen (Bikard and Barrangou, 2017). In the present study, we showed an active killing of *C. difficile* cells by CRISPR self-targeting via expression of the mini-array from a plasmid vector. Therefore, in perspective, this approach could be promising for future developments of alternative strategies for *C. difficile* infection treatment.

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

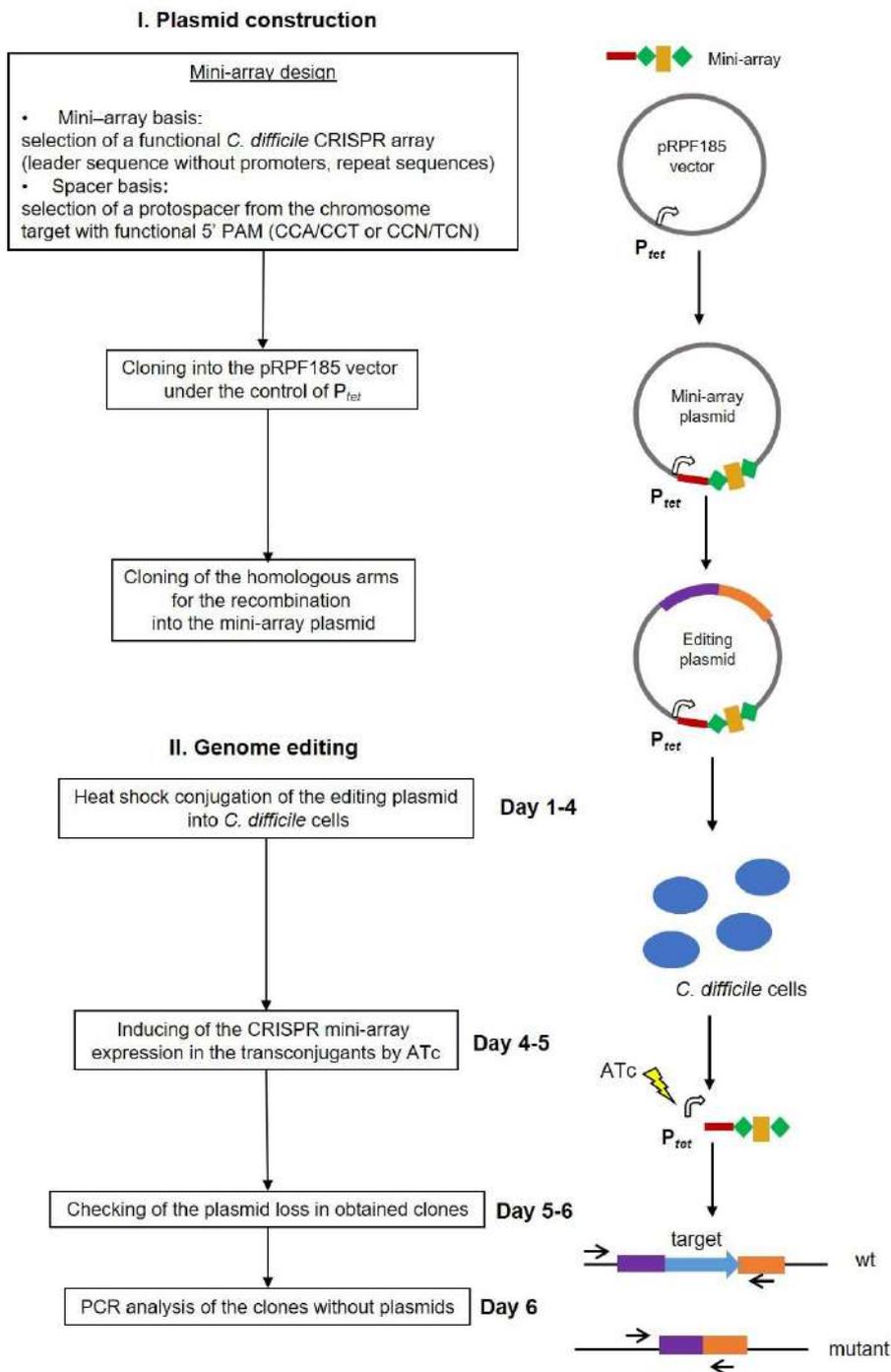


Figure 4.6. The general workflow for application of endogenous CRISPR-Cas-based genome editing method in *C. difficile*.

wt – wild type.

4.5 Supplementary materials

A

5' AATTATATGATAGGTTTTTTTATTAAGCATACTAGCTGGTGTATATCAGCTTATATTTTACAAAAATAAAAAACACCCAGACGCCAATAAGGGTGATTTTAA
 -10
 +1 → +1 →
 AAAAAATAATCTTTAAATCAATTTTGATGGAAATAGCTACTCTGTATAAAGTAAATATTTCCCTGGCTTTTATATACCACAAATGGTACAGATATTCAAAAAT
 AATATTTTATGATATAAAAAATGTAGAGATTTTGACGTGAGCAATATTCGATAAAATGAAGTTTAAACAATTGGAATACAAGGTATGAGGGTGTGTATAA
 ATGTATCAATTCACACTACTCTGCTGCTGCAAAATTTGAGAGAGGTGTGTATGTGTAGATATTGAAATACTAAGTTTATTTGGGGTTTTAGATTAACTATAT
 GGAATGTAATGTTACTATATATTAGTAAATGGAGTCAAGTTAAAGGTTTTAGATTAACTATATGGAATGTAAT 3'

B

5' TGAGCAATATTGCGCAATAAATGAAGTTTAAACAATTGGAATACAAGGTATGAGGGTGTGTATAAATGTTATCAATTCACACTACTCATGGTCTACTGCAAAAT
 TGAGAGAGGTGTGTATGTGTAGATATTGAAATACTAAGTTTATTTGGGGTTTTAGATTAACTATATGGAATGTAATGTTACTATATATTAGTAAATGGAGT
 CAAGTTAAAGGTTTTAGATTAACTATATGGAATGTAAT 3'

C

630Δerm left arm

5' CGATATTGAAATAAAAAAGTTATTGGAACAGCTAGAAAAGTGTGAGAATCCATTACATGCTCCTCATGGAAGACCTATAATGGTCGAAATATCTAAGACAGAAATAGAAAAATGT
 TTAAGAAGATATGATAAATATTTGGATATTTAAAATATATGGAAGGAATGTTATGAAAAAATACCACTTATAATACCTACAGGTCCAACCTGCTTGGAAAAACAGATTTGTC
 TAAAATAGCAAAAGATATGAATGCAGAGATAATTTCACTGATCTTATGCAAAATATAAATATATGGATATTGGAAGTGCAAAAGTACAGAGAAAGAGATGCAAGGAGTGAAGCA
 TTACTTTATAGATGAAGTAAAGCTGACTACTCAATTTTCAGTTTCAGAGTTCAACAAGAGCAAAAAATATATACATGAAATAAATAAAGGAAAAATGTTTTCAGTAACTGG
 GGCACAGGACTTTAATAACTCTTTAATATATAATATGGATTTGCAACTCTGATGCCAATAATGAATTAAGAGAAGAACCTCAAAAAACCTTGC TGAAGGGGAATAGATTACA
 TGCATAAATGTTAAAAGAACCTGATGAAAGATCTGCTAATAGAATCCATAAAAACACTAAAAGGGTCAAGAAGCTCTTGAAGTTTGTCTAAGTGGTAAAGAAATGAATGATTT
 TTCAGTGATTTAAAATTTAATGAAGAATATCAACCAATAAATAGTCTTAAA TAGAGATAGAGAACAATTTGTATCAAGAATAAATAAGAGTTGATATATGATTAATAAATGGA
 TTGGTTGAAGAGGTAAAAAATCTTTAAGTATGGGATTTAAAAAAGATATGATATCAATGCAAGGTATAGTTATAAAGAGATATAAAGTATCTGGATGGAGAAATATACATACGAA
 AAGCTATTGAAATAAATAAAGAGATCTAGAAAGTATGCAAAACGTCAGATAACATGGTTTAAAGATATAAACAAGCAAAATGGTTGATTTAGACCAATATGAAATATAGACGA
 ATTAATAAATGAAATAATCTTATATATAAAGATCTATAAATAAATATATAAATTTTAAAGTATACAACCTAATATTGAAATTTGTCATAAATATAGAATAAAGGTAGGAAT
 ATTTTGAAGTCAATTTAATCTTGGGAGGTACAAATCTA 3'

630Δerm right arm

5' ATAATTAATTAATTTAAGATGATTGAGAGGAGACTTACTAACCTATAAATATAATTTATTTGTTTTAGTAAAATAGTCTCCTTTACATGTTTTAAAATAAATAAAGAGAC
 AGGTGATAAATGCTTAATGAGACAAAAGAAATTTAAAAGACTACTATGGAATAGATGATGACTTTTAAACTATCGCAAGAAATAATGGAAGAAATAAAAGATAAGTTTGAAGAA
 ATAAAAGAAATAGAGAAATAAATAAATAAAGTTTTAAAAGCTATGCAAGAATCTAAACTAAGTGATATGCACTTTAACTGGACTACAGGATATGGATATAATGATATGGTCTGT
 AAAAAATAGAGAAATTTTATCTAAAGTATTAAATACAGAAGATGCTCTTGTAAAGCCAATAAATTTGAAATGGTACACATGCACTTACTTTATGATACAAAGGTAATGTTAGCCAGG
 AGATGAAATTTTATCTGTTACTGGTAGACCATATGATACTTTAGAGGGAGTTATAGGCATAAAGAGAAGAAAAAGGTTTCAATGAAAGAGATAGGTGTTACATATGATGATGATGATTT
 TTAGAGGATGGAACCTTAGATTTAGAGGAAATTAAGAAGCAAAATAAATGATAGAACAATAACTCGTTATGATACAAAGGTCAAAAGGATATCTTGGAGAAATCACTTTCAATTAGTG
 ATATAAAGAGAACTATAGAAAGTAAATAAGTCTGTAAACCAAGAAGCTATAGTAATGGTTGATGATTTGTTATGGTGAAGTTTTAGATACCAAAAGAGCTACTGATGTGGAGCAGATGT
 TATGGCTGGCTCTTAAATAAAAAATCTTGGTGGTGGCTTGCATTTGACAGGTGGAATATAGCTGGTGAAGAAAGATTTGATGAGCTTATATCTTATAGAATGACTCTCTCGGAATA
 GGAAAGAAATGGTCTTACTTTTGAACAACATGAAATGACTCTCAGGGTTTTTCTTGGCCCTTATATAGTATCTCAGGCTGTAATGGGAGCTATCTTTTGC TCAAGAGCTTTTG
 AAAAAATAGGATATGATGATTTACCAAAATATGATGACTTAAAGAAGTATATTATCAATGATTAGACTTAAATGCTGATGAAAGTAATAAGTTTTTGTGAAGGTATACAAGAAGC
 TGCACAGTCTGATTTTATGTAACCTGTTCC 3'

D

R20291 left arm

5' CGATATTGAAATAAAAAAGTTATTGGAACAGCTAGAAAAGTGTGAGAATCCATTACATGCTCCTCATGGAAGACCTATAATGGTCGAAATATCTAAGACAGAAATAGAAAAATGT
 TTAAGAAGATATGATAAATATTTGGATATTTAAAATATATGGAAGGAATGTTATGAAAAAATACCACTTATAATACCTACAGGTCCAACCTGCTTGGAAAAACAGATTTGTC
 TAAAATAGCAAAAGATATGAATGCAGAGATAATTTCACTGATCTTATGCAAAATATAAATATATGGATATTGGAAGTGCAAAAGTACAGAGAAAGAGATGCAAGGAGTGAAGCA
 TTACTTTATAGATGAAGTAAAGCTGACTACTCAATTTTCAGTTTCAGAGTTTCAACAAGAGCAAAAAATATATACATGAAATAAATAAAGGAAAAATGTTTTCAGTAACTGG
 GGCACAGGACTTTAATAACTCTTTAATATATAATATGGATTTGCAACTCTGATGCCAATAATGAATTAAGAGAAGAACCTCAAAAAACCTTGC TGAAGGGGAATAGATTACA
 TGCATAAATGTTAAAAGAACCTGATGAAAGATCTGCTAATAGAATCCATAAAAACACTAAAAGGGTCAAGAAGCTCTTGAAGTTTGTCTAAGTGGTAAAGAAATGAATGATTT
 TTCAGTGATTTAAAATTTAATGAAGAATATCAACCAATAAATAGTCTTAAA TAGAGATAGAGAACAATTTGTATCAAGAATAAATAAGAGTTGATATATGATTAATAAATGGA
 TTGGTTGAAGAGGTAAAAAATCTTTAAGTATGGGATTTAAAAAAGATATGATATCAATGCAAGGTATAGTTATAAAGAGATATAAAGTATCTGGATGGAGAAATATACATACGAA
 AAGCTATTGAAATAAATAAAGAGATCTAGAAAGTATGCAAAACGTCAGATAACATGGTTTAAAGATATAAACAAGCAAAATGGTTGATTTAGACCAATATGAAATATAGACGA
 ATTAATAAATGAAATAATCTTATATATAAAGATCTATAAATAAATATAAATTTTAAAGTATACAACCTAATATTGAAATTTGTCATAAATATAGAATAAAGGTAGGAAT
 ATTTTGAAGTCAATTTAATCTTGGGAGGTACAAATCTA 3'

R20291 right arm

5' ATAATTAATTAATTTAAGATGATTGAGAGGAGACTTACTAACCTATAAATATAATTTATTTGTTTTAGTAAAATAGTCTCCTTTACATGTTTTAAAATAAATAAAGAGAC
 AGGTGATAAATGCTTAATGAGACAAAAGAAATTTAAAAGACTACTATGGAATAGATGATGACTTTTAAACTATCGCAAGAAATAATGGAAGAAATAAAAGATAAGTTTGAAGAA
 ATAAAAGAAATAGAGAAATAAATAAATAAAGTTTTAAAAGCTATGCAAGAATCTAAACTAAGTGATATGCACTTTAACTGGACTACAGGATATGGATATAATGATATGGTCTGT
 AAAAAATAGAGAAATTTTATCTAAAGTATTAAATACAGAAGATGCTCTTGTAAAGCCAATAAATTTGAAATGGTACACATGCACTTACTTTATGATACAAAGGTAATGTTAGCCAGG
 AGATGAAATTTTATCTGTTACTGGTAGACCATATGATACTTTAGAGGGAGTTATAGGCATAAAGAGAAGAAAAAGGTTTCAATGAAAGAGATAGGTGTTACATATGATGATGATGATTT
 TTAGAGGATGGAACCTTAGATTTAGAGGAAATTAAGAAGCAAAATAAATGATAGAACAATAACTCGTTATGATACAAAGGTCAAAAGGATATCTTGGAGAAATCACTTTCAATTAGTG
 ATATAAAGAGAACTATAGAAAGTAAATAAGTCTGTAAACCAAGAAGCTATAGTAATGGTTGATGATTTGTTATGGTGAAGTTTTAGATACCAAAAGAGCTACTGATGTGGAGCAGATGT
 TATGGCTGGCTCTTAAATAAAAAATCTTGGTGGTGGCTTGCATTTGACAGGTGGAATATAGCTGGTGAAGAAAGATTTGATGAGCTTATATCTTATAGAATGACTCTCTCGGAATA
 GGAAAGAAATGGTCTTACTTTTGAACAACATGAAATGACTCTCAGGGTTTTTCTTGGCCCTTATATAGTATCTCAGGCTGTAATGGGAGCTATCTTTTGC TCAAGAGCTTTTG
 AAAAAATAGGATATGATGATTTACCAAAATATGATGACTTAAAGAAGTATATTATCAATGATTAGACTTAAATGCTGATGAAAGTAATAAGTTTTTGTGAAGGTATACAAGAAGC
 TGCACAGTCTGATTTTATGTAACCTGTTCC 3'

Figure S4.1. Sequences used to construct mini-array and editing plasmids to delete *hfq* gene in *C. difficile*. **A** – the sequence of the mini-array containing the full leader sequence with all the native promoters. **B** – the sequence of the mini-array containing the partial leader sequence. **C** – sequences of 1200-bp long regions flanking *hfq* gene of the *C. difficile* 630Δerm used as

(Figure S4.1. Continue) homologous arms. **D** – sequences of 1200-bp long regions flanking *hfq* gene of the *C. difficile* R202091 used as homologous arms. Repeat sequences in the mini-arrays are marked with green color and spacer sequence is marked with blue color. Promoters and transcriptional start sites (+1) are marked with dark red color.

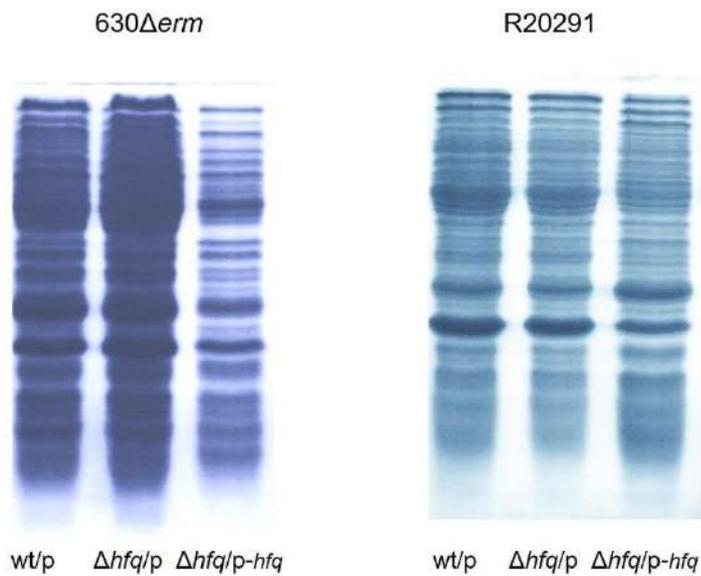


Figure S4.2. Protein gels stained with the InstantBlue and used as a loading control for the Western blot analysis of wt/p, Δ hfq-p, and Δ hfq/p-hfq *C. difficile* strains.

Chapter 5. Conclusions and perspectives

Clostridium difficile is an opportunistic enteropathogen and the main cause of nosocomial diarrhea in adults. During its infection cycle, the bacterium survives inside the complex colon communities possibly by using defense mechanisms of bacterial immune systems. CRISPR-Cas systems are prokaryotic adaptive defense systems against phages and other foreign genetic agents. *C. difficile* possesses I-B subtype CRISPR-Cas system with several unusual features: a large set of actively expressed arrays, some of them are localized inside prophage regions, and multiple *cas* operons (Boudry et al., 2015). This original CRISPR-Cas system may play a crucial role in *C. difficile* adaptation inside the host.

In this work, we have investigated all general functional aspects of *C. difficile* CRISPR-Cas system, and the main conclusions of this part of the research are as follows:

1. Enlarged PAM (YCN) sequences were identified for *C. difficile* 630 Δ *erm* and R20291 strains;
2. Active interference and different contribution to the defense of all 12 CRISPR arrays in 630 strain were demonstrated;
3. Active interference and enlarged PAMs (CCC/CCG, CCA/CCT) were experimentally confirmed for the R20291 strain;
4. The deletion of full *cas* operon did not completely abolish interference in the 630 Δ *erm* strain;
5. New spacer acquisition was demonstrated for 2 CRISPR arrays in 630 Δ *erm* strain, and the naïve adaptation seems not to be as active as interference in *C. difficile*.

Despite this, many characteristics of *C. difficile* CRISPR-Cas system still remain to be investigated. In particular, we identified the enlarged PAM (YCN) sequences, but an additional experimental verification of all the PAM nucleotide positions (especially TCN motifs) is still required. Additionally, the functionality of remaining *C. difficile* R20291 CRISPR arrays and the role of all the *cas* operons of both 630 Δ *erm* and R20291 strains in interference and *C. difficile* infection are yet to be assessed. Moreover, further detailed quantitative analysis of the CRISPR arrays expression is required for both strains for better understanding the link between CRISPR arrays transcriptional levels and their

different contribution to the defense. Interestingly, for at least three CRISPR arrays from *C. difficile* 630 Δ *erm* strain (CRISPR12, CRISPR15/16 and CRISPR17), the interference efficiency could be correlated with their expression level estimated by RNAseq, Northern blotting and qRT-PCR analysis. The general function of *C. difficile* CRISPR-Cas system during CDI is needed to be explored by additional experiments. In particular, the complete inactivation of the system by deletion of all *cas* operons will allow investigating the fitness of mutant strain inside the host using available animal models. Our results also revealed only the naïve type of CRISPR adaptation. Further experiments with phages and primed adaptation assays will enrich our knowledge about CRISPR immunization mechanisms in *C. difficile*. This work also raised a question about the function of the Cas4 protein in the process of new spacer acquisition and overall *C. difficile* physiology. Finally, it is not clear why *C. difficile* CRISPR-Cas system is not highly active in adaptation. It could be hypothesized, that uncharacterized anti-CRISPR proteins (Pawluk et al., 2018), potentially encoded inside the chromosomal prophage regions, may inhibit adaptation process. These interesting points should be explored in future studies.

Another goal of the present Ph.D. thesis was to study the mechanisms of *C. difficile* CRISPR-Cas system regulation, in particular, in biofilm conditions and in response to various stresses. During this work, we discovered a unique feature of this enteropathogenic CRISPR-Cas system – an association of type I toxin-antitoxin systems with CRISPR arrays in the majority of sequenced *C. difficile* strains. We chose two of these TA modules and investigated their functionality and a possible link with the CRISPR-Cas system regulation. The general outcomes of this study are as follows:

1. The functionality of type I TA systems was demonstrated with the growth arrest induced by toxin overexpression and the neutralization of toxins by antitoxins;
2. Co-regulation of CRISPR arrays and adjacent type I TA systems was suggested in biofilms and under stress conditions, potentially associated with the presence of *sigB* promoters.

However, the direct link between type I TA modules and CRISPR-Cas system function has not been shown. This aspect and other unanswered questions about the role of the TA modules in stress response, prophage stability, and stabilization of chromosomal regions carrying CRISPR arrays need to be explored by future studies.

We also investigated the role of the bacterial second messenger c-di-GMP in *C. difficile* CRISPR-Cas regulation. We found a c-di-GMP-dependent riboswitch associated with the CRISPR12 array in 630 Δ *erm* strain, which can indicate the direct impact of c-di-GMP-dependent regulation on this array function. In general, the global effect of c-di-GMP on the expression of other CRISPR-Cas system components has been explored. These experiments showed:

1. a slight induction of interference in *C. difficile* 630 Δ *erm* CRISPR12 by high levels of c-di-GMP;
2. induction of expression of both *cas* operons and several CRISPR arrays in *C. difficile* 630 Δ *erm* by high levels of c-di-GMP.

More detailed analysis of *C. difficile* CRISPR-Cas system regulation needs to be performed in the future, in particular, the regulation by other biofilm-related stimuli and stresses and the molecular mechanisms of these regulatory processes.

Study of *C. difficile* CRISPR-Cas system functionality allowed us to explore its biotechnological potential. In this Thesis, we described the application of native *C. difficile* CRISPR-Cas system as a novel tool for genome editing in this bacterium. Thus, Chapter 4 describes the utilization of endogenous *C. difficile* CRISPR-Cas system as a novel technique for genome editing in *C. difficile*. Conclusions of this part of the research are as follows:

1. The same CRISPR mini-array can be used in both 630 Δ *erm* and R20291 strains and could be probably extended to other *C. difficile* strains;
2. An efficient editing plasmid loss was demonstrated;
3. An efficient CRISPR autoimmunity was observed;
4. Δ *hfq* mutants in 630 Δ *erm* and R20291 strains were created, which could not be obtained using other genome editing methods before.

This new genome editing approach enlarges the set of genome editing tools available for *C. difficile*, and the method can be used in developing new antimicrobials against CDI.

The present Ph.D. Thesis extends our knowledge about *C. difficile* physiology and its genetic features, and also opens new perspectives in biotechnological applications of

C. difficile CRISPR-Cas system. The summary of this study conclusions and future research perspectives is presented in Figure 5.1.

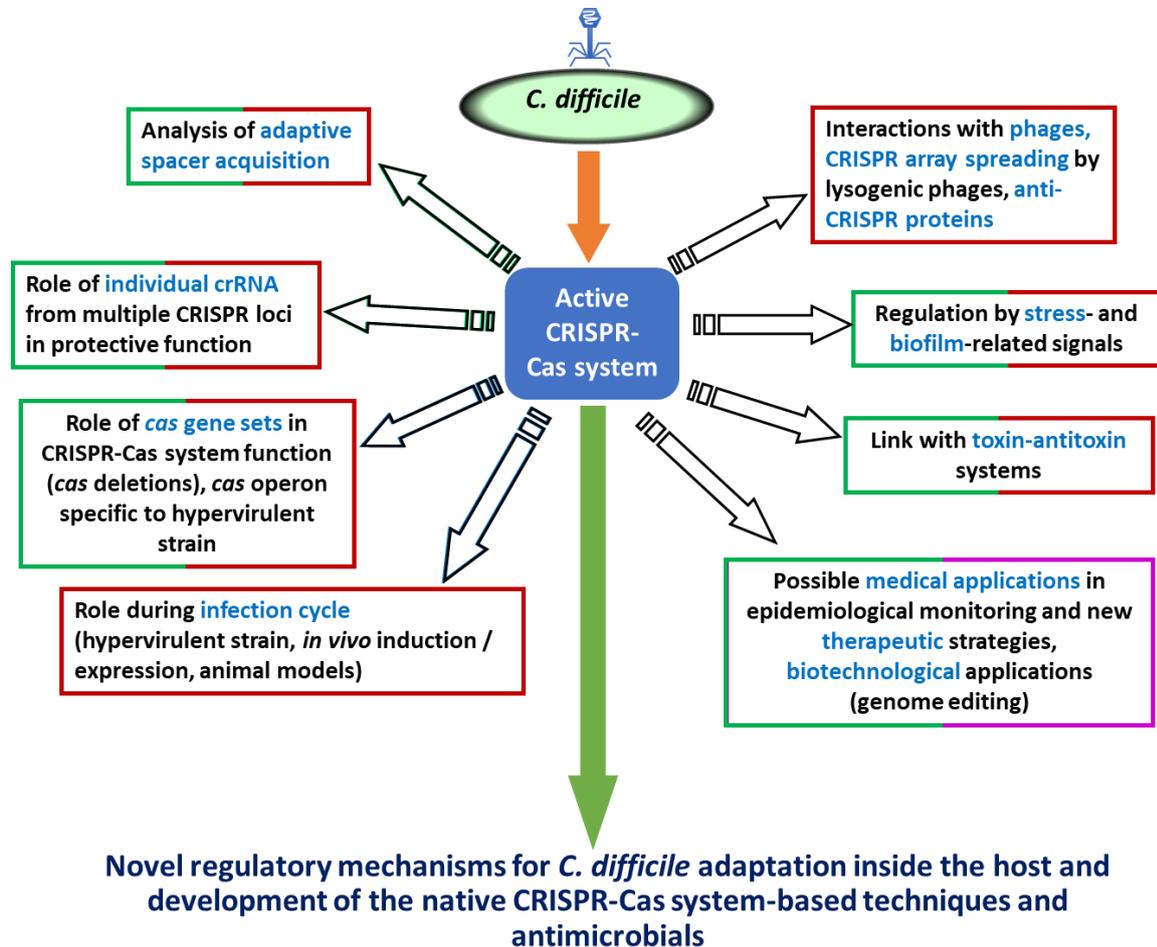


Figure 5.1. Conclusions and perspectives of *C. difficile* CRISPR-Cas system research. Dark green color designates results, obtained in this PhD Thesis, dark red color designates further perspectives of *C. difficile* CRISPR-Cas system study, and purple color designates possible applications of *C. difficile* CRISPR-Cas system.

Bibliography

- Abedon, S. (2012). Spatial vulnerability: bacterial arrangements, microcolonies, and biofilms as responses to low rather than high phage densities. *Viruses* 4, 663–687. doi:10.3390/v4050663.
- Abt, M. C., McKenney, P. T., and Pamer, E. G. (2016). *Clostridium difficile* colitis: pathogenesis and host defence. *Nat. Rev. Microbiol.* 14, 609–620. doi:10.1038/nrmicro.2016.108.
- Amitai, G., and Sorek, R. (2016). CRISPR–Cas adaptation: insights into the mechanism of action. *Nat. Rev. Microbiol.* advance on, 67–76. doi:10.1038/nrmicro.2015.14.
- Andersen, J. M., Shoup, M., Robinson, C., Britton, R., Olsen, K. E. P., and Barrangou, R. (2016). CRISPR diversity and microevolution in *Clostridium difficile*. *Genome Biol. Evol.* 8, 2841–2855. doi:10.1093/gbe/evw203.
- André, G., Even, S., Putzer, H., Burguière, P., Croux, C., Danchin, A., et al. (2008). S-box and T-box riboswitches and antisense RNA control a sulfur metabolic operon of *Clostridium acetobutylicum*. *Nucleic Acids Res.* 36, 5955–5969. doi:10.1093/nar/gkn601.
- Antunes, A., Camiade, E., Monot, M., Courtois, E., Barbut, F., Sernova, N. V., et al. (2012). Global transcriptional control by glucose and carbon regulator CcpA in *Clostridium difficile*. *Nucleic Acids Res.* 40, 10701–10718. doi:10.1093/nar/gks864.
- Arakawa, K., and Tomita, M. (2007). Selection effects on the positioning of genes and gene structures from the interplay of replication and transcription in bacterial genomes. *Evol. Bioinform. Online* 3, 279–86.
- Atmadjaja, A. N., Holby, V., Harding, A. J., Krabben, P., Smith, H. K., and Jenkinson, E. R. (2019). CRISPR-Cas, a highly effective tool for genome editing in *Clostridium saccharoperbutylacetonicum* N1-4(HMT). *FEMS Microbiol. Lett.* 366, fnz059. doi:10.1093/femsle/fnz059.
- Babic, A., Berkmen, M. B., Lee, C. A., and Grossman, A. D. (2011). Efficient gene transfer in bacterial cell chains. *MBio* 2, e00027-11. doi:10.1128/mBio.00027-11.
- Banawas, S. S. (2018). *Clostridium difficile* infections: a global overview of drug sensitivity and resistance mechanisms. *Biomed Res. Int.* 2018, 1–9. doi:10.1155/2018/8414257.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–77. doi:10.1089/cmb.2012.0021.
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., et al. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709–12.

doi:10.1126/science.1138140.

- Barrangou, R., and Horvath, P. (2017). A decade of discovery: CRISPR functions and applications. *Nat. Microbiol.* 2, 17092. doi:10.1038/nmicrobiol.2017.92.
- Barros, M. P. S., França, C. T., Lins, R. H. F. B., Santos, M. D. V., Silva, E. J., Oliveira, M. B. M., et al. (2014). Dynamics of CRISPR loci in microevolutionary process of *Yersinia pestis* strains. *PLoS One* 9, e108353. doi:10.1371/journal.pone.0108353.
- Bartlett, J. G., Chang, T. W., Moon, N., and Onderdonk, A. B. (1978). 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, 1525–30.
- Bauer, M. P., Notermans, D. W., van Benthem, B. H., Brazier, J. S., Wilcox, M. H., Rupnik, M., et al. (2011). *Clostridium difficile* infection in Europe: a hospital-based survey. *Lancet* 377, 63–73. doi:10.1016/S0140-6736(10)61266-4.
- Bertani, G. (1951). Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J. Bacteriol.* 62, 293–300.
- Bhaya, D., Davison, M., and Barrangou, R. (2011). CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu. Rev. Genet.* 45, 273–297. doi:10.1146/annurev-genet-110410-132430.
- Bikard, D., and Barrangou, R. (2017). Using CRISPR-Cas systems as antimicrobials. *Curr. Opin. Microbiol.* 37, 155–160. doi:10.1016/j.mib.2017.08.005.
- Bikard, D., Hatoum-Aslan, A., Mucida, D., and Marraffini, L. A. (2012). CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. *Cell Host Microbe* 12, 177–186. doi:10.1016/j.chom.2012.06.003.
- Bloom-Ackermann, Z., Steinberg, N., Rosenberg, G., Oppenheimer-Shaanan, Y., Pollack, D., Ely, S., et al. (2016). Toxin-Antitoxin systems eliminate defective cells and preserve symmetry in *Bacillus subtilis* biofilms. *Environ. Microbiol.* 18, 5032–5047. doi:10.1111/1462-2920.13471.
- Blosser, T. R., Loeff, L., Westra, E. R., Vlot, M., Künne, T., Sobota, M., et al. (2015). Two distinct DNA binding modes guide dual roles of a CRISPR-Cas protein complex. *Mol. Cell* 58, 60–70. doi:10.1016/j.molcel.2015.01.028.
- Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S. D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 151, 2551–61. doi:10.1099/mic.0.28048-0.
- Bordeleau, E., Fortier, L.-C., Malouin, F., and Burrus, V. (2011). c-di-GMP turn-over in *Clostridium difficile* is controlled by a plethora of diguanylate cyclases and phosphodiesterases. *PLoS Genet.* 7, e1002039. doi:10.1371/journal.pgen.1002039.

- Bordeleau, E., Purcell, E. B., Lafontaine, D. A., Fortier, L.-C., Tamayo, R., and Burrus, V. (2015). Cyclic di-GMP riboswitch-regulated type IV pili contribute to aggregation of *Clostridium difficile*. *J. Bacteriol.* 197, 819–832. doi:10.1128/JB.02340-14.
- Boudry, P., Gracia, C., Monot, M., Caillet, J., Saujet, L., Hajnsdorf, E., et al. (2014). Pleiotropic role of the RNA chaperone protein Hfq in the human pathogen *Clostridium difficile*. *J. Bacteriol.* 196, 3234–3248. doi:10.1128/JB.01923-14.
- Boudry, P., Semenova, E., Monot, M., Datsenko, K. A., Lopatina, A., Sekulovic, O., et al. (2015). Function of the CRISPR-cas system of the human pathogen *Clostridium difficile*. *MBio* 6, e01112-15. doi:10.1128/mBio.01112-15.
- Bourgogne, A., Garsin, D. A., Qin, X., Singh, K. V., Sillanpaa, J., Yerrapragada, S., et al. (2008). Large scale variation in *Enterococcus faecalis* illustrated by the genome analysis of strain OG1RF. *Genome Biol.* 9, R110. doi:10.1186/gb-2008-9-7-r110.
- Brantl, S. (2007). Regulatory mechanisms employed by *cis*-encoded antisense RNAs. *Curr. Opin. Microbiol.* 10, 102–109. doi:10.1016/j.mib.2007.03.012.
- Brantl, S. (2012a). Acting antisense: plasmid- and chromosome-encoded sRNAs from Gram-positive bacteria. *Future Microbiol.* 7, 853–871. doi:10.2217/fmb.12.59.
- Brantl, S. (2012b). Bacterial type I toxin-antitoxin systems. *RNA Biol.* 9, 1488–1490. doi:10.4161/rna.23045.
- Brantl, S., and Jahn, N. (2015). sRNAs in bacterial type I and type III toxin-antitoxin systems. *FEMS Microbiol. Rev.* 39, 413–427. doi:10.1093/femsre/fuv003.
- Brantl, S., and Jahn, N. (2016). Heat-shock-induced refolding entails rapid degradation of *bsrG* toxin mRNA by RNases Y and J1. *Microbiology* 162, 590–599. doi:10.1099/mic.0.000247.
- Braun, V., Hundsberger, T., Leukel, P., Sauerborn, M., and von Eichel-Streiber, C. (1996). Definition of the single integration site of the pathogenicity locus in *Clostridium difficile*. *Gene* 181, 29–38. doi:10.1016/s0378-1119(96)00398-8.
- Brouns, S. J. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J. H., Snijders, A. P. L., et al. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* (80-.). 321, 960–4. doi:10.1126/science.1159689.
- Brown, K. A., Khanafer, N., Daneman, N., and Fisman, D. N. (2013). Meta-analysis of antibiotics and the risk of community-associated *Clostridium difficile* infection. *Antimicrob. Agents Chemother.* 57, 2326–2332. doi:10.1128/AAC.02176-12.
- Brown, S. D., Nagaraju, S., Utturkar, S., De Tissera, S., Segovia, S., Mitchell, W., et al. (2014). Comparison of single-molecule sequencing and hybrid approaches for finishing the genome of *Clostridium autoethanogenum* and analysis of CRISPR systems in industrial relevant

- Clostridia. *Biotechnol. Biofuels* 7, 40. doi:10.1186/1754-6834-7-40.
- Cairns, M. D., Preston, M. D., Hall, C. L., Gerding, D. N., Hawkey, P. M., Kato, H., et al. (2017). Comparative genome analysis and global phylogeny of the toxin variant *Clostridium difficile* PCR ribotype 017 reveals the evolution of two independent sublineages. *J. Clin. Microbiol.* 55, 865–876. doi:10.1128/JCM.01296-16.
- Calabi, E., Calabi, F., Phillips, A. D., and Fairweather, N. F. (2002). Binding of *Clostridium difficile* surface layer proteins to gastrointestinal tissues. *Infect. Immun.* 70, 5770–8. doi:10.1128/iai.70.10.5770-5778.2002.
- Carroll, K. C., and Bartlett, J. G. (2011). Biology of *Clostridium difficile* : implications for epidemiology and diagnosis. *Annu. Rev. Microbiol.* 65, 501–521. doi:10.1146/annurev-micro-090110-102824.
- Carter, G. P., Lyras, D., Allen, D. L., Mackin, K. E., Howarth, P. M., O'Connor, J. R., et al. (2007). Binary toxin production in *Clostridium difficile* is regulated by CdtR, a LytTR family response regulator. *J. Bacteriol.* 189, 7290–301. doi:10.1128/JB.00731-07.
- Carter, G. P., Rood, J. I., and Lyras, D. (2012). The role of toxin A and toxin B in the virulence of *Clostridium difficile*. *Trends Microbiol.* 20, 21–29. doi:10.1016/j.tim.2011.11.003.
- Cartman, S. T., Kelly, M. L., Heeg, D., Heap, J. T., and Minton, N. P. (2012). 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. doi:10.1128/AEM.00249-12.
- Charpentier, E., Richter, H., van der Oost, J., and White, M. F. (2015). Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. *FEMS Microbiol. Rev.*, 428–441. doi:10.1093/femsre/fuv023.
- Chen, A. G. Y., Sudarsan, N., and Breaker, R. R. (2011). Mechanism for gene control by a natural allosteric group I ribozyme. *RNA* 17, 1967–1972. doi:10.1261/rna.2757311.
- Cheng, F., Gong, L., Zhao, D., Yang, H., Zhou, J., Li, M., et al. (2017). Harnessing the native type I-B CRISPR-Cas for genome editing in a polyploid archaeon. *J. Genet. Genomics* 44, 541–548. doi:10.1016/j.jgg.2017.09.010.
- Cohen, J. E., Wang, R., Shen, R.-F., Wu, W. W., and Keller, J. E. (2017). Comparative pathogenomics of *Clostridium tetani*. *PLoS One* 12, e0182909. doi:10.1371/journal.pone.0182909.
- Coignard, B., Barbut, F., Blanckaert, K., Thiolet, J. M., Poujol, I., Carbonne, A., et al. (2006). Emergence of *Clostridium difficile* toxinotype III, PCR-ribotype 027-associated disease, France, 2006. *Eurosurveillance J.* 11, 3044. doi:10.2807/esw.11.37.03044-en.

- Collignon, A., Bourlioux, P., Waligora-Dupriet, A.-J., Karjalainen, T., Barc, M.-C., Hennequin, C., et al. (2001). GroEL (Hsp60) of *Clostridium difficile* is involved in cell adherence. *Microbiology* 147, 87–96. doi:10.1099/00221287-147-1-87.
- Collins, T. J. (2007). ImageJ for microscopy. *Biotechniques* 43, S25–S30. doi:10.2144/000112517.
- Coray, D. S., Wheeler, N. E., Heinemann, J. A., and Gardner, P. P. (2017). Why so narrow: Distribution of anti-sense regulated, type I toxin-antitoxin systems compared with type II and type III systems. *RNA Biol.* 14, 275–280. doi:10.1080/15476286.2016.1272747.
- Crooks, G. E., Hon, G., Chandonia, J.-M., and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–90. doi:10.1101/gr.849004.
- Crowther, G. S., Chilton, C. H., Todhunter, S. L., Nicholson, S., Freeman, J., Baines, S. D., et al. (2014). Comparison of planktonic and biofilm-associated communities of *Clostridium difficile* and indigenous gut microbiota in a triple-stage chemostat gut model. *J. Antimicrob. Chemother.* 69, 2137–2147. doi:10.1093/jac/dku116.
- Cui, Y., Li, Y., Gorgé, O., Platonov, M. E., Yan, Y., Guo, Z., et al. (2008). Insight into microevolution of *Yersinia pestis* by Clustered Regularly Interspaced Short Palindromic Repeats. *PLoS One* 3, e2652. doi:10.1371/journal.pone.0002652.
- Dapa, T., Leuzzi, R., Ng, Y. K., Baban, S. T., Adamo, R., Kuehne, S. A., et al. (2013). Multiple factors modulate biofilm formation by the anaerobic pathogen *Clostridium difficile*. *J. Bacteriol.* 195, 545–555. doi:10.1128/JB.01980-12.
- Datsenko, K. A., Pougach, K., Tikhonov, A., Wanner, B. L., Severinov, K., and Semenova, E. (2012). Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat. Commun.* 3, 945. doi:10.1038/ncomms1937.
- Dawson, L. F., Valiente, E., Faulds-Pain, A., Donahue, E. H., and Wren, B. W. (2012). Characterisation of *Clostridium difficile* biofilm formation, a role for Spo0A. *PLoS One* 7, e50527. doi:10.1371/journal.pone.0050527.
- Dembek, M., Barquist, L., Boinett, C. J., Cain, A. K., Mayho, M., Lawley, T. D., et al. (2015). High-throughput analysis of gene essentiality and sporulation in *Clostridium difficile*. *MBio* 6, e02383. doi:10.1128/mBio.02383-14.
- Denève, C., Janoir, C., Poilane, I., Fantinato, C., and Collignon, A. (2009). New trends in *Clostridium difficile* virulence and pathogenesis. *Int. J. Antimicrob. Agents* 33, S24–S28. doi:10.1016/S0924-8579(09)70012-3.
- Dillingham, M. S., and Kowalczykowski, S. C. (2008). RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiol. Mol. Biol. Rev.* 72, 642–671. doi:10.1128/MMBR.00020-08.

- Dineen, S. S., Villapakkam, A. C., Nordman, J. T., and Sonenshein, A. L. (2007). Repression of *Clostridium difficile* toxin gene expression by CodY. *Mol. Microbiol.* 66, 206–219. doi:10.1111/j.1365-2958.2007.05906.x.
- Dingle, K. E., Didelot, X., Ansari, M. A., Eyre, D. W., Vaughan, A., Griffiths, D., et al. (2013). 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. doi:10.1093/infdis/jis734.
- Dingle, K. E., Elliott, B., Robinson, E., Griffiths, D., Eyre, D. W., Stoesser, N., et al. (2014). Evolutionary history of the *Clostridium difficile* pathogenicity locus. *Genome Biol. Evol.* 6, 36–52. doi:10.1093/gbe/evt204.
- Domka, J., Lee, J., Bansal, T., and Wood, T. K. (2007). Temporal gene-expression in *Escherichia coli* K-12 biofilms. *Environ. Microbiol.* 9, 332–346. doi:10.1111/j.1462-2920.2006.01143.x.
- Donegan, N. P., and Cheung, A. L. (2009). Regulation of the mazEF toxin-antitoxin module in *Staphylococcus aureus* and its impact on sigB expression. *J. Bacteriol.* 191, 2795–2805. doi:10.1128/JB.01713-08.
- Dupuis, M.-È., Villion, M., Magadán, A. H., and Moineau, S. (2013). CRISPR-Cas and restriction-modification systems are compatible and increase phage resistance. *Nat. Commun.* 4, 2087. doi:10.1038/ncomms3087.
- Dupuy, B., and Sonenshein, A. L. (1998). Regulated transcription of *Clostridium difficile* toxin genes. *Mol. Microbiol.* 27, 107–120. doi:10.1046/j.1365-2958.1998.00663.x.
- Durand, S., Gilet, L., and Condon, C. (2012a). The essential function of *B. subtilis* RNase III is to silence foreign toxin genes. *PLoS Genet.* 8, e1003181. doi:10.1371/journal.pgen.1003181.
- Durand, S., Jahn, N., Condon, C., and Brantl, S. (2012b). Type I toxin-antitoxin systems in *Bacillus subtilis*. *RNA Biol.* 9, 1491–1497. doi:10.4161/rna.22358.
- Eyre, D. W., Cule, M. L., Wilson, D. J., Griffiths, D., Vaughan, A., O'Connor, L., et al. (2013). Diverse sources of *C. difficile* infection identified on whole-genome sequencing. *N. Engl. J. Med.* 369, 1195–1205. doi:10.1056/NEJMoa1216064.
- Fagan, R. P., and Fairweather, N. F. (2011). *Clostridium difficile* has two parallel and essential Sec secretion systems. *J. Biol. Chem.* 286, 27483–27493. doi:10.1074/JBC.M111.263889.
- Fire, A., and Mello, C. (2006). The Nobel Prize in Physiology or Medicine 2006. Available at: <https://www.nobelprize.org/prizes/medicine/2006/summary/>.
- Fozo, E. M., Makarova, K. S., Shabalina, S. A., Yutin, N., Koonin, E. V., and Storz, G. (2010). Abundance of type I toxin-antitoxin systems in bacteria: searches for new candidates and discovery of novel families. *Nucleic Acids Res.* 38, 3743–59. doi:10.1093/nar/gkq054.

- Garneau, J. E., Dupuis, M.-V., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., et al. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468, 67–71. doi:10.1038/nature09523.
- Garnier, F., Taourit, S., Glaser, P., Courvalin, P., and Galimand, M. (2000). Characterization of transposon Tn1549, conferring VanB-type resistance in *Enterococcus spp.* *Microbiology* 146, 1481–1489.
- Georgiades, K., and Raoult, D. (2011). Genomes of the most dangerous epidemic bacteria have a virulence repertoire characterized by fewer genes but more toxin-antitoxin modules. *PLoS One* 6, e17962. doi:10.1371/journal.pone.0017962.
- Gerdes, K., Christensen, S. K., and Løbner-Olesen, A. (2005). Prokaryotic toxin–antitoxin stress response loci. *Nat. Rev. Microbiol.* 3, 371–382. doi:10.1038/nrmicro1147.
- Gerdes, K., and Maisonneuve, E. (2012). Bacterial persistence and toxin-antitoxin loci. *Annu. Rev. Microbiol.* 66, 103–123. doi:10.1146/annurev-micro-092611-150159.
- Gerding, D. N., Johnson, S., Rupnik, M., and Aktories, K. (2014). *Clostridium difficile* binary toxin CDT. *Gut Microbes* 5, 15–27. doi:10.4161/gmic.26854.
- Ghigo, J.-M. (2001). Natural conjugative plasmids induce bacterial biofilm development. *Nature* 412, 442–445. doi:10.1038/35086581.
- Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345. doi:10.1038/nmeth.1318.
- Gil, F., Pizarro-Guajardo, M., Álvarez, R., Garavaglia, M., and Paredes-Sabja, D. (2015). *Clostridium difficile* recurrent infection: possible implication of TA systems. *Future Microbiol.* 10, 1649–57. doi:10.2217/fmb.15.94.
- Gomaa, A. A., Klumpe, H. E., Luo, M. L., Selle, K., Barrangou, R., and Beisel, C. L. (2014). Programmable removal of bacterial strains by use of genome-targeting CRISPR-Cas systems. *MBio* 5, e00928-13. doi:10.1128/mBio.00928-13.
- Govind, R., and Dupuy, B. (2012). Secretion of *Clostridium difficile* toxins A and B requires the holin-like protein TcdE. *PLoS Pathog.* 8, e1002727. doi:10.1371/journal.ppat.1002727.
- Gripenland, J., Netterling, S., Loh, E., Tiensuu, T., Toledo-Arana, A., and Johansson, J. (2010). RNAs: regulators of bacterial virulence. *Nat. Rev. Microbiol.* 8, 857–866. doi:10.1038/nrmicro2457.
- Grissa, I., Vergnaud, G., and Pourcel, C. (2007). The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* 8, 172. doi:10.1186/1471-2105-8-172.

- Groenen, P. M., Bunschoten, A. E., van Soolingen, D., and van Embden, J. D. (1993). Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. *Mol. Microbiol.* 10, 1057–65.
- Gross, C., Peters, J. M., Ph, D., Peters, J. M., Ph, D., Larson, M. H., et al. (2016). A comprehensive, CRISPR-based functional analysis of essential genes in bacteria. *Cell* 165, 1493–1506. doi:10.1016/j.cell.2016.05.003.
- Hall-Stoodley, L., Costerton, J. W., and Stoodley, P. (2004). Bacterial biofilms: from the natural environment to infectious diseases. *Nat. Rev. Microbiol.* 2, 95–108. doi:10.1038/nrmicro821.
- Hall, I. C., and O’Toole, E. (1935). Intestinal flora in new-born infants: with a description of a new pathogenic anaerobe, *Bacillus difficilis*. *Am. J. Dis. Child.* 49, 390. doi:10.1001/archpedi.1935.01970020105010.
- Hargreaves, K. R., and Clokie, M. R. J. (2014). *Clostridium difficile* phages: Still difficult? *Front. Microbiol.* 5, 1–14. doi:10.3389/fmicb.2014.00184.
- Hargreaves, K. R., Flores, C. O., Lawley, T. D., Clokie, R. J., Trevor, D., Hargreaves, K. R., et al. (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, 1–10. doi:10.1128/mBio.01045-13.Invited.
- Hayes, F. (2003). Toxins-Antitoxins: Plasmid maintenance, programmed cell death, and cell cycle arrest. *Science (80-)*. 301, 1496–1499. doi:10.1126/science.1088157.
- He, M., Miyajima, F., Roberts, P., Ellison, L., Pickard, D. J., Martin, M. J., et al. (2013). Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*. *Nat. Genet.* 45, 109–113. doi:10.1038/ng.2478.
- He, M., Sebahia, M., Lawley, T. D., Stabler, R. A., Dawson, L. F., Martin, M. J., et al. (2010). Evolutionary dynamics of *Clostridium difficile* over short and long time scales. *Proc. Natl. Acad. Sci. U. S. A.* 107, 7527–32. doi:10.1073/pnas.0914322107.
- Heap, J. T., Kuehne, S. A., Ehsaan, M., Cartman, S. T., Cooksley, C. M., Scott, J. C., et al. (2010). The Clostron: mutagenesis in *Clostridium* refined and streamlined. *J. Microbiol. Methods* 80, 49–55. doi:10.1016/J.MIMET.2009.10.018.
- Hennequin, C., Janoir, C., Barc, M.-C., Collignon, A., and Karjalainen, T. (2003). Identification and characterization of a fibronectin-binding protein from *Clostridium difficile*. *Microbiology* 149, 2779–2787. doi:10.1099/mic.0.26145-0.
- Heussler, G. E., and O’Toole, G. A. (2016). Friendly fire: biological functions and consequences of chromosomal targeting by CRISPR-Cas systems. *J. Bacteriol.* 198, 1481–1486. doi:10.1128/JB.00086-16.

- Hochstrasser, M. L., Taylor, D. W., Kornfeld, J. E., Nogales, E., and Doudna, J. A. (2016). DNA targeting by a minimal CRISPR RNA-guided Cascade. *Mol. Cell* 63, 840–851. doi:10.1016/j.molcel.2016.07.027.
- Hong, W., Zhang, J., Cui, G., Wang, L., and 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. doi:10.1021/acssynbio.8b00087.
- Horvath, P., Romero, D. A., Coute-Monvoisin, A.-C., Richards, M., Deveau, H., Moineau, S., et al. (2008). Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J. Bacteriol.* 190, 1401–1412. doi:10.1128/JB.01415-07.
- Høyland-Kroghsbo, N. M., Paczkowski, J., Mukherjee, S., Broniewski, J., Westra, E., Bondy-Denomy, J., et al. (2016). Quorum sensing controls the *Pseudomonas aeruginosa* CRISPR-Cas adaptive immune system. *Proc. Natl. Acad. Sci.*, 201617415. doi:10.1073/pnas.1617415113.
- Hsu, P. D., Lander, E. S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157, 1262–1278. doi:10.1016/J.CELL.2014.05.010.
- Hussain, H. A., Roberts, A. P., and 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. doi:10.1099/jmm.0.45790-0.
- Inés, I., Cañ, C., Groothuis, D., Zygouropoulou, M., Rodrigues, R., and Minton, N. P. (2019). RiboCas: A universal CRISPR-based editing tool for *Clostridium*. *ACS Synth. Biol.* 8, 1379–1390. doi:10.1021/acssynbio.9b00075.
- Ingle, P., Groothuis, D., Rowe, P., Huang, H., Cockayne, A., Kuehne, S. A., et al. (2019). Generation of a fully erythromycin-sensitive strain of *Clostridioides difficile* using a novel CRISPR-Cas9 genome editing system. *Sci. Rep.* 9, 8123. doi:10.1038/s41598-019-44458-y.
- Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakata, A. (1987). Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* 169, 5429–5433. doi:10.1128/jb.169.12.5429-5433.1987.
- Jackson, S. A., McKenzie, R. E., Fagerlund, R. D., Kieper, S. N., Fineran, P. C., and Brouns, S. J. J. (2017). CRISPR-Cas: Adapting to change. *Science* 356, eaal5056. doi:10.1126/science.aal5056.
- Jahn, N., Brantl, S., and Strahl, H. (2015). Against the mainstream: The membrane-associated type I toxin BsrG from *Bacillus subtilis* interferes with cell envelope biosynthesis without

- increasing membrane permeability. *Mol. Microbiol.* 98, 651–666. doi:10.1111/mmi.13146.
- Jahn, N., Preis, H., Wiedemann, C., and Brantl, S. (2012). BsrG/SR4 from *Bacillus subtilis*- the first temperature-dependent type I toxin-antitoxin system. *Mol. Microbiol.* 83, 579–598. doi:10.1111/j.1365-2958.2011.07952.x.
- Janoir, C. (2016). Virulence factors of *Clostridium difficile* and their role during infection. *Anaerobe* 37, 13–24. doi:10.1016/j.anaerobe.2015.10.009.
- Jansen, R., Embden, J. D. A. van, Gaastra, W., and Schouls, L. M. (2002a). Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* 43, 1565–75.
- Jansen, R., van Embden, J. D. A., Gaastra, W., and Schouls, L. M. (2002b). Identification of a novel family of sequence repeats among prokaryotes. *Omi. A J. Integr. Biol.* 6, 23–33. doi:10.1089/15362310252780816.
- Jore, M. M., Lundgren, M., van Duijn, E., Bultema, J. B., Westra, E. R., Waghmare, S. P., et al. (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat. Struct. Mol. Biol.* 18, 529–536. doi:10.1038/nsmb.2019.
- Just, I., Selzer, J., Wilm, M., Eichel-Streiber, C. von, Mann, M., and Aktories, K. (1995). Glucosylation of Rho proteins by *Clostridium difficile* toxin B. *Nature* 375, 500–503. doi:10.1038/375500a0.
- Kawano, M. (2012). Divergently overlapping *cis*-encoded antisense RNA regulating toxin-antitoxin systems from *E. coli*. *RNA Biol.* 9, 1520–1527. doi:10.4161/rna.22757.
- Kawano, M., Aravind, L., and Storz, G. (2007). An antisense RNA controls synthesis of an SOS-induced toxin evolved from an antitoxin. *Mol. Microbiol.* 64, 738–54. doi:10.1111/j.1365-2958.2007.05688.x.
- Kazanowski, M., Smolarek, S., Kinnarney, F., and Grzebieniak, Z. (2014). *Clostridium difficile*: epidemiology, diagnostic and therapeutic possibilities—a systematic review. *Tech. Coloproctol.* 18, 223–232. doi:10.1007/s10151-013-1081-0.
- Kieper, S. N., Almendros, C., Behler, J., McKenzie, R. E., Nobrega, F. L., Haagsma, A. C., et al. (2018). Cas4 facilitates PAM-compatible spacer selection during CRISPR adaptation. *Cell Rep.* 22, 3377–3384. doi:10.1016/J.CELREP.2018.02.103.
- Kint, N., Janoir, C., Monot, M., Hoys, S., Soutourina, O., Dupuy, B., et al. (2017). The alternative sigma factor σ B plays a crucial role in adaptive strategies of *Clostridium difficile* during gut infection. *Environ. Microbiol.* 19, 1933–1958. doi:10.1111/1462-2920.13696.
- Kirk, J. A., and Fagan, R. P. (2016). Heat shock increases conjugation efficiency in *Clostridium difficile*. *Anaerobe* 42, 1–5. doi:10.1016/j.anaerobe.2016.06.009.
- Koonin, E. V. (2017). Evolution of RNA- and DNA-guided antiviral defense systems in

- prokaryotes and eukaryotes: common ancestry vs convergence. *Biol. Direct* 12, 5.
doi:10.1186/s13062-017-0177-2.
- Koonin, E. V., and Zhang, F. (2016). Coupling immunity and programmed cell suicide in prokaryotes: Life-or-death choices. *BioEssays* 1600186, 1–9. doi:10.1002/bies.201600186.
- Koonin, E. V., Makarova, K. S., and Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.* 37, 67–78. doi:10.1016/j.mib.2017.05.008.
- Koskenniemi, K., Laakso, K., Koponen, J., Kankainen, M., Greco, D., Auvinen, P., et al. (2011). Proteomics and transcriptomics characterization of bile stress response in probiotic *Lactobacillus rhamnosus* GG. *Mol. Cell. Proteomics* 10, M110.002741.
doi:10.1074/mcp.M110.002741.
- Kuehne, S. A., Heap, J. T., Cooksley, C. M., Cartman, S. T., and Minton, N. P. (2011). “ClosTron-mediated engineering of *Clostridium*,” in *Methods in molecular biology (Clifton, N.J.)*, 389–407. doi:10.1007/978-1-61779-197-0_23.
- Kurka, H., Ehrenreich, A., Ludwig, W., Monot, M., Rupnik, M., Barbut, F., et al. (2014). Sequence similarity of *Clostridium difficile* strains by analysis of conserved genes and genome content is reflected by their ribotype affiliation. *PLoS One* 9, e86535.
doi:10.1371/journal.pone.0086535.
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948.
doi:10.1093/bioinformatics/btm404.
- Lechner, M., Findeiß, S., Steiner, L., Marz, M., Stadler, P. F., and Prohaska, S. J. (2011). Proteinortho: Detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics* 12, 124.
doi:10.1186/1471-2105-12-124.
- Lee, E. R., Baker, J. L., Weinberg, Z., Sudarsan, N., and Breaker, R. R. (2010). An allosteric self-splicing ribozyme triggered by a bacterial second messenger. *Science* 329, 845–848.
doi:10.1126/science.1190713.
- Lee, H., Dhingra, Y., and Sashital, D. G. (2019). The Cas4-Cas1-Cas2 complex mediates precise prespacer processing during CRISPR adaptation. *Elife* 8, e44248. doi:10.7554/eLife.44248.
- Lee, H., Zhou, Y., Taylor, D. W., and Sashital, D. G. (2018). Cas4-dependent prespacer processing ensures high-fidelity programming of CRISPR arrays. *Mol. Cell* 70, 48-59.e5.
doi:10.1016/j.molcel.2018.03.003.
- Lee, K.-Y., and Lee, B.-J. (2016). Structure, biology, and therapeutic application of toxin–antitoxin systems in pathogenic bacteria. *Toxins (Basel)*. 8, 305. doi:10.3390/toxins8100305.
- Leenay, R. T., Maksimchuk, K. R., Slotkowski, R. A., Agrawal, R. N., Gomaa, A. A., Briner, A.

- E., et al. (2016). Identifying and visualizing functional PAM diversity across CRISPR-Cas systems. *Mol. Cell* 62, 137–147. doi:10.1016/j.molcel.2016.02.031.
- Levy, A., Goren, M. G., Yosef, I., Auster, O., Manor, M., Amitai, G., et al. (2015). CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* 520, 505–510. doi:10.1038/nature14302.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi:10.1093/bioinformatics/btp324.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi:10.1093/bioinformatics/btp352.
- Li, Y., Pan, S., Zhang, Y., Ren, M., Feng, M., Peng, N., et al. (2015). Harnessing Type I and Type III CRISPR-Cas systems for genome editing. *Nucleic Acids Res.* 44, e34. doi:10.1093/nar/gkv1044.
- Lin, Y.-P., Kuo, C.-J., Koleci, X., McDonough, S. P., and Chang, Y.-F. (2011). Manganese binds to *Clostridium difficile* Fbp68 and is essential for fibronectin binding. *J. Biol. Chem.* 286, 3957–3969. doi:10.1074/jbc.M110.184523.
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* 25, 402–408. doi:10.1006/METH.2001.1262.
- Lobato-Márquez, D., Moreno-Córdoba, I., Figueroa, V., Díaz-Orejás, R., and García-del Portillo, F. (2015). Distinct type I and type II toxin-antitoxin modules control *Salmonella* lifestyle inside eukaryotic cells. *Sci. Rep.* 5, 9374. doi:10.1038/srep09374.
- Louwen, R., Staals, R. H. J., Endtz, H. P., van Baarlen, P., and van der Oost, J. (2014). The role of CRISPR-Cas systems in virulence of pathogenic bacteria. *Microbiol. Mol. Biol. Rev.* 78, 74–88. doi:10.1128/MMBR.00039-13.
- Lucado, J., Gould, C., and Elixhauser, A. (2006). *Clostridium Difficile Infections (CDI) in Hospital Stays, 2009: Statistical Brief #124*.
- Luo, M. L., Mullis, A. S., Leenay, R. T., and Beisel, C. L. (2015). Repurposing endogenous type I CRISPR-Cas systems for programmable gene repression. *Nucleic Acids Res.* 43, 674–681. doi:10.1093/nar/gku971.
- Mackin, K. E., Carter, G. P., Howarth, P., Rood, J. I., and Lyras, D. (2013). Spo0A differentially regulates toxin production in evolutionarily diverse strains of *Clostridium difficile*. *PLoS One* 8, e79666. doi:10.1371/journal.pone.0079666.
- Maier, L.-K., Dyall-Smith, M., and Marchfelder, A. (2015). The adaptive immune system of

- Haloferax volcanii*. *Life* 5, 521–537. doi:10.3390/life5010521.
- Maier, L.-K., Stoll, B., Brendel, J., Fischer, S., Pfeiffer, F., Dyall-Smith, M., et al. (2013). The ring of confidence: a haloarchaeal CRISPR/Cas system. *Biochem. Soc. Trans.* 41, 374–8. doi:10.1042/BST20120263.
- Maikova, A., Kreis, V., Boutserin, A., Severinov, K., and Soutourina, O. (2019). Using endogenous CRISPR-Cas system for genome editing in the human pathogen *Clostridium difficile*. *Appl. Environ. Microbiol.*, AEM.01416-19. doi:10.1128/AEM.01416-19.
- Maikova, A., Peltier, J., Boudry, P., Hajnsdorf, E., Kint, N., Monot, M., et al. (2018a). Discovery of new type I toxin–antitoxin systems adjacent to CRISPR arrays in *Clostridium difficile*. *Nucleic Acids Res.* 46, 4733–4751. doi:10.1093/nar/gky124.
- Maikova, A., Severinov, K., and Soutourina, O. (2018b). New insights into functions and possible applications of *Clostridium difficile* CRISPR-Cas system. *Front. Microbiol.* 9, 1740. doi:10.3389/fmicb.2018.01740.
- Maisonneuve, E., Castro-Camargo, M., and Gerdes, K. (2013). (p)ppGpp controls bacterial persistence by stochastic induction of toxin-antitoxin activity. *Cell* 154, 1140–1150. doi:10.1016/J.CELL.2013.07.048.
- Makarova, K. S., Aravind, L., Wolf, Y. I., and Koonin, E. V (2011a). Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biol. Direct* 6, 38. doi:10.1186/1745-6150-6-38.
- Makarova, K. S., Grishin, N. V, Shabalina, S. A., Wolf, Y. I., and Koonin, E. V (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct* 1, 7. doi:10.1186/1745-6150-1-7.
- Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J. J., Charpentier, E., Horvath, P., et al. (2011b). Evolution and classification of the CRISPR–Cas systems. *Nat. Rev. Microbiol.* 9, 467–477. doi:10.1038/nrmicro2577.
- Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., et al. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* 13, 722–736. doi:10.1038/nrmicro3569.
- Makarova, K. S., Wolf, Y. I., and Koonin, E. V. (2013). Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Res.* 41, 4360–4377. doi:10.1093/nar/gkt157.
- Makarova, K. S., Wolf, Y. I., Snir, S., and Koonin, E. V. (2011c). Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *J. Bacteriol.* 193, 6039–6056. doi:10.1128/JB.05535-11.

- Maldarelli, G. A., Piepenbrink, K. H., Scott, A. J., Freiberg, J. A., Song, Y., Achermann, Y., et al. (2016). Type IV pili promote early biofilm formation by *Clostridium difficile*. *Pathog. Dis.* 74, ftw061. doi:10.1093/femspd/ftw061.
- Mandin, P., Repoila, F., Vergassola, M., Geissmann, T., and Cossart, P. (2007). Identification of new noncoding RNAs in *Listeria monocytogenes* and prediction of mRNA targets. *Nucleic Acids Res.* 35, 962–74. doi:10.1093/nar/gkl1096.
- Mani, N., and Dupuy, B. (2001). Regulation of toxin synthesis in *Clostridium difficile* by an alternative RNA polymerase sigma factor. *Proc. Natl. Acad. Sci.* 98, 5844–5849. doi:10.1073/pnas.101126598.
- Marraffini, L. A. (2015). CRISPR-Cas immunity in prokaryotes. *Nature* 526, 55–61. doi:10.1038/nature15386.
- Matamouros, S., England, P., and Dupuy, B. (2007). *Clostridium difficile* toxin expression is inhibited by the novel regulator TcdC. *Mol. Microbiol.* 64, 1274–1288. doi:10.1111/j.1365-2958.2007.05739.x.
- McAllister, K. N., Bouillaut, L., Kahn, J. N., Self, W. T., and Sorg, J. A. (2017). Using CRISPR-Cas9-mediated genome editing to generate *C. difficile* mutants defective in selenoproteins synthesis. *Sci. Rep.* 7, 14672. doi:10.1038/s41598-017-15236-5.
- McKee, R. W., Mangalea, M. R., Purcell, E. B., Borchardt, E. K., and Tamayo, R. (2013). The second messenger cyclic di-GMP regulates *Clostridium difficile* toxin production by controlling expression of *sigD*. *J. Bacteriol.* 195, 5174–5185. doi:10.1128/JB.00501-13.
- Melnikow, E., Schoenfeld, C., Spehr, V., Warrass, R., Gunkel, N., Duszenko, M., et al. (2008). A compendium of antibiotic-induced transcription profiles reveals broad regulation of *Pasteurella multocida* virulence genes. *Vet. Microbiol.* 131, 277–292. doi:10.1016/j.vetmic.2008.03.007.
- Metcalf, D., Sharif, S., and Weese, J. S. (2010). Evaluation of candidate reference genes in *Clostridium difficile* for gene expression normalization. *Anaerobe* 16, 439–443. doi:10.1016/j.anaerobe.2010.06.007.
- Mick, E., Stern, A., and Sorek, R. (2013). Holding a grudge: persisting anti-phage CRISPR immunity in multiple human gut microbiomes. *RNA Biol.* 10, 900–6. doi:10.4161/rna.23929.
- Miller, M. B., and Bassler, B. L. (2001). Quorum sensing in bacteria. *Annu. Rev. Microbiol.* 55, 165–199. doi:10.1146/annurev.micro.55.1.165.
- Miniño, A. M., Xu, J., and Kochanek, K. D. (2010). Deaths: preliminary data for 2008. *Natl. Vital Stat. Rep.* 59, 1–52.
- Minot, S., Bryson, A., Chehoud, C., Wu, G. D., Lewis, J. D., and Bushman, F. D. (2013). Rapid

- evolution of the human gut virome. *Proc. Natl. Acad. Sci. U. S. A.* 110, 12450–5. doi:10.1073/pnas.1300833110.
- Mojica, F. J., Díez-Villaseñor, C., Soria, E., and Juez, G. (2000). Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol. Microbiol.* 36, 244–6.
- Mojica, F. J., Ferrer, C., Juez, G., and Rodríguez-Valera, F. (1995). Long stretches of short tandem repeats are present in the largest replicons of the Archaea *Haloferax mediterranei* and *Haloferax volcanii* and could be involved in replicon partitioning. *Mol. Microbiol.* 17, 85–93.
- Mojica, F. J. M., Díez-Villasenor, C., Garcia-Martinez, J., and Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* 60, 174–182. doi:10.1007/s00239-004-0046-3.
- Monot, M., Boursaux-Eude, C., Thibonnier, M., Vallenet, D., Moszer, I., Medigue, C., et al. (2011). Reannotation of the genome sequence of *Clostridium difficile* strain 630. *J. Med. Microbiol.* 60, 1193–1199. doi:10.1099/jmm.0.030452-0.
- Monot, M., Orgeur, M., Camiade, E., Brehier, C., and Dupuy, B. (2014). COV2HTML: a visualization and analysis tool of bacterial next generation sequencing (NGS) data for postgenomics life scientists. *OMICS* 18, 184–195. doi:10.1089/omi.2013.0119.
- Morgan, M., Anders, S., Lawrence, M., Aboyoun, P., Pagès, H., and Gentleman, R. (2009). ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinforma. Appl. NOTE* 25, 2607–2608. doi:10.1093/bioinformatics/btp450.
- Moura, I., Monot, M., Tani, C., Spigaglia, P., Barbanti, F., Norais, N., et al. (2014). Multidisciplinary analysis of a nontoxigenic *Clostridium difficile* strain with stable resistance to metronidazole. *Antimicrob. Agents Chemother.* 58, 4957–60. doi:10.1128/AAC.02350-14.
- Müh, U., Pannullo, A. G., Weiss, D. S., and Ellermeier, C. D. (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. doi:10.1128/jb.00711-18.
- Mukhopadhyay, A., Redding, A. M., Joachimiak, M. P., Arkin, A. P., Borglin, S. E., Dehal, P. S., et al. (2007). Cell-wide responses to low-oxygen exposure in *Desulfovibrio vulgaris* Hildenborough. *J. Bacteriol.* 189, 5996–6010. doi:10.1128/JB.00368-07.
- Müller, P., Jahn, N., Ring, C., Maiwald, C., Neubert, R., Meißner, C., et al. (2016). A multistress responsive type I toxin-antitoxin system: *bsrE* /SR5 from the *B. subtilis* chromosome. *RNA Biol.* 13, 511–523. doi:10.1080/15476286.2016.1156288.
- Nale, J. Y., Chutia, M., Carr, P., Hickenbotham, P. T., Clokie, M. R. J., and Abedon, S. T. (2016).

- ‘Get in early’; Biofilm and wax moth (*Galleria mellonella*) models reveal new insights into the therapeutic potential of *Clostridium difficile* bacteriophages. *7*, 1–16.
doi:10.3389/fmicb.2016.01383.
- Nam, K. H., Haitjema, C., Liu, X., Ding, F., Wang, H., DeLisa, M. P., et al. (2012). Cas5d protein processes pre-crRNA and assembles into a Cascade-like interference complex in subtype I-C/Dvulg CRISPR-Cas system. *Structure* *20*, 1574–1584. doi:10.1016/j.str.2012.06.016.
- Negahdaripour, M., Nezafat, N., Hajighahramani, N., Rahmatabadi, S. S., and Ghasemi, Y. (2017). Investigating CRISPR-Cas systems in *Clostridium botulinum* via bioinformatics tools. *Infect. Genet. Evol.* *54*, 355–373. doi:10.1016/J.MEEGID.2017.06.027.
- Ng, Y. K., Ehsaan, M., Philip, S., Collery, M. M., Janoir, C., Collignon, A., et al. (2013). Expanding the repertoire of gene tools for precise manipulation of the *Clostridium difficile* genome: allelic exchange using pyrE alleles. *PLoS One* *8*, e56051.
doi:10.1371/journal.pone.0056051.
- Niewoehner, O., Jinek, M., and Doudna, J. A. (2014). Evolution of CRISPR RNA recognition and processing by Cas6 endonucleases. *Nucleic Acids Res.* *42*, 1341–53. doi:10.1093/nar/gkt922.
- Nozawa, T., Furukawa, N., Aikawa, C., Watanabe, T., Haobam, B., Kurokawa, K., et al. (2011). CRISPR inhibition of prophage acquisition in *Streptococcus pyogenes*. *PLoS One* *6*, e19543.
doi:10.1371/journal.pone.0019543.
- Nudler, E., and Mironov, A. S. (2004). The riboswitch control of bacterial metabolism. *Trends Biochem. Sci.* *29*, 11–17. doi:10.1016/j.tibs.2003.11.004.
- Nuñez, J. K., Bai, L., Harrington, L. B., Hinder, T. L., and Doudna, J. A. (2016). CRISPR immunological memory requires a host factor for specificity. *Mol. Cell* *62*, 824–833.
doi:10.1016/J.MOLCEL.2016.04.027.
- Nuñez, J. K., Harrington, L. B., Kranzusch, P. J., Engelman, A. N., and Doudna, J. A. (2015a). Foreign DNA capture during CRISPR–Cas adaptive immunity. *Nature* *527*, 535–538.
doi:10.1038/nature15760.
- Nuñez, J. K., Kranzusch, P. J., Noeske, J., Wright, A. V, Davies, C. W., and Doudna, J. A. (2014). Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. *Nat. Struct. Mol. Biol.* *21*, 528–534. doi:10.1038/nsmb.2820.
- Nuñez, J. K., Lee, A. S. Y., Engelman, A., and Doudna, J. A. (2015b). Integrase-mediated spacer acquisition during CRISPR–Cas adaptive immunity. *Nature* *519*, 193–198.
doi:10.1038/nature14237.
- O’Connor, J. R., Lyras, D., Farrow, K. A., Adams, V., Powell, D. R., Hinds, J., et al. (2006). Construction and analysis of chromosomal *Clostridium difficile* mutants. *Mol. Microbiol.* *61*,

- 1335–1351. doi:10.1111/j.1365-2958.2006.05315.x.
- Oren, A., and Rupnik, M. (2018). *Clostridium difficile* and *Clostridioides difficile*: two validly published and correct names. *Anaerobe* 52, 125–126. doi:10.1016/J.ANAEROBE.2018.07.005.
- Owens, Jr., R. C., Donskey, C. J., Gaynes, R. P., Loo, V. G., and Muto, C. A. (2008). Antimicrobial-associated risk factors for *Clostridium difficile* infection. *Clin. Infect. Dis.* 46, S19–S31. doi:10.1086/521859.
- Page, R., and Peti, W. (2016). Toxin-antitoxin systems in bacterial growth arrest and persistence. *Nat. Chem. Biol.* 12, 208–214. doi:10.1038/nchembio.2044.
- Palmer, K. L., and Gilmore, M. S. (2010). Multidrug-resistant enterococci lack CRISPR-cas. *MBio* 1, e00227-10. doi:10.1128/mBio.00227-10.
- Patel, S., and Weaver, K. E. (2006). Addiction toxin Fst has unique effects on chromosome segregation and cell division in *Enterococcus faecalis* and *Bacillus subtilis*. *J. Bacteriol.* 188, 5374–5384. doi:10.1128/JB.00513-06.
- Patterson, A. G., Jackson, S. A., Taylor, C., Evans, G. B., Salmond, G. P. C., Przybilski, R., et al. (2016). Quorum sensing controls adaptive immunity through the regulation of multiple CRISPR-Cas Systems. *Mol. Cell* 64, 1102–1108. doi:10.1016/J.MOLCEL.2016.11.012.
- Pawluk, A., Davidson, A. R., and Maxwell, K. L. (2018). Anti-CRISPR: discovery, mechanism and function. *Nat. Rev. Microbiol.* 16, 12–17. doi:10.1038/nrmicro.2017.120.
- Pearson, B. M., Louwen, R., van Baarlen, P., and van Vliet, A. H. M. (2015). Differential distribution of Type II CRISPR-Cas systems in agricultural and nonagricultural *Campylobacter coli* and *Campylobacter jejuni* isolates correlates with lack of shared environments. *Genome Biol. Evol.* 7, 2663–2679. doi:10.1093/gbe/evv174.
- Peltier, J., Shaw, H. A., Couchman, E. C., Dawson, L. F., Yu, L., Choudhary, J. S., et al. (2015). Cyclic diGMP regulates production of sortase substrates of *Clostridium difficile* and their surface exposure through ZmpI protease-mediated cleavage. *J. Biol. Chem.* 290, 24453–24469. doi:10.1074/jbc.M115.665091.
- Peng, L., Pei, J., Pang, H., Guo, Y., Lin, L., and Huang, R. (2014). Whole genome sequencing reveals a novel CRISPR system in industrial *Clostridium acetobutylicum*. *J. Ind. Microbiol. Biotechnol.* 41, 1677–1685. doi:10.1007/s10295-014-1507-3.
- Pepin, J., Valiquette, L., Alary, M.-E., Villemure, P., Pelletier, A., Forget, K., et al. (2004). *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. doi:10.1503/cmaj.1041104.
- Pettengill, J. B., Timme, R. E., Barrangou, R., Toro, M., Allard, M. W., Strain, E., et al. (2014).

- The evolutionary history and diagnostic utility of the CRISPR-Cas system within *Salmonella enterica* ssp. *enterica*. *PeerJ* 2, e340. doi:10.7717/peerj.340.
- Pettit, L. J., Browne, H. P., Yu, L., Smits, W., Fagan, R. P., Barquist, L., et al. (2014). Functional genomics reveals that *Clostridium difficile* Spo0A coordinates sporulation, virulence and metabolism. *BMC Genomics* 15, 160. doi:10.1186/1471-2164-15-160.
- Pichon, C., and Felden, B. (2007). Proteins that interact with bacterial small RNA regulators. *FEMS Microbiol. Rev.* 31, 614–625. doi:10.1111/j.1574-6976.2007.00079.x.
- Pinel-Marie, M.-L., Brielle, R., and Felden, B. (2014). Dual toxic-peptide-coding *Staphylococcus aureus* RNA under antisense regulation targets host cells and bacterial rivals unequally. *Cell Rep.* 7, 424–435. doi:10.1016/j.celrep.2014.03.012.
- Poehlein, A., Zverlov, V. V., Daniel, R., Schwarz, W. H., and Liebl, W. (2013). Complete genome sequence of *Clostridium stercorarium* subsp. *stercorarium* strain DSM 8532, a thermophilic degrader of plant cell wall fibers. *Genome Announc.* 1, e0007313. doi:10.1128/genomeA.00073-13.
- Pourcel, C., Salvignol, G., and Vergnaud, G. (2005). CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* 151, 653–663. doi:10.1099/mic.0.27437-0.
- Pruitt, R. N., and Lacy, D. B. (2012). Toward a structural understanding of *Clostridium difficile* toxins A and B. *Front. Cell. Infect. Microbiol.* 2, 28. doi:10.3389/fcimb.2012.00028.
- Purcell, E. B., McKe/e, R. W., McBride, S. M., Waters, C. M., and Tamayo, R. (2012). Cyclic diguanylate inversely regulates motility and aggregation in *Clostridium difficile*. *J. Bacteriol.* 194, 3307–3316. doi:10.1128/JB.00100-12.
- Pyne, M. E., Bruder, M. R., Moo-Young, M., Chung, D. A., and Chou, C. P. (2016). Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in *Clostridium*. *Sci. Rep.* 6, 25666. doi:10.1038/srep25666.
- Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., et al. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152, 1173–1183. doi:10.1016/J.CELL.2013.02.022.
- Ratner, H. K., Sampson, T. R., and Weiss, D. S. (2015). I can see CRISPR now, even when phage are gone. *Curr. Opin. Infect. Dis.* 28, 267–274. doi:10.1097/QCO.0000000000000154.
- Richter, C., Chang, J. T., and Fineran, P. C. (2012). Function and regulation of clustered regularly interspaced short palindromic repeats (CRISPR) / CRISPR associated (Cas) systems. *Viruses* 4, 2291–311. doi:10.3390/v4102291.
- Rodriguez, A. M., Callahan, J. E., Fawcett, P., Ge, X., Xu, P., and Kitten, T. (2011). Physiological

- and molecular characterization of genetic competence in *Streptococcus sanguinis*. *Mol. Oral Microbiol.* 26, 99–116. doi:10.1111/j.2041-1014.2011.00606.x.
- Romby, P., and Charpentier, E. (2010). An overview of RNAs with regulatory functions in gram-positive bacteria. *Cell. Mol. Life Sci.* 67, 217–237. doi:10.1007/s00018-009-0162-8.
- Römling, U. (2012). Cyclic di-GMP, an established secondary messenger still speeding up. *Environ. Microbiol.* 14, 1817–1829. doi:10.1111/j.1462-2920.2011.02617.x.
- Römling, U., Galperin, M. Y., and Gomelsky, M. (2013). Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. *Microbiol. Mol. Biol. Rev.* 77, 1–52. doi:10.1128/MMBR.00043-12.
- Rothenbacher, F. P., Suzuki, M., Hurley, J. M., Montville, T. J., Kirn, T. J., Ouyang, M., et al. (2012). *Clostridium difficile* MazF toxin exhibits selective, not global, mRNA cleavage. *J. Bacteriol.* 194, 3464–74. doi:10.1128/JB.00217-12.
- Rupnik, M., Wilcox, M. H., and Gerding, D. N. (2009). *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. *Nat. Rev. Microbiol.* 7, 526–536. doi:10.1038/nrmicro2164.
- Rutkauskas, M., Sinkunas, T., Songailiene, I., Tikhomirova, M. S., Siksnys, V., and Seidel, R. (2015). Directional R-Loop formation by the CRISPR-Cas surveillance complex Cascade provides efficient off-target site rejection. *Cell Rep.* 10, 1534–1543. doi:10.1016/j.celrep.2015.01.067.
- Ryjenkov, D. A., Tarutina, M., Moskvina, O. V., and Gomelsky, M. (2005). Cyclic diguanylate is a ubiquitous signaling molecule in bacteria: insights into biochemistry of the GGDEF protein domain. *J. Bacteriol.* 187, 1792–1798. doi:10.1128/JB.187.5.1792-1798.2005.
- Safari, F., Zare, K., Negahdaripour, M., Barekati-Mowahed, M., and Ghasemi, Y. (2019). CRISPR Cpf1 proteins: structure, function and implications for genome editing. *Cell Biosci.* 9, 36. doi:10.1186/s13578-019-0298-7.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). *Molecular cloning : a laboratory manual, second edition*. Cold Spring Harbor, N. Y: Cold Spring Harbor Laboratory.
- Sampson, T. R., and Weiss, D. S. (2013). Alternative roles for CRISPR/Cas systems in bacterial pathogenesis. *PLoS Pathog.* 9, 9–11. doi:10.1371/journal.ppat.1003621.
- Saujet, L., Monot, M., Dupuy, B., Soutourina, O., and Martin-Verstraete, I. (2011). The key sigma factor of transition phase, σ^{HNS} , controls sporulation, metabolism, and virulence factor expression in *Clostridium difficile*. *J. Bacteriol.* 193, 3186–3196. doi:10.1128/JB.00272-11.
- Sayed, N., Jousselin, A., and Felden, B. (2012a). A *cis*-antisense RNA acts in trans in *Staphylococcus aureus* to control translation of a human cytolytic peptide. *Nat. Struct. Mol.*

- Biol.* 19, 105–112. doi:10.1038/nsmb.2193.
- Sayed, N., Nonin-Lecomte, S., Réty, S., and Felden, B. (2012b). Functional and structural insights of a *Staphylococcus aureus* apoptotic-like membrane peptide from a toxin-antitoxin module. *J. Biol. Chem.* 287, 43454–43463. doi:10.1074/jbc.M112.402693.
- Schmidt, A. J., Ryjenkov, D. A., and Gomelsky, M. (2005). The ubiquitous protein domain EAL is a cyclic diguanylate-specific phosphodiesterase: enzymatically active and inactive EAL domains. *J. Bacteriol.* 187, 4774–4781. doi:10.1128/JB.187.14.4774-4781.2005.
- Sebahia, M., Wren, B. W., Mullany, P., Fairweather, N. F., Minton, N., Stabler, R., et al. (2006). The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nat. Genet.* 38, 779–786. doi:10.1038/ng1830.
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi:10.1093/bioinformatics/btu153.
- Sekulovic, O., Garneau, J. R., Néron, A., and Fortier, L. C. (2014). Characterization of temperate phages infecting *Clostridium difficile* isolates of human and animal origins. *Appl. Environ. Microbiol.* 80, 2555–2563. doi:10.1128/AEM.00237-14.
- Semenova, E., Jore, M. M., Datsenko, K. A., Semenova, A., Westra, E. R., Wanner, B., et al. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci.* 108, 10098–10103. doi:10.1073/pnas.1104144108.
- Semenyuk, E. G., Laning, M. L., Foley, J., Johnston, P. F., Knight, K. L., Gerding, D. N., et al. (2014). Spore formation and toxin production in *Clostridium difficile* biofilms. *PLoS One* 9, e87757. doi:10.1371/journal.pone.0087757.
- Serrano, M., Kint, N., Pereira, F. C., Saujet, L., Boudry, P., Dupuy, B., et al. (2016). A recombination directionality factor controls the cell type-specific activation of σ K and the fidelity of spore development in *Clostridium difficile*. *PLOS Genet.* 12, e1006312. doi:10.1371/journal.pgen.1006312.
- Sesto, N., Touchon, M., Andrade, J. M., Kondo, J., Rocha, E. P. C., Arraiano, C. M., et al. (2014). A PNPase dependent CRISPR system in *Listeria*. *PLoS Genet.* 10, e1004065. doi:10.1371/journal.pgen.1004065.
- Shah, S. A., Erdmann, S., Mojica, F. J. M., and Garrett, R. A. (2013). Protospacer recognition motifs. *RNA Biol.* 10, 891–899. doi:10.4161/rna.23764.
- Shen, Z., Patil, R. D., Sahin, O., Wu, Z., Pu, X.-Y., Dai, L., et al. (2016). Identification and functional analysis of two toxin-antitoxin systems in *Campylobacter jejuni*. *Mol. Microbiol.* 101, 909–923. doi:10.1111/mmi.13431.

- Shields, K., Araujo-Castillo, R. V., Theethira, T. G., Alonso, C. D., and Kelly, C. P. (2015). Recurrent *Clostridium difficile* infection: From colonization to cure. *Anaerobe* 34, 59–73. doi:10.1016/j.anaerobe.2015.04.012.
- Shimori, M., Garrett, S. C., Chambers, D. P., Glover, C. V. C., Graveley, B. R., and Terns, M. P. (2017). Role of free DNA ends and protospacer adjacent motifs for CRISPR DNA uptake in *Pyrococcus furiosus*. *Nucleic Acids Res.*, 1–14. doi:10.1093/nar/gkx839.
- Shmakov, S. A., Sitnik, V., Makarova, K. S., Wolf, Y. I., Severinov, K. V., and Koonin, E. V. (2017). The CRISPR spacer space is dominated by sequences from species-specific mobilomes. *MBio* 8, e01397-17. doi:10.1128/mBio.01397-17.
- Shmakov, S., Savitskaya, E., Semenova, E., Logacheva, M. D., Datsenko, K. A., and Severinov, K. (2014). Pervasive generation of oppositely oriented spacers during CRISPR adaptation. *Nucleic Acids Res.* 42, 5907–5916. doi:10.1093/nar/gku226.
- Silas, S., Lucas-Elio, P., Jackson, S. A., Aroca-Crevillén, A., Hansen, L. L., Fineran, P. C., et al. (2017). Type III CRISPR-Cas systems can provide redundancy to counteract viral escape from type I systems. *Elife* 6, e27601. doi:10.7554/eLife.27601.
- Silvaggi, J. M., Perkins, J. B., and Losick, R. (2005). Small untranslated RNA antitoxin in *Bacillus subtilis*. *J. Bacteriol.* 187, 6641–6650. doi:10.1128/JB.187.19.6641-6650.2005.
- Smits, W. K., Lyras, D., Lacy, D. B., Wilcox, M. H., and Kuijper, E. J. (2016). *Clostridium difficile* infection. *Nat. Rev. Dis. Prim.* 2, 16020. doi:10.1038/nrdp.2016.20.
- Soavelomandroso, A. P., Gaudin, F., Hoys, S., Nicolas, V., Vedantam, G., Janoir, C., et al. (2017). Biofilm structures in a mono-associated mouse model of *Clostridium difficile* Infection. *Front. Microbiol.* 8, 2086. doi:10.3389/fmicb.2017.02086.
- Sobrero, P., and Valverde, C. (2012). The bacterial protein Hfq: much more than a mere RNA-binding factor. *Crit. Rev. Microbiol.* 38, 276–299. doi:10.3109/1040841X.2012.664540.
- Sorek, R., Lawrence, C. M., and Wiedenheft, B. (2013). CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annu. Rev. Biochem.* 82, 237–66. doi:10.1146/annurev-biochem-072911-172315.
- Soutourina, O. (2017). RNA-based control mechanisms of *Clostridium difficile*. *Curr. Opin. Microbiol.* 36, 62–68. doi:10.1016/j.mib.2017.01.004.
- Soutourina, O., Monot, M., Boudry, P., Saujet, L., Pichon, C., Sismeiro, O., et al. (2013). Genome-wide identification of regulatory RNAs in the human pathogen *Clostridium difficile*. *PLoS Genet.* 9, e1003493. doi:10.1371/journal.pgen.1003493.
- Stabler, R. A., He, M., Dawson, L., Martin, M., Valiente, E., Corton, C., et al. (2009). Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the

- evolution of a hypervirulent bacterium. *Genome Biol.* 10, R102. doi:10.1186/gb-2009-10-9-r102.
- Stachler, A.-E., and Marchfelder, A. (2016). Gene repression in *Haloarchaea* using the CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas I-B system. *J. Biol. Chem.* 291, 15226–42. doi:10.1074/jbc.M116.724062.
- Steiner, C., Barrett, M., and Weiss, A. (2012). *Clostridium difficile* hospitalizations 2001 to 2013. HCUP Projections Report# 2014-01. Rockville Available at: <http://hcuptus.ahrq.gov/reports/projections/2014T03.pdf>.
- Stern, A., Mick, E., Tirosh, I., Sagy, O., and Sorek, R. (2012). CRISPR targeting reveals a reservoir of common phages associated with the human gut microbiome. *Genome Res.* 22, 1985–94. doi:10.1101/gr.138297.112.
- Sudarsan, N., Lee, E. R., Weinberg, Z., Moy, R. H., Kim, J. N., Link, K. H., et al. (2008). Riboswitches in eubacteria sense the second messenger cyclic di-GMP. *Science* 321, 411–3. doi:10.1126/science.1159519.
- Takeuchi, N., Wolf, Y. I., Makarova, K. S., and Koonin, E. V (2012). Nature and intensity of selection pressure on CRISPR-associated genes. *J. Bacteriol.* 194, 1216–25. doi:10.1128/JB.06521-11.
- Tamayo, R., Tischler, A. D., and Camilli, A. (2005). The EAL domain protein VieA is a cyclic diguanylate phosphodiesterase. *J. Biol. Chem.* 280, 33324–33330. doi:10.1074/jbc.M506500200.
- Theophilou, E. S., Vohra, P., Gallagher, M. P., Poxton, I. R., and Blakely, G. W. (2019). Generation of markerless deletions in the nosocomial pathogen *Clostridium difficile* by induction of DNA double-strand breaks. *Appl. Environ. Microbiol.* 85, e02055-18. doi:10.1128/AEM.02055-18.
- Timme, R. E., Pettengill, J. B., Allard, M. W., Strain, E., Barrangou, R., Wehnes, C., et al. (2013). Phylogenetic diversity of the enteric pathogen *Salmonella enterica* subsp. *enterica* inferred from genome-wide reference-free SNP characters. *Genome Biol. Evol.* 5, 2109–23. doi:10.1093/gbe/evt159.
- Tulli, L., Marchi, S., Petracca, R., Shaw, H. A., Fairweather, N. F., Scarselli, M., et al. (2013). CbpA: a novel surface exposed adhesin of *Clostridium difficile* targeting human collagen. *Cell. Microbiol.* 15, n/a-n/a. doi:10.1111/cmi.12139.
- Vedantam, G., Clark, A., Chu, M., McQuade, R., Mallozzi, M., and Viswanathan, V. K. (2012). *Clostridium difficile* infection: Toxins and non-toxin virulence factors, and their contributions to disease establishment and host response. *Gut Microbes* 3, 121–134.

- doi:10.4161/gmic.19399.
- Vengadesan, K., and Narayana, S. V. L. (2011). Structural biology of gram-positive bacterial adhesins. *Protein Sci.* 20, 759–772. doi:10.1002/pro.613.
- Vercoe, R. B., Chang, J. T., Dy, R. L., Taylor, C., Gristwood, T., Clulow, J. S., et al. (2013). Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape bacterial genomes and expel or remodel pathogenicity islands. *PLoS Genet.* 9, e1003454. doi:10.1371/journal.pgen.1003454.
- Vogel, J., and Luisi, B. F. (2011). Hfq and its constellation of RNA. *Nat. Rev. Microbiol.* 9, 578–589. doi:10.1038/nrmicro2615.
- Vorontsova, D., Datsenko, K. A., Medvedeva, S., Bondy-Denomy, J., Savitskaya, E. E., Pougach, K., et al. (2015). Foreign DNA acquisition by the I-F CRISPR-Cas system requires all components of the interference machinery. *Nucleic Acids Res.* 43, 10848–60. doi:10.1093/nar/gkv1261.
- Wagner, E. G. H., and Romby, P. (2015). Small RNAs in bacteria and archaea: who they are, what they do, and how they do it. *Adv. Genet.* 90, 133–208. doi:10.1016/BS.ADGEN.2015.05.001.
- Waligora, A.-J., Hennequin, C., Mullany, P., Bourlioux, P., Collignon, A., and Karjalainen, T. (2001). Characterization of a cell surface protein of *Clostridium difficile* with adhesive properties. *Infect. Immun.* 69, 2144. doi:10.1128/IAI.69.4.2144-2153.2001.
- Wang, H., La Russa, M., and Qi, L. S. (2016). CRISPR/Cas9 in genome editing and beyond. *Annu. Rev. Biochem.* 85, 227–264. doi:10.1146/annurev-biochem-060815-014607.
- Wang, S., Hong, W., Dong, S., Zhang, Z.-T., Zhang, J., Wang, L., et al. (2018). Genome engineering of *Clostridium difficile* using the CRISPR-Cas9 system. *Clin. Microbiol. Infect.* 0. doi:10.1016/j.cmi.2018.03.026.
- Wang, X., and Wood, T. K. (2011). Toxin-antitoxin systems influence biofilm and persister cell formation and the general stress response. *Appl. Environ. Microbiol.* 77, 5577–83. doi:10.1128/AEM.05068-11.
- Wassarman, K. M. (2007). 6S RNA: a small RNA regulator of transcription. *Curr. Opin. Microbiol.* 10, 164–168. doi:10.1016/j.mib.2007.03.008.
- Waters, L. S., and Storz, G. (2009). Regulatory RNAs in bacteria. *Cell* 136, 615–28. doi:10.1016/j.cell.2009.01.043.
- Wen, Y., Behiels, E., and Devreese, B. (2014). Toxin-Antitoxin systems: their role in persistence, biofilm formation, and pathogenicity. doi:10.1111/2049-632X.12145.
- Westra, E. R., Buckling, A., and Fineran, P. C. (2014). CRISPR-Cas systems: beyond adaptive immunity. *Nat. Rev. Microbiol.* 12, 317–326. doi:10.1038/nrmicro3241.

- Westra, E. R., van Erp, P. B. G., Künne, T., Wong, S. P., Staals, R. H. J., Seegers, C. L. C., et al. (2012). CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol. Cell* 46, 595–605. doi:10.1016/j.molcel.2012.03.018.
- Wickham, H. (2009). *Ggplot2 : elegant graphics for data analysis*. Springer.
- Wiedenheft, B., van Duijn, E., Bultema, J. B., Waghmare, S. P., Zhou, K., Barendregt, A., et al. (2011). RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc. Natl. Acad. Sci.* 108, 10092–10097. doi:10.1073/pnas.1102716108.
- Wozniak, R. A. F., Fouts, D. E., Spagnoletti, M., Colombo, M. M., Ceccarelli, D., Garriss, G., et al. (2009). Comparative ICE genomics: Insights into the evolution of the SXT/R391 family of ICEs. *PLoS Genet.* 5, e1000786. doi:10.1371/journal.pgen.1000786.
- Wright, A., Wait, R., Begum, S., Crossett, B., Nagy, J., Brown, K., et al. (2005). Proteomic analysis of cell surface proteins from *Clostridium difficile*. *Proteomics* 5, 2443–2452. doi:10.1002/pmic.200401179.
- Yamaguchi, Y., and Inouye, M. (2011). Regulation of growth and death in *Escherichia coli* by toxin–antitoxin systems. *Nat. Rev. Microbiol.* 9, 779–790. doi:10.1038/nrmicro2651.
- Yosef, I., Goren, M. G., and Qimron, U. (2012). Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res.* 40, 5569–5576. doi:10.1093/nar/gks216.
- Zegans, M. E., Wagner, J. C., Cady, K. C., Murphy, D. M., Hammond, J. H., and O’Toole, G. A. (2009). Interaction between bacteriophage DMS3 and host CRISPR region inhibits group behaviors of *Pseudomonas aeruginosa*. *J. Bacteriol.* 191, 210–9. doi:10.1128/JB.00797-08.
- Zhang, J., Kasciukovic, T., and White, M. F. (2012). The CRISPR associated protein Cas4 is a 5’ to 3’ DNA exonuclease with an iron-sulfur cluster. *PLoS One* 7. doi:10.1371/journal.pone.0047232.
- Zhang, J., Zong, W., Hong, W., Zhang, Z.-T., and 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. doi:10.1016/j.ymben.2018.03.007.
- Zhang, Z., Pan, S., Liu, T., Li, Y., and Peng, N. (2019). Cas4 nucleases can effect specific integration of CRISPR spacers. *J. Bacteriol.* 201, e00747-18. doi:10.1128/JB.

Résumé

Introduction

Clostridium difficile est un entéropathogène opportuniste et la principale cause de diarrhée nosocomiale chez l'adulte. Au cours de son cycle d'infection, la bactérie survit dans les communautés complexes du côlon, éventuellement en utilisant les mécanismes de défense des systèmes immunitaires bactériens. Les systèmes CRISPR-Cas sont des systèmes de défense adaptatifs procaryotes contre les phages et d'autres agents génétiques étrangers. Le *C. difficile* possède un système CRISPR-Cas de sous-type I-B avec plusieurs caractéristiques inhabituelles: un grand nombre de cassettes CRISPR activement exprimées, dont certaines sont localisées à l'intérieur de régions de prophage, et de multiples opérons *cas* (Boudry et al., 2015). Ce système CRISPR-Cas original peut jouer un rôle crucial dans l'adaptation de *C. difficile* à l'intérieur de l'hôte.

Buts de la recherche

1. Etudier le rôle et la fonctionnalité du système CRISPR-Cas de *C. difficile* dans les interactions avec des éléments d'ADN étrangers (tels que les plasmides) ;
2. Révéler la manière dont le système CRISPR-Cas de *C. difficile* est régulé et fonctionne dans des conditions de culture bactérienne différentes, incluant la réponse aux stress.

Résultats

Dans ce travail, nous avons étudié tous les aspects fonctionnels généraux du système CRISPR-Cas de *C. difficile* et les principales conclusions de cette partie de la recherche sont les suivantes:

1. Des séquences consensus de motifs PAM élargies (CCN / TCN) ont été identifiées pour les souches de *C. difficile* 630 Δ_{erm} et R20291 ;
2. L'interférence active et la contribution différente à la défense des 12 cassettes CRISPR dans la souche du laboratoire 630 Δ_{erm} ont été démontrées ;
3. Le processus d'interférence actif et des motifs PAM élargies (CCC / CCG, CCA / CCT) ont été confirmés expérimentalement pour la souche epidémique R20291;
4. La délétion de l'opéron *cas* complet n'a pas complètement supprimé les capacités d'interférence dans la souche 630 Δ_{erm} Δ_{erm} ;

5. Une nouvelle acquisition de séquence « *spacer* » a été démontrée pour 2 cassettes CRISPR de la souche 630 Δ *erm*, et l'adaptation naïve ne semble pas être aussi active que l'interférence chez *C. difficile*.

Malgré cela, de nombreuses caractéristiques du système CRISPR-Cas de *C. difficile* restent à étudier. En particulier, la fonctionnalité des cassettes CRISPR restants de *C. difficile* R20291 et le rôle de tous les opérons *cas* des souches 630 Δ *erm* et R20291 dans les processus d'interférence et les infections à *C. difficile* n'ont pas encore été évalués. De plus, la fonction générale du système CRISPR-Cas de *C. difficile* au cours de l'infection à *C. difficile* doit être explorée par des expériences supplémentaires. En particulier, l'inactivation complète du système par la délétion de tous les opérons *cas* permettra d'étudier la « *fitness* » de la souche mutante à l'intérieur de l'hôte à l'aide de modèles animaux disponibles. Nos résultats ont également révélé que le type naïf d'adaptation CRISPR. D'autres expériences avec des tests sur des phages et des essais d'adaptation enrichiront nos connaissances sur les mécanismes d'adaptation du système CRISPR chez *C. difficile*. Ce travail a également soulevé une question sur la fonction de la protéine Cas4 dans le processus d'acquisition des nouvelles séquences « *spacer* » et la physiologie globale de *C. difficile*. Enfin, on ne comprend pas pourquoi le système CRISPR-Cas de *C. difficile* n'est pas très actif en matière d'adaptation. On pourrait émettre l'hypothèse que des protéines anti-CRISPR non caractérisées (Pawluk et al., 2018), potentiellement codées à l'intérieur des régions du prophage chromosomique, pourraient inhiber le processus d'adaptation. Ces points intéressants devraient être explorés dans les études futures.

Un autre objectif de cette thèse de doctorat visait à étudier les mécanismes de régulation du système CRISPR-Cas de *C. difficile*, en particulier dans des conditions de biofilm et en réponse à divers stress. Au cours de ce travail, nous avons découvert une caractéristique unique de ce système CRISPR-Cas chez une bactérie entéropathogène - une association de systèmes toxine-antitoxine (TA) de type I avec des cassettes CRISPR dans la majorité des souches de *C. difficile* séquencées (Maikova et al., 2018). Nous avons choisi deux de ces modules TA et avons étudié leur fonctionnalité et un lien possible avec la régulation du système CRISPR-Cas. Les principaux résultats de cette étude sont les suivants :

1. La fonctionnalité des systèmes TA de type I a été démontrée avec l'arrêt de la croissance induit par la surexpression de la toxine et la neutralisation des toxines par les antitoxines ;
2. La co-régulation des cassettes CRISPR et des systèmes TA adjacents de type I a été suggérée dans les biofilms et dans des conditions de stress, potentiellement associées à la présence de promoteurs dépendant du facteur sigma de stress général, SigB.

Cependant, le lien direct entre les modules TA de type I et la fonction du système CRISPR-Cas n'a pas été démontré. Cet aspect et d'autres questions sans réponse sur le rôle des modules TA dans la réponse au stress, la stabilité du prophage et la stabilisation des régions chromosomiques portant des cassettes CRISPR doivent être explorés lors d'études ultérieures.

Nous avons également étudié le rôle d'un messager secondaire bactérien c-di-GMP dans la régulation CRISPR-Cas de *C. difficile*. Nous avons trouvé un riboswitch dépendant de c-di-GMP associé à la cassette CRISPR12 dans la souche 630 Δ erm, ce qui peut indiquer l'impact direct de la régulation dépendant de c-di-GMP sur l'expression de cassette CRISPR. En général, l'effet global de c-di-GMP sur l'expression d'autres composants du système CRISPR-Cas a été étudié. Ces expériences ont montré :

1. une légère induction d'efficacité d'interférence dans la souche 630 Δ erm de *C. difficile* pour la cassette CRISPR12 par des niveaux élevés de c-di-GMP ;
2. induction de l'expression des deux opérons *cas* et de plusieurs cassettes CRISPR chez la souche 630 Δ erm de *C. difficile* par des taux élevés de c-di-GMP.

Une analyse plus détaillée de la régulation du système CRISPR-Cas de *C. difficile* doit être effectuée à l'avenir, en particulier de la régulation par d'autres stimuli et stress liés au biofilm et des mécanismes moléculaires de ces processus de régulation.

L'étude de la fonctionnalité du système CRISPR-Cas pour *C. difficile* nous a permis d'explorer son potentiel biotechnologique. Dans cette thèse, nous avons décrit l'application du système CRISPR-Cas natif de *C. difficile* comme nouvel outil de la rédaction du génome de cette bactérie. Ainsi, le chapitre 4 décrit l'utilisation du système CRISPR-Cas endogène de *C. difficile* comme nouvelle technique de la rédaction du génome chez *C. difficile* (Maikova et al., 2019). Les conclusions de cette partie de la recherche sont les suivantes:

1. La même mini-cassette CRISPR peut être utilisée dans les souches 630 Δ erm et R20291 et pourrait probablement être étendu à d'autres souches de *C. difficile* ;
2. Une perte du plasmide d'édition efficace a été démontrée ;
3. Une auto-immunité CRISPR efficace a été observée ;
4. Des mutants Δ hfq ont été créés dans les souches 630 Δ erm et R20291, qui ne pouvaient être obtenus auparavant à l'aide d'autres méthodes de manipulation du génome.

Cette nouvelle approche de la rédaction du génome élargit l'ensemble des outils génétiques disponibles pour *C. difficile*, et cette méthode peut être utilisée pour développer de nouveaux agents antimicrobiens contre l'infection à *C. difficile*.

Cette thèse de doctorat élargit nos connaissances sur la physiologie de *C. difficile* et ses caractéristiques génétiques, et ouvre de nouvelles perspectives pour les applications biotechnologiques du système CRISPR-Cas de *C. difficile*.

Références

- Boudry, P., Semenova, E., Monot, M., Datsenko, K. A., Lopatina, A., Sekulovic, O., et al. (2015). Function of the CRISPR-cas system of the human pathogen *Clostridium difficile*. *MBio* 6, e01112-15. doi:10.1128/mBio.01112-15.
- Pawluk, A., Davidson, A. R., and Maxwell, K. L. (2018). Anti-CRISPR: discovery, mechanism and function. *Nat. Rev. Microbiol.* 16, 12–17. doi:10.1038/nrmicro.2017.120.
- Maikova, A., Peltier, J., Boudry, P., Hajnsdorf, E., Kint, N., Monot, M., et al. (2018). Discovery of new type I toxin–antitoxin systems adjacent to CRISPR arrays in *Clostridium difficile*. *Nucleic Acids Res.* 46, 4733–4751. doi:10.1093/nar/gky124.
- Maikova, A., Kreis, V., Boutserin, A., Severinov, K., and Soutourina, O. (2019). Using endogenous CRISPR-Cas system for genome editing in the human pathogen *Clostridium difficile*. *Appl. Environ. Microbiol.*, AEM.01416-19. doi:10.1128/AEM.01416-19.

Annex. Article 1

Discovery of new type I toxin–antitoxin systems adjacent to CRISPR arrays in *Clostridium difficile*

Anna Maikova^{1,2,3,4,†}, Johann Peltier^{1,2,†}, Pierre Boudry^{1,2}, Eliane Hajnsdorf⁵, Nicolas Kint^{1,2}, Marc Monot^{1,2,6}, Isabelle Poquet^{1,7}, Isabelle Martin-Verstraete^{1,2}, Bruno Dupuy^{1,2} and Olga Soutourina^{1,2,8,*}

¹Laboratoire Pathogénèse des Bactéries Anaérobies, Institut Pasteur, 75724 Paris Cedex 15, France, ²Université Paris Diderot, Sorbonne Paris Cité, 75724 Paris Cedex 15, France, ³Center for Data-Intensive Biomedicine and Biotechnology, Skolkovo Institute of Science and Technology, Moscow 143028, Russia, ⁴Peter the Great St.Petersburg Polytechnic University, Saint Petersburg 195251, Russia, ⁵UMR8261 (CNRS-Univ. Paris Diderot, Sorbonne Paris Cité), Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie, 75005 Paris, France, ⁶Département de Microbiologie et d'infectiologie, Faculté de Médecine et des Sciences de la Santé, Université de Sherbrooke, J1E 4K8, Sherbrooke, QC, Canada, ⁷INRA, UMR1319 Micalis (Microbiologie de l'Alimentation au service de la Santé), Domaine de Vilvert, 78352, Jouy-en-Josas Cedex, France and ⁸Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ. Paris-Sud, Université Paris-Saclay, 91198, Gif-sur-Yvette cedex, France

Received July 21, 2017; Revised February 08, 2018; Editorial Decision February 09, 2018; Accepted February 12, 2018

ABSTRACT

Clostridium difficile, a major human enteropathogen, must cope with foreign DNA invaders and multiple stress factors inside the host. We have recently provided an experimental evidence of defensive function of the *C. difficile* CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR-associated) system important for its survival within phage-rich gut communities. Here, we describe the identification of type I toxin–antitoxin (TA) systems with the first functional antisense RNAs in this pathogen. Through the analysis of deep-sequencing data, we demonstrate the general colocalization with CRISPR arrays for the majority of sequenced *C. difficile* strains. We provide a detailed characterization of the overlapping convergent transcripts for three selected TA pairs. The toxic nature of small membrane proteins is demonstrated by the growth arrest induced by their overexpression. The co-expression of antisense RNA acting as an antitoxin prevented this growth defect. Co-regulation of CRISPR-Cas and type I TA genes by the general stress response Sigma B and biofilm-related factors further suggests a possible link between these systems with a role in recurrent *C. difficile* infections. Our results provide the first description of genomic links between CRISPR and type I TA systems within

defense islands in line with recently emerged concept of functional coupling of immunity and cell dormancy systems in prokaryotes.

INTRODUCTION

All living organisms need to survive in changing environments by adapting their physiology. Horizontal gene transfer contributes to the acquisition of new adaptive traits important for survival. However, these foreign DNA elements can be deleterious and even lead to cell death in the case of phage infection. The constant need to maintain the balance between DNA uptake and defense processes would drive the genome evolution. To cope with the presence of invaders, prokaryotes have developed efficient defense systems including recently discovered CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR-associated) systems (1).

The CRISPR-Cas systems are found in about half of sequenced bacterial genomes and in almost all archaeal genomes (2). These prokaryotic adaptive immunity systems provide defense against foreign nucleic acids (3,4). CRISPR loci are arranged in arrays of almost identical direct repeats of ~30 bp separated by similarly sized variable sequences called spacers. Some spacers match viral or plasmid DNA and have been acquired during prior encounters with mobile genetic elements in 'adaptation' step (5). CRISPR arrays are transcribed as single RNA transcripts (pre-crRNA) that are processed to generate small CRISPR RNAs (crRNAs). In complex with Cas proteins, these crRNAs interfere with bacteriophage infection and plasmid transfer

*To whom correspondence should be addressed. Tel: +33 169 154 631; Fax: +33 169 157 808; Email: olga.soutourina@i2bc.paris-saclay.fr

†These authors contributed equally to the paper as first authors.

by recognizing foreign nucleic acids. This complementary base-pairing recognition leads to the destruction of targeted nucleic acids during an ‘interference’ process, thus protecting cells from the invasion by foreign genetic elements. The Cas proteins are involved in all stages of CRISPR-Cas activity. Universal Cas1 and Cas2 components are required for adaptation, while Cas6 proteins are necessary for crRNA processing. Either a single Cas protein or a multisubunit Cas proteins complex together with the mature crRNA achieve the interference step (1). During bacterial infection, vegetative cells survive in phage-rich gut communities and the bacteria could control the genetic exchanges favoured within this environment by relying on efficient anti-invader defense systems including CRISPR-Cas (6–8).

The human pathogen *Clostridium difficile* is an anaerobic spore-forming bacterium constituting the major cause of antibiotic-therapy-associated nosocomial diarrhoea in adults (9). This enteropathogen can lead to a variety of pathologies ranging from diarrhoea to pseudomembranous colitis, a potentially lethal disease. Transmission of *C. difficile* is mediated by contamination of the gut by spores. Antimicrobial therapy disturbs the colonic microflora allowing colonization of the intestinal tract by *C. difficile* from pre-existing or acquired spores (10,11). After spore germination and multiplication of vegetative cells, the pathogen produces either one or both of the two toxins (TcdA and TcdB) that are the major virulence factors. These two large toxins induce alterations in the actin cytoskeleton of intestinal epithelial cells (12,13). Yet, many aspects of *C. difficile* pathogenicity and its regulation still remain poorly understood. Regulatory RNAs may contribute to several steps during infection. It is increasingly recognized that bacterial non-coding RNAs (ncRNAs) play a critical role in adaptive responses and in various metabolic, physiological and pathogenic processes (14). By combining *in silico* analysis, RNAseq and genome-wide promoter mapping, we have recently identified more than 200 ncRNAs in *C. difficile* (15). This includes riboswitches, *trans*-acting riboregulators and *cis*-acting antisense RNAs, but also CRISPR RNAs that are among the most abundant RNAs revealed by our deep sequencing analysis.

We have recently provided experimental evidence for the function of CRISPR-Cas system in *C. difficile* (16). The *C. difficile* CRISPR-Cas system is characterized by the presence of an unusually large set of CRISPR arrays (an average of 8.5 per genome), the presence of two sets of *cas* genes conserved in almost all sequenced *C. difficile* strains and the prophage location of several CRISPR arrays (16–18). Both complete and partial *C. difficile cas* gene operons belong to the less characterized I-B subtype. Phage genome sequencing and CRISPR spacer homology analysis revealed a correlation with host range of several newly sequenced *C. difficile* phages (16,18). We demonstrated the role of *cas* genes in an heterologous host, *Escherichia coli*, and the defensive function of the *C. difficile* CRISPR-Cas system in an active interference process by analysis of plasmid conjugation efficiency in *C. difficile* (16). CRISPR arrays location within prophages in *C. difficile* is rather unique. High transmissibility of CRISPR systems and their association with plasmids, megaplasmids and in some cases prophages have been suggested (19). However, why CRISPR arrays are lo-

calated within phages and plasmids remains unknown. One hypothesis would be the stabilization of loci against loss and competition with other invaders (19).

We report here that most of the CRISPR arrays are colocalised with toxin-antitoxin (TA) systems in the *C. difficile* genome. TA modules encode two-component systems consisting of a stable ‘toxin’ and an unstable ‘antitoxin’ (20). The overexpression of toxin either kills cells or confers growth stasis. TA systems have been initially discovered on plasmids where they confer stability of maintenance through post-segregation killing (21). Plasmid loss results in a rapid decrease in levels of the unstable antitoxin, which allows the stable toxin to kill the plasmid-free cells. TA systems have also been found on bacterial and archaeal chromosomes, sometimes in great numbers but their function remains largely unclear. Among suggested functions are prophage maintenance, chromosomal region stabilization, prevention of phage infection, stress response and persister formation (20,22–27).

TA systems are classified into six types depending on the nature and action of the antitoxin that can be either a protein or a small antisense RNA (20). In type I systems, the antitoxin is a small antisense RNA that forms RNA duplex with the toxin-encoding mRNA (28,29). Most studies are devoted to type II TA systems, in which the protein antitoxin sequestering the toxin is more easily defined than the RNA antitoxin of type I TA (30). Numerous identified TA modules, generally of type II, are part of the mobilome including phages, plasmids, transposons and integrative and conjugative elements that can be shared by distant bacteria thus contributing to bacterial evolution (21,31). RNA antitoxins belong to the largest and most extensively studied set of sRNA regulators that act by modulating the translation and/or stability of their mRNA targets. Most of type I toxins are small hydrophobic proteins of <60 amino acids containing a potential transmembrane domain and charged amino acids at the C-terminus (32). In many cases, they seem to act like phage holins by inducing pores into cell membranes and thus impairing adenosine triphosphate synthesis (29). Replication, transcription and translation are consequently inhibited, which leads to cell death. However, alternative mechanisms of action have been also suggested for both membrane-associated toxins like BsrG and cytoplasmic toxins for example RalR with nuclease activity (33–35).

In this work, we describe the first identification of type I TA systems in *C. difficile* and highlight their association with CRISPR-Cas defense system. Through the analysis of deep-sequencing data for the strain 630 Δ *erm*, six potential TA loci were identified in the close proximity of transcribed CRISPR arrays. Three of these TA loci have been selected and the structure of their overlapping transcripts confirmed by Northern blot, 5’/3’RACE and reverse transcription and quantitative real-time polymerase chain reaction (qRT-PCR) analyses. The small proteins of unknown function in the proximity of CRISPR arrays have all the characteristic sequence features of type I toxins. We provide experimental evidence for the membrane localization and the toxic nature of these small proteins. Inducible toxin overexpression led to growth arrest of *C. difficile*, which was abolished by the co-expression of the RNA antitoxin

in cis and *in trans*. Using half-life measurements to assess the toxin mRNA and the RNA antitoxin stability, we show that the CRISPR-associated TA loci encode a rather stable toxin and unstable antitoxin RNA. By RNA band shift analysis we show an efficient duplex formation between TA transcripts *in vitro*. Finally, our results suggest that both the CRISPR-Cas genes and the type I TA modules could be co-regulated by stress- and biofilm-related factors, supporting a possible link between these two systems. The genomic analysis of more than 2,500 *C. difficile* strains revealed the general co-localisation of CRISPR arrays with potential TA modules, some of them located within the prophage regions, expanding our conclusions to the majority of the sequenced *C. difficile* strains.

MATERIALS AND METHODS

Plasmid and bacterial strain construction and growth conditions

C. difficile and *E. coli* strains and plasmids used in this study are presented in Supplementary Table S1. *C. difficile* strains were grown anaerobically (5% H₂, 5% CO₂ and 90% N₂) in TY (36) or Brain Heart Infusion (BHI, Difco) media in an anaerobic chamber (Jacomex). When necessary, cefoxitin (Cfx; 25 µg/ml) and thiamphenicol (Tm; 15 µg/ml) were added to *C. difficile* cultures. *E. coli* strains were grown in LB broth (37), and when needed, ampicillin (100 µg/ml) or chloramphenicol (15 µg/ml) was added to the culture medium. The non-antibiotic analog anhydrotetracycline (ATc) was used for induction of the *P*_{tet} promoter of pRPF185 vector derivatives in *C. difficile* (38). Strains carrying pRPF185 derivatives were generally grown in TY medium in the presence of 250 ng/ml ATc and 7.5 µg/ml Tm for 7.5 h. Growth curves were obtained using a GloMax plate reader (Promega).

All routine plasmid constructions were carried out using standard procedures (39). All primers used in this study are listed in Supplementary Table S2. For inducible expression of *C. difficile* genes, we used the pDIA6103 derivative of pRPF185 vector expression system lacking a *gusA* gene (15,38). The *CD2517.1* gene (−89 to +178 relative to the translational start site), the *CD2907.1* gene (−84 to +223 relative to the translational start site), *CD2517.1*-RCd8 TA region with RCd8 promoter (−306 to +504 relative to the translational start site of *CD2517.1*) and *CD2907.1*-RCd9 TA region with RCd9 promoter (−294 to +456 relative to the translational start site of *CD2907.1*) were amplified by PCR and cloned into StuI and BamHI sites of pDIA6103 vector under the control of the ATc-inducible *P*_{tet} promoter giving pDIA6319, pDIA6195, pDIA6202 and pDIA6196, respectively.

The knockdown antisense system on pRPF185 vector derivative was used to deplete the *C. difficile* 630Δ*erm* strain for the specific ribonucleases RNase III, RNase J and RNase Y. The *mecS* gene fragment comprising part of 5' untranslated region (UTR) and the beginning of the *mecS* coding part (−39 to +188 relative to the translational start site) was amplified by PCR on *C. difficile* 630Δ*erm* strain genomic DNA and cloned into StuI and BamHI sites of pRPF185 vector in antisense orientation under the control

of the ATc-inducible *P*_{tet} promoter giving pDIA6126. Similar strategy was used to construct plasmids pDIA5975 and pDIA5977 for inducible expression of antisense RNA for RNase J and RNase Y genes (+7 to +217 and +55 to +210 relative to the transcriptional start site (TSS) identified by deep sequencing, respectively).

For subcellular localization of toxins we used reverse PCR approach to construct *CD2517.1*-HA and *CD2907.1*-HA-expressing plasmids on the basis of corresponding pDIA6103-derivatives with primers designed to introduce the HA-tag sequence at the C-terminal part of coding toxin regions, directly upstream the stop codon (Supplementary Table S2). To compare the action of short and long forms of antitoxins on cognate and non-cognate toxins when co-expressed either *in cis* or *in trans* (from a site distant from the vector MCS), we used reverse PCR approach and Gibson assembly to construct different plasmids on the basis of the corresponding pDIA6103-derivatives (pT) (Supplementary Tables S1, S2 and Supplementary methods). DNA sequencing was performed to verify plasmid constructs using pRPF185-specific primers IMV507 and IMV508. The resulting derivative pRPF185 plasmids were transformed into the *E. coli* HB101 (RP4) and subsequently mated with *C. difficile* 630Δ*erm* (40) (Supplementary Table S1). *C. difficile* transconjugants were selected by sub-culturing on BHI agar containing Tm (15 µg/ml) and Cfx (25 µg/ml).

Light microscopy

For light microscopy, bacterial cells were observed at 100× magnification on an Axioskop Zeiss Light Microscope. Cell length was estimated for more than 100 cells for each strain using ImageJ software (41).

RNA extraction, quantitative real-time PCR, northern blot and 5'/3'RACE

Total RNA was isolated from *C. difficile* strains grown 7.5 h in TY medium containing 7.5 µg/ml of Tm and 250 ng/ml of ATc as previously described (42). For biofilm samples *C. difficile* 630Δ*erm* strain was grown for 72 h in TY medium using continuous-flow microfermentor culture system (43). The 24-h planktonic culture in TY medium was used for comparative analysis. The cDNA synthesis by reverse transcription and qRT-PCR analysis were performed as previously described (44). In each sample, the relative expression for a gene was calculated relatively to the 16S rRNA gene or *dnaF* gene (*CD1305*) encoding DNA polymerase III or *ccpA* gene encoding catabolite control protein. The relative change in gene expression was recorded as the ratio of normalized target concentrations (ΔΔCt) (45). Northern blot analysis and 5'/3'RACE experiments were performed as previously described (15).

RNA band-shift assay and *in vitro* processing by RNase III

Templates for the synthesis of RNA probes were obtained by PCR amplification using the Term and T7 oligonucleotides (Supplementary Table S2). RNAs were synthesized by T7 RNA polymerase with [α-³²P] UTP as a tracer

and were then gel purified. RNA concentrations were monitored by counting out the radioactivity and the RNA samples were stored until use (46). This gives a 298-nt long *CD2907.1* and a 132-nt long RCd9 transcripts with three additional G at the 5' extremity. Just before use, RCd9 RNA was 5'-radiolabeled and incubated with increasing concentrations of *CD2907.1* mRNA under two different conditions referred as N (Native) and F (Full RNA duplex) conditions, respectively. Radiolabeled RCd9 was incubated either alone or with the unlabeled *CD2907.1* RNA to allow them to anneal in Tris-Mg-acetate-Na-acetate (TMN) buffer for 5 min at 37°C (20 mM Tris acetate, pH 7.5, 10 mM magnesium acetate, 100 mM sodium acetate). Alternatively, after denaturation at 90°C for 2 min, labeled RCd9 RNAs in 1×TE were incubated 30 min at 37°C with the unlabeled *CD2907.1* to allow them to anneal. The complexes were immediately loaded on native polyacrylamide gels to control for hybridization efficiency (47) or submitted to *in vitro* processing by RNase III of *E. coli*. RNase III digestion of free or complexed RCd9 was performed at 37°C in TMN buffer containing 1 µg tRNA from 1 min to 15 min with 0.05 units of RNase III (Epicentre). After precipitation, addition of loading buffer and heat denaturation, samples were analyzed on 8% polyacrylamide-Urea gels.

Subcellular localization of HA-tagged toxins by cell fractionation and western blotting

The *C. difficile* cultures were inoculated from overnight grown cells in 10 ml of TY medium at OD 600 nm of 0.05, allowed to grow for 3 h before addition of 250 ng/ml ATc and incubation for 90 min followed by centrifugation and protein extraction. Cell lysis, fractionation and protein analysis were performed as previously described (48). Coomassie staining was performed for loading and fractionation control. Western blotting was performed as previously described (49) with anti-HA antibodies.

Measurement of RNA decay by rifampicin assay

For determination of toxin and antitoxin RNA half-lives the *C. difficile* strains were grown in TY medium supplemented with 250 ng/ml ATc and 7.5 µg/ml Tm for 7.5 h at 37°C. Samples were taken at different times after addition of 200 µg/ml rifampicin (0, 2, 5, 10, 20, 40, 60 and 120 min) and subjected to RNA preparation and northern blotting.

In silico screening for potential new TA genes and CRISPR arrays co-localization

The raw sequencing read data of 2,584 *C. difficile* strains were downloaded for this genomic analysis (16,50). For each strain, we realized an assembly with Spades (51) and an automatic annotation using PROKKA (52). Then we selected small proteins from 40 to 60 amino acids in length, adjacent to CRISPR arrays and performed an orthology analysis using proteinortho5 (53). Multiple alignment was done using ClustalW (54).

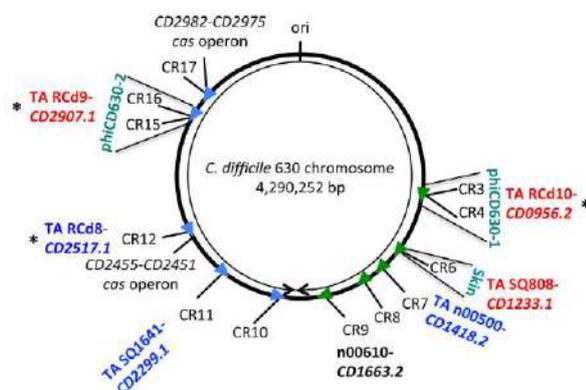


Figure 1. Genomic map of potential type I TA loci in association with CRISPR arrays in *Clostridium difficile* strain 630. Schematic view of the genomic location of expressed CRISPR arrays in strain 630. CRISPR arrays are numbered according to CRISPRdb database (2). Arrowheads indicate the array position and the transcriptional orientation. The location of the associated TA modules, the *cas* operons, the prophage regions and the replication origin (*ori*) are indicated. The right and left replichores are shown by arrows. The n00610 antisense RNA overlaps the *CD1663.2* gene, which encodes a small protein with a divergent sequence associated with CRISPR 9 array. The CRISPR-associated TA modules within prophage regions are RCd9-*CD2907.1*, RCd10-*CD0956.2* and SQ808-*CD1233.1*. * indicates the three TA modules that were selected for detailed analysis.

RESULTS

Identification of toxin-antitoxin system candidates in *C. difficile* genome

We have revisited our previously reported deep sequencing data (15) and observed an unusual transcriptional unit organization in the close proximity of CRISPR loci in the genome of *C. difficile* strain 630Δ*erm* (Supplementary Figure S1). The presence of several overlapping transcripts was detected by comparison of Tobacco Acid Pyrophosphatase treated (TAP+) and non-treated (TAP-) samples for TSS mapping. This analysis combined with the RNA-seq data for whole transcript coverage revealed that the majority of CRISPR RNA loci are associated with potential antisense RNAs of genes encoding small proteins of unknown function (Figure 1). A more detailed analysis of the nature of the overlapping convergent transcripts allowed us to identify candidates for six type I TA systems that co-localized with CRISPR 3/4, CRISPR 6, CRISPR 7, CRISPR 11, CRISPR 12 and CRISPR 16/15 arrays in *C. difficile* (Figure 1 and Supplementary Table S3) (32). An additional pair of antisense RNA and small protein gene near CRISPR 9 array had divergent sequence without common type I TA features. Supplementary Table S3 summarizes the data on the candidate antitoxin RNAs including the 5'-end determination of the transcripts by global TSS mapping and the position of the 3'-ends deduced from RNA-seq data.

Interestingly, three potential TA modules together with associated CRISPR arrays are located within prophage regions (Figure 1). The pathogenicity-island location of type I TA modules has been reported in *Staphylococcus aureus* (55,56). The CRISPR 3/4 array, the associated potential toxin gene *CD0956.2* and the antitoxin gene are located

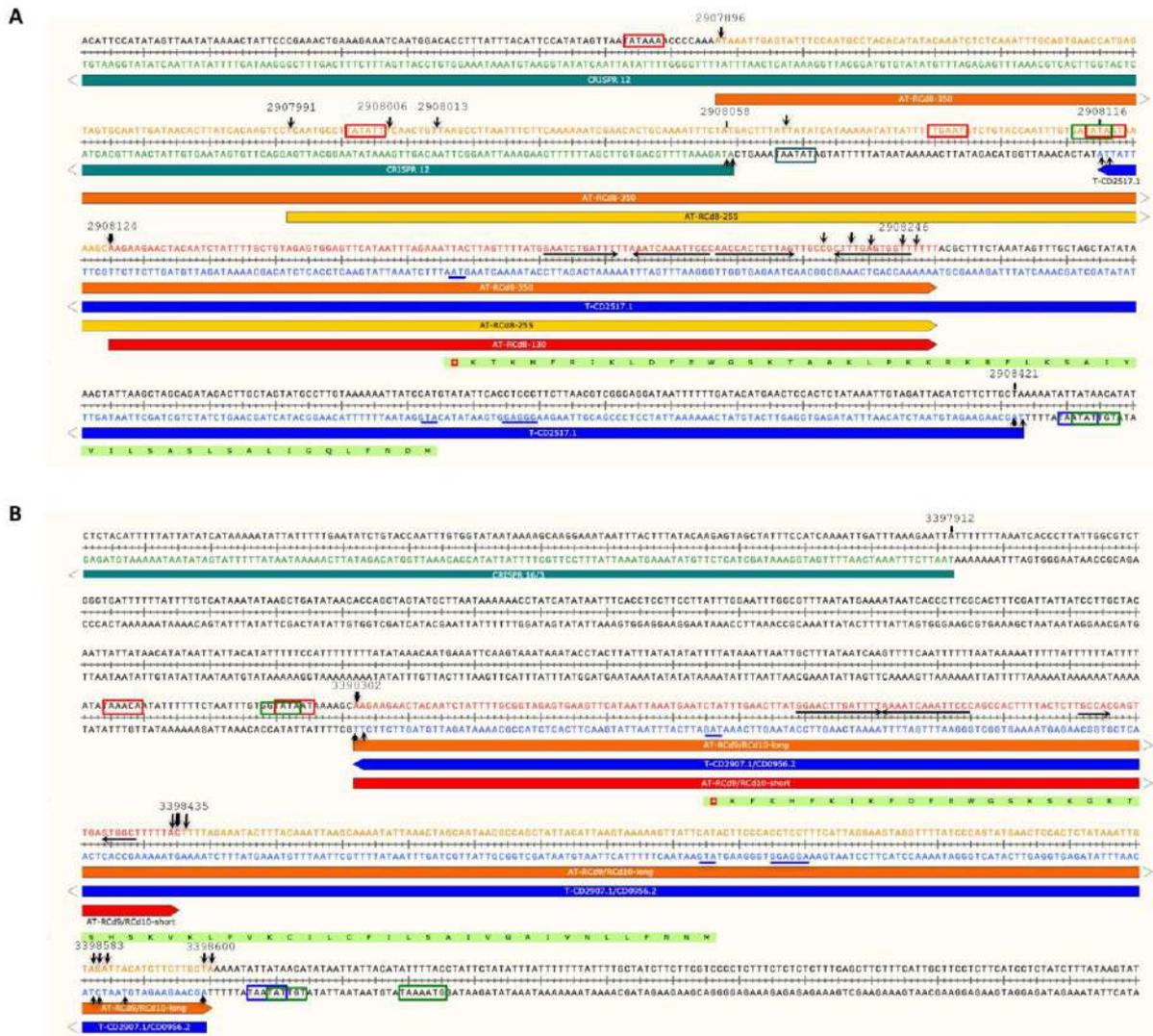


Figure 2. Schematic representation of potential type I TA loci in *C. difficile* chromosome: (A) CRISPR 12 region, (B) CRISPR 16/15 (CRISPR 3/4) regions. The horizontal arrows below the double-stranded sequences show the toxin, antitoxin and CRISPR transcripts, the direction of transcription is indicated by arrowheads. The transcriptional start sites for sense and antisense transcripts identified by 5'/3'RACE and TSS mapping are indicated by vertical arrows with their genomic location. Line thickness corresponds to the proportion of observed extremities. The genomic location of 5'- and 3'-ends of the transcripts are indicated above the sequence. The inverted repeats at the position of transcriptional terminators are indicated by black arrows. The positions of Sigma A-dependent promoter -10 and -35 elements of antitoxin (AT) and toxin (T) are shown in boxes. The positions of ribosome binding site, translation initiation codon and stop codon of toxin (T) mRNA are underlined. The positions of Sigma B-dependent promoter elements are shown in boxes for both TA genes.

within the phiCD630-1 prophage region while the identical CRISPR 16/15 array and the potential TA module containing *CD2907.1* toxin and antitoxin genes are located within the phiCD630-2 prophage region (Figure 1). Similar to the *txpA/RatA* type I TA module in *Bacillus subtilis* (57), the *CD1233.1/SQ808* pair is located within the *skin* element of *C. difficile* strain 630 (58), yet there is no sequence homology between the two loci. This *CD1233.1/SQ808* pair is located near the CRISPR 6 array in the *C. difficile skin* element.

We have chosen three representative type I TA modules for further detailed analysis. The RCd8-*CD2517.1* module is located near the CRISPR 12 array, which is associated with a partial *cas* operon. The RCd9-*CD2907.1* and RCd10-*CD0956.2* modules are located near the CRISPR 16/15 and CRISPR 3/4 arrays. They lie respectively within the phiCD630-1 and phiCD630-2 prophage regions, which have identical sequences and are thus indistinguishable from each other through gene expression analysis. These two highly similar prophages phiCD630-1 (1088001-1143874)

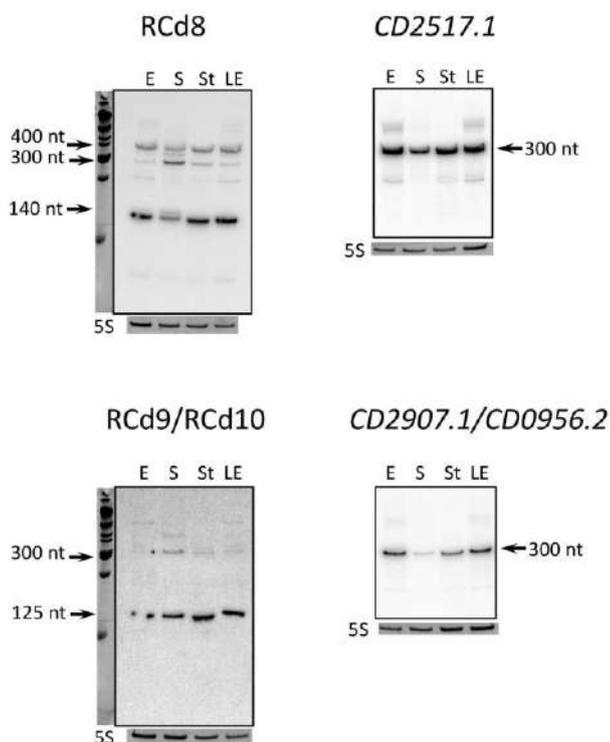


Figure 3. Detection of antisense RNA and toxin mRNA by northern blot. RNA samples were extracted from 630 Δ erm strain grown at exponential phase (E, 4 h of growth), late exponential phase (LE, 6 h of growth), entry to stationary phase (S, 10 h of growth) or under nutrient starvation conditions (St). 5S RNA at the bottom serves as loading control. The arrows show the detected transcripts with their size estimated by comparison with RNA molecular weight standards. As indicated at the top, the blots were hybridized either with antitoxin- or toxin-specific probe. The same 5S control panel is shown when reprobings of the same membrane was performed.

and phiCD630-2 (3377033-3434358) of 55.9 and 57.3 kb in length, respectively, are located in an inverted orientation on different replichores of the *C. difficile* chromosome. The regions encoding TA modules and CRISPR arrays are identical. We assigned the names RCd8 (previously named SQ1781), RCd9 (previously named CD630_n01000) and RCd10 (previously named CD630_n00370) to the putative antitoxin RNAs (Figure 1). We have previously demonstrated the active expression of the CRISPR arrays associated with the chosen modules and their functionality for interference with plasmid DNA (16).

We first mapped by 5'/3'RACE analysis the transcriptional start and termination sites of the genes corresponding to the potential toxin and the antitoxin RNAs for selected loci. Figure 2 shows the chromosomal organization of these genes and the position of 5' and 3' ends of overlapping transcripts identified by 5'/3'RACE (Table 1, Supplementary Table S3 and Figure S2). The alignment of the TA genomic regions revealed the presence of conserved sequences upstream of the TSS for both the putative toxin and antitoxin genes and allowed the identification of consensus elements for Sigma A-dependent promoters upstream of their TSS

(Supplementary Table S3, indicated in blue and red in Supplementary Figure S2). Moreover, the consensus sequence promoters recognized by the alternative Sigma factor of the general stress response, Sigma B, could be identified upstream of the TSS of both the potential antitoxin and toxin genes (indicated in green in Supplementary Figure S2).

We then confirmed the detection of the candidate RNA antitoxin transcripts by northern blot under several growth conditions (Figure 3). Transcript length deduced from TSS mapping, RNA-seq and 5'/3'RACE analysis agreed generally well with the size of RNAs detected by northern blotting. The RCd8 RNA is transcribed in antisense direction to the *CD2517.1* gene and overlaps with the CRISPR 12 array (Figure 2A). An abundant transcript of about 140 nt was stably detected by northern blotting under all tested conditions (exponential growth phase, late exponential phase, onset of stationary phase or starvation) (Figure 3). The presence of additional less abundant longer transcripts of about 300 and 400 nt was consistent with 5'/3'RACE results (Table 1) and suggests that they could result from a detected alternative TSS (Figure 2A). The major 140-nt transcript starts with a transcriptional +1 associated with a conserved promoter sequence and stops at the rho-independent terminator (Figure 2 and Supplementary Figure S2).

The RCd9 RNA is transcribed in antisense direction to the *CD2907.1* gene adjacent to CRISPR 16/15 array (Figure 2B). An identical region within the homologous phiCD630-1 prophage encodes the RCd10 RNA transcribed in antisense orientation to the *CD0956.2* gene adjacent to CRISPR 3/4 array. An abundant transcript of about 125 nt was detected with RCd9/RCd10-specific probe under all tested conditions by northern blotting in addition to a hardly detectable longer transcript of about 300 nt in accordance with the 134 and 283-nt transcript sizes deduced from the 5'/3'RACE data (Figure 3, Table 1 and Supplementary Table S3). The relative abundance of these transcripts is consistent with the 5'/3'RACE data where most of the 3' ends mapped to the rho-independent terminator position for a short transcript (Figure 2).

To monitor the toxin mRNA expression, we rehybridized the northern blots with the probes matching to *CD2517.1* and *CD2907.1*. In accordance with the 5'/3'RACE data, an abundant transcript of about 300 nt was detected under all tested conditions (Figure 3). A decreased intensity of the toxin mRNA signal was observed for the onset of stationary phase. It is worth noting that a slight decrease in the signal of the antitoxin corresponding to short transcripts was also detected under these conditions with the concomitant appearance of longer transcripts (Figure 3).

From these data we infer that several RNA transcripts for RCd8 of 140, 255 and 350-nt long are detected while the *CD2517.1* mRNA is 306-nt long. The RCd8 and *CD2517.1* RNAs overlap by 131 nt. The RCd9/RCd10 RNA transcripts are 134 and 283-nt long and overlap the 298-nt long *CD2907.1/CD0956.2* mRNAs by 283 or 134 nt for long and short transcripts, respectively. Secondary structure prediction by Mfold revealed that potential antitoxin RCd8 and RCd9/RCd10 RNAs are highly structured (data not shown). Corresponding overlapping mRNA are also highly structured and contain double-stranded secondary structure regions sequestering their ribosome binding sites that

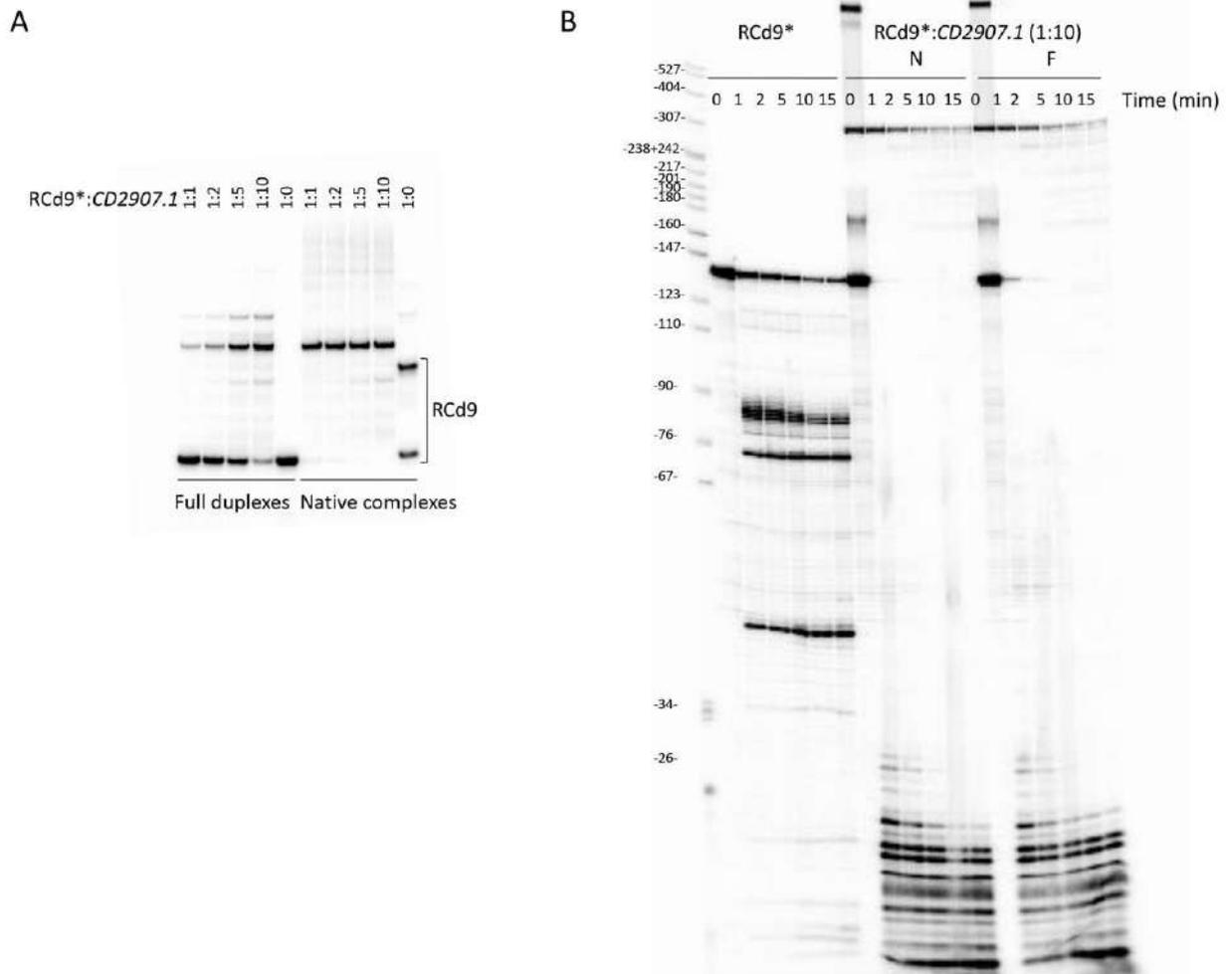


Figure 4. Analysis of TA RNA duplex formation by RNA band-shift assay and *in vitro* processing by RNase III. (A) 5'-radiolabeled RCd9 RNA was incubated with increasing concentrations of CD2907.1 under two different conditions referred as N (Native) and F (Full RNA duplex) conditions, respectively. Native RCd9:CD2907.1 complexes were formed at 37°C for 5 min in TMN buffer, and full duplexes were obtained after a denaturation-annealing treatment in TE Buffer (2 min 90°C, 30 min 37°C). The complexes were immediately loaded on native polyacrylamide gels to control for hybridization efficiency or submitted to *in vitro* processing by RNase III (B). RNase III digestion of free or complexed RCd9 was performed at 37°C in TMN buffer containing 1 µg tRNA from 1 min to 15 min with 0.05 units of RNase III per sample. Samples were analyzed on 8% polyacrylamide-Urea gels.

Table 1. The antisense RNA and associated toxin mRNA extremity identification by 5'/3'RACE

Name	Description	5'-end RACE position	5'-end TSS mapping position	Strand	3'-end RACE position	Size, nt
RCd10 (CD630_n00370)	Antitoxin of TA associated with CRISPR 3/4	1124339	1124339	-	1124206, 1124058, 1124041	134, 283
CD0956.2	Toxin of TA associated with CRISPR 3/4	1124042	1124042	+	1124339	298
RCd9 (CD630_n01000)	Antitoxin of TA associated with CRISPR 16/15	3398302	3398302	+	3398435, 3398583, 3398600	134, 283
CD2907.1	Toxin of TA associated with CRISPR 16/15	3398599	3398599	-	3398302	298
RCd8 (SQ1781)	Antitoxin of TA associated with CRISPR 12	2907896, 2907991, 2908066	2908006, 2908013, 2908124	+	2908246	130, 255, 350
CD2517.1	Toxin of TA associated with CRISPR 12	2908421	2908421	-	2908116	306

The positions of 5'-start and 3'-end of these RNAs were identified by 5'/3'RACE analysis and compared with 5'-end identified by 5'-end RNA-seq analysis (TSS mapping).

could be important for the regulatory process as observed for *B. subtilis* type I toxin mRNAs (data not shown) (59–61). Moreover, the presence of 5' UTRs of 88 and 83 nt in length was observed for *CD2517.1* and *CD2907.1/CD0956.2* mRNAs, respectively (Supplementary Figure S2). The comparison between the secondary structure predictions for the TA pair transcripts suggests potential loop–loop interactions between stem–loop structures that could initiate the RNA duplex formation (data not shown) (59).

We have chosen representative abundant transcripts from TA pair RCd9/*CD2907.1* for the analysis of duplex formation between toxin and antitoxin RNA by gel retardation assays. We investigated the interaction between RCd9 and *CD2907.1* to determine whether they form a kissing complex as in the case of the antisense RNA, CopA and its target CopT (62). Figure 4A shows that under native conditions RCd9 harbors two conformations both proficient to duplex formation upon incubation with equimolar amount of *CD2907.1* transcript suggesting that the primary binding intermediate could involve a loop–loop interaction.

We then investigated RNase III-dependent cleavages of RCd9 alone and in interaction with *CD2907.1* under conditions of native complex formation or after extensive denaturation to define the extent of duplex formation (Figure 4B). RNase III cleaves RCd9 at 3 sites located about 82, 72 and 42 nt from the 5' extremity. In contrast, RNase III very efficiently and rapidly degrades the duplexes formed under native conditions or after extensive denaturation. This indicates that RCd9 is fully hybridized to the *CD2907.1* toxin mRNA and that the kissing intermediate formed under native conditions of hybridization could be converted into a full duplex leading to the RNase III cleavage of the entire duplex.

Functionality of toxin–antitoxin systems in *C. difficile*

Type I toxins are generally small hydrophobic proteins of <60 amino acids containing a potential transmembrane domain and charged amino acids at the C-terminus (32). The alignment of proteins from the potential TA modules encoded in the proximity of CRISPR arrays revealed that these small proteins have all characteristic features of type I toxins. Indeed, as shown in Figure 5A, the potential toxic proteins are from 50 to 53 amino acids in length, carry a conserved hydrophobic region at their N-terminal part and a lysine-rich, positively charged region at their C-terminal part in agreement with the hydrophobicity profile predictions by Kyte and Doolittle algorithm (data not shown). Transmembrane domain location in N-terminal moiety was predicted by TMHMM program (data not shown). To experimentally identify the expression and localization of these small proteins in *C. difficile* we constructed plasmids expressing under inducible P_{tet} promoter either *CD2517.1* or *CD2907.1/CD0956.2* fused with a HA tag at the C-terminus (Supplementary Table S1). By western blotting with anti-HA antibodies, no signal was detected for whole cell extracts from control strains expressing untagged proteins while a specific signal was detected for strains expressing HA-tagged proteins (Figure 5B). To precise the subcellular localization of these proteins we then performed cell fractionation and examined supernatant, cell

wall, membrane and cytosolic fractions by western blotting. As shown in Figure 5B, HA-tagged *CD2517.1* and *CD2907.1* (*CD0956.2*) were only detected in the membrane fractions of *C. difficile* cell extracts suggesting the association of these small proteins with the cell membrane in *C. difficile*.

To show the toxic nature of these small proteins, we analyzed the effect of their overexpression on the growth of *C. difficile* cells in liquid and solid media. HA-tagged proteins *CD2517.1* and *CD2907.1/CD0956.2* conserved their toxic activity on cell growth when overexpressed from plasmids used for determination of their subcellular localization by western blotting (Supplementary Figure S3 and Figure 5B). This result suggests that despite the presence of HA-tag these small proteins remain active for cell growth inhibition.

We then generated plasmids allowing either inducible overexpression of an untagged version of one of the small, potentially toxic proteins or simultaneous expression of both the potential toxin and the antisense RNA for the TA modules near the CRISPR 12 and CRISPR 16/15 (CRISPR 3/4) arrays. For this purpose, we cloned either the small protein-coding region with its ribosome-binding site (RBS) (*CD2517.1* or *CD2907.1/CD0956.2*) under the control of the inducible P_{tet} promoter (pT) or the entire potential TA module (pTA). pTA constructs allow both the inducible overexpression of the putative toxin under the control of the P_{tet} promoter and the expression of the antisense RNA from its own strong promoter (Figure 6A). *C. difficile* strain 630 Δerm carrying an empty vector (p) was used as a control. No growth difference was observed for any of the three strains on BHI plates in the absence of AnhydroTetracycline (ATc) inducer for both potential TA modules (Figures 6A and 7A). By contrast, a dramatic growth defect was observed on BHI plates in the presence of ATc inducer for the strain overexpressing the genes *CD2517.1* or *CD2907.1/CD0956.2* (Figures 6A and 7A). Co-expression of these potential toxins with the associated RNA antitoxins led to the full or partial reversion of the growth defect for both TA modules (Figures 6A and 7A). Consistently, northern blotting with RCd8 and RCd9/RCd10-specific probes using RNA extracted from the strain 630 Δerm carrying pT or pTA confirmed the important overexpression of the antitoxin RNA from the pTA constructs as compared to the level of expression from their chromosomal location in control strain carrying an empty vector (p) (Figures 6B and 7B).

The overexpression of toxins from selected TA modules (Figures 6B and 7B) also induced rapid growth arrest in liquid culture. As shown in Figure 6C for *CD2517.1*-RCd8 TA module, the addition of ATc inducer after 3h of exponential growth led to rapid growth arrest for the strain carrying the pT plasmid but allowed near normal growth of the *C. difficile* 630 Δerm strain carrying pTA. Similar deleterious growth effects were observed for the strain carrying the pT plasmid when strains pre-grown overnight in the absence of inducer and then diluted in an ATc-containing medium were allowed to grow for 24 h in an automatic plate reader (Figure 6D). For the *CD2907.1*-RCd9/*CD0956.2*-RCd10 TA module, we observed only a partial reversion of the growth defect in liquid culture associated with the toxin gene expression when both toxin and antitoxin were co-expressed on pTA plasmid (Figure 7C). This partial restora-

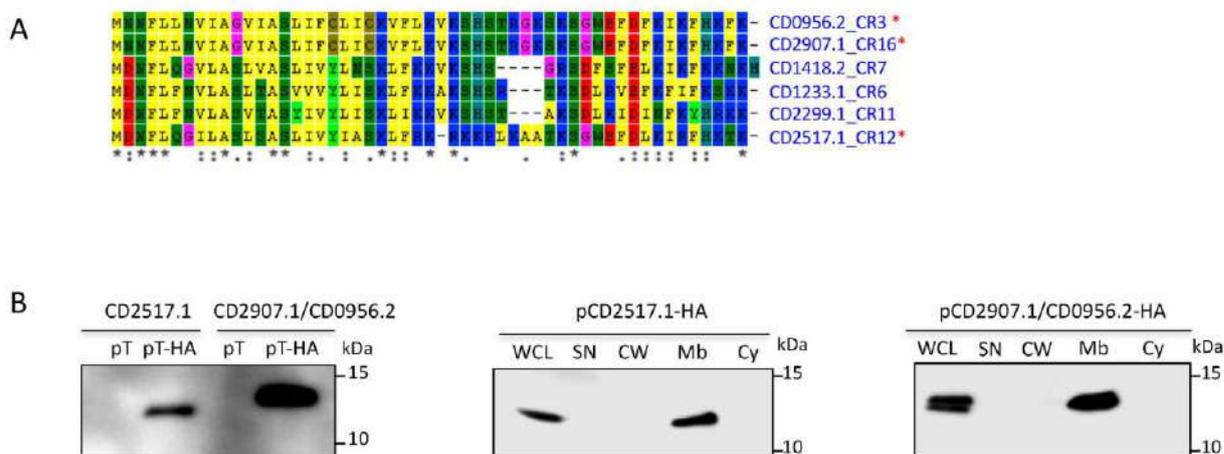


Figure 5. Potential type I toxin proteins alignment and analysis. (A) Proteins alignment using ClustalW. "*" on the right indicates toxins from three TA modules selected for detailed analysis. "*" at the bottom indicates conserved residues. (B) Western-blot detection and localization of HA-tagged small proteins in the membrane fraction of *C. difficile* cell extracts. WCL: whole cell lysate; SN: supernatant; CW: cell wall; Mb: membrane; Cy: cytosolic fraction. Immunoblotting with anti-HA antibodies detected a major polypeptide of ~10 kDa in whole cell lysates of the strain carrying P_{tet} -T(CD2517.1 or CD2907.1/CD0956.2)-HA (pT-HA) construct grown in the presence of the 250 ng/ml ATc inducer but not in extracts of strains expressing non-tagged toxins (pT) (left panel). The culture of strains carrying P_{tet} -T-HA plasmids induced with 250 ng/ml ATc was fractionated into cell wall (CW), membrane (Mb) and cytosolic (Cy) compartments and immunoblotted with anti-HA antibodies (middle and right panels). Proteins were separated on 12% Bis-Tris polyacrylamide gels in MES buffer.

tion of growth could be due to an unbalance in the relative level of toxin and antitoxin expression. Interestingly, northern blotting revealed a reverse correlation between the relative toxin and antitoxin transcript abundance under inducing conditions. Toxin overexpression after ATc induction led to more than 2-fold decrease in the amount of antitoxin transcript expressed from chromosomal location compared to the strain $630\Delta erm$ containing the vector alone (lanes 'pT' versus 'p' in Figures 6B and 7B).

Toxins from TA modules in *B. subtilis* and *Enterococcus faecalis* have been reported to affect cell envelope biosynthesis, nucleoid condensation, cell division and chromosome segregation (35,63). To assess whether the changes in cell morphology could be induced by toxin overexpression in *C. difficile*, we analyzed by light microscopy liquid cultures of strain $630\Delta erm$ carrying the vector, pT or pTA 1 h after ATc addition. For both TA modules (CD2517.1 and CD2907.1/CD0956.2), the overexpression of the toxins in strain $630/pT$ led to a significant increase in cell length for about 9 and 5.4% of the cells, respectively. The length of these cells was above the value of $630/p$ mean length with two standard deviations ($10.5 \mu\text{m}$) (Figure 7D and Supplementary Figure S4). For control strain $630/p$ the length of only 1.7% of cells exceeded this value. Co-expression of the entire TA module (pTA) led to a partial reversion of this phenotype to the control culture morphology.

To get further insights into the requirements for repression and the function of abundant short and less abundant full-length antitoxin transcripts, we co-expressed *in trans* different antitoxin transcripts from their own promoter with the native or HA-tagged toxin proteins under the control of inducible P_{tet} promoter for both TA modules (Supplementary Table S1). For each combination, toxin and antitoxin were co-expressed from the same plasmid but

from distant locations. In the presence of ATc inducer, both short and full-length antitoxins expressed *in trans* were able to rescue the strains from the toxicity associated with overexpression of the cognate native toxin both on plates and in liquid culture (Figure 8). The reversion of growth defect was similar to that observed with a control strain carrying the plasmid co-expressing the toxin from the inducible P_{tet} promoter and the cognate antitoxin *in cis* from the native convergent configuration (Figure 8). Cognate short antitoxin co-expression *in trans* also led to the reversion of growth defect induced by the overexpression of the HA-tagged toxin (Supplementary Figure S5). These results suggest that the abundant short form of antitoxins is sufficient to repress the associated toxin and that the C-terminal HA-tag does not interfere with antitoxin action. By contrast, a dramatic growth defect was still observed when the non-cognate antitoxins were co-expressed with the toxins (Figure 8). These results demonstrate that RCd8 and RCd9/RCd10 antitoxins act on a highly specific manner to repress their associated toxin not only when they are expressed from the native convergent TA configuration (Figures 6–8) but also when expressed *in trans* from a distant plasmid location (Figure 8).

Altogether our data demonstrate that functional type I TA modules are present in the proximity of CRISPR arrays in *C. difficile*.

Stability of the RNAs of the toxin and antitoxin genes

To determine the half-lives of toxin and antitoxin RNAs, *C. difficile* strains were grown in TY medium until the late-exponential growth phase and rifampicin was added to block transcription. Samples were taken at different time points after rifampicin addition for total RNA preparation and Northern blots were performed using probes tar-

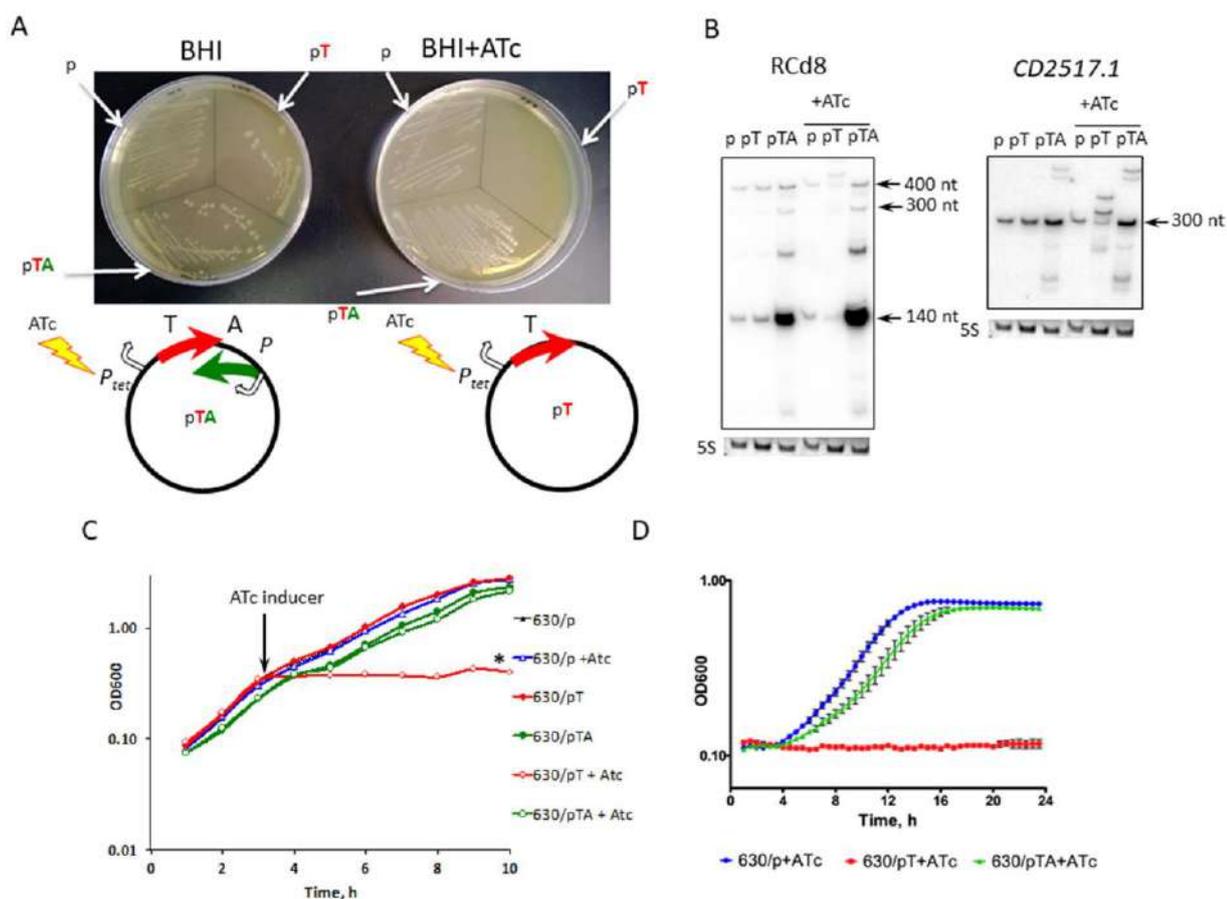


Figure 6. Effect of inducible toxin and TA overexpression for CD2517.1-RCd8 TA module near CRISPR 12 on growth in solid (A) and liquid medium (B–D). (A) Growth phenotype of *C. difficile* strains CDIP369 (630/p), CDIP357 (630/pT) and CDIP332 (630/pTA) on BHI agar plates supplemented with Tm alone (on the left) or with the addition of 500 ng/ml of ATc inducer (on the right) after 24 h of incubation at 37°C. Schematic representations of pT and pTA constructs are shown. The 630 strain carrying an empty vector (p) is used as a control. (B) Detection of RCd8 and *CD2517.1* transcripts. For northern blot analysis, RNA samples were extracted from 630/p control strain, from 630/pT strain overexpressing the CD2517.1 toxin and from 630/pTA strain overexpressing the entire TA module grown at late exponential growth phase in the presence of 250 ng/ml ATc (+ATc) or the absence of inducer. As indicated at the top, the blots were hybridized either with antitoxin- or toxin-specific probe. The same 5S control panel is shown when reprobing of the same membrane was performed. (C) Growth of 630/p strain (triangles), 630/pT strain (diamond) and 630/pTA strain (circle) in TY medium at 37°C in the presence (open symbols) or absence (closed symbols) of 250 ng/ml ATc. The time point of ATc addition is indicated by an arrow. (D) Growth curves for 630/p strain, 630/pT strain and 630/pTA strain in TY medium at 37°C in the presence of 250 ng/ml ATc using a GloMax plate reader (Promega). The mean values and standard deviations are shown for three independent experiments.

getting the toxin or antitoxin RNAs of one representative TA module. We have used a control strain 630 Δ erm carrying an empty vector (CDIP369) to allow further comparison with strains depleted for ribonucleases (see below). In this strain, quantification of the northern blots with corresponding probes allowed us to estimate the half-life of the mRNA of the *CD2907.1/CD0956.2* toxin gene to about 35 min (Figure 9A). The half-life of the 125-nt transcript for RCd9/RCd10 antitoxin RNA was estimated to be about 13 min (Figure 9B). These results further confirm that the identified type I TA modules produce a rather stable toxin mRNA and a less stable antitoxin RNA.

In model Gram-positive bacterium *B. subtilis*, the double-strand-specific enzyme RNase III plays an essential role in the degradation of toxin mRNA from prophage-

encoded type I TA modules (59,64) while the single-strand specific endoribonuclease RNase Y and the 5'-3' exoribonuclease RNase J1 participate in antitoxin RNA degradation (59). To analyse the possible contribution of these ribonucleases to the degradation of toxin and antitoxin RNAs, we tested the effect of RNase depletion on the stability of the corresponding transcripts. We also evaluated the effect of the depletion for the RNA chaperone protein Hfq. *C. difficile* 630 encodes CD1289 (*rni*), CD1329 (*rny*) and CD1248 (*rncS*) proteins homologous to *B. subtilis* RNase J, RNase Y and RNase III. No transposon insertions were identified in a previously reported TraDIS (transposon-directed insertion site sequencing) experiment for these RNase genes suggesting their crucial role for *C. difficile* physiology (65). We have previously reported the use of a knock-down strategy

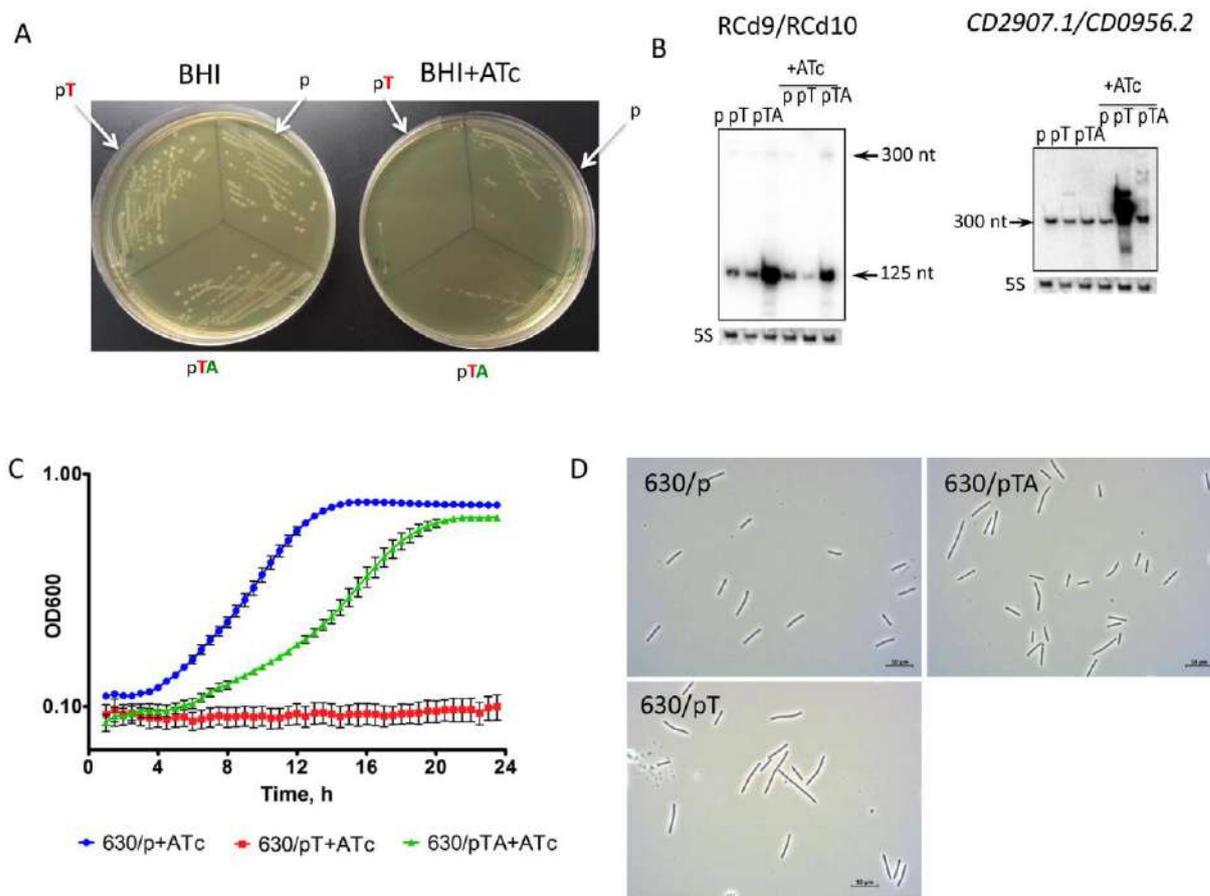


Figure 7. Effect of inducible toxin and TA overexpression for CD2907.1-RCd9/CD0956.2-RCd10 TA module near CRISPR 16/15 (CRISPR 3/4) on growth in solid (A) and liquid (B–D) medium. (A) Growth of *C. difficile* strains CDIP369 (630/p), CDIP317 (630/pT) and CDIP319 (630/pTA) on BHI agar plates supplemented with Tm alone (on the left) or with the addition of 500 ng/ml of ATc inducer (on the right) after 24 h of incubation at 37°C. Under inducing conditions 630/pT strain overexpresses CD2907.1 toxin and 630/pTA strain overexpresses entire TA module. The 630 strain carrying an empty vector (p) is used as a control. (B) Detection of RCd9/RCd10 and CD2907.1/CD0956.2 transcripts. For northern blot analysis, RNA samples were extracted from 630/p control strain, 630/pT strain overexpressing CD2907.1 and 630/pTA strain overexpressing the entire TA module grown at late exponential growth phase in the presence of 250 ng/ml ATc (+ATc) or the absence of inducer. As indicated at the top, the blots were hybridized either with antitoxin- or toxin-specific probe. The same 5S control panel is shown when reprobing of the same membrane was performed. (C) Growth of 630/p strain, 630/pT strain and 630/pTA strain in TY medium at 37°C in the presence of 250 ng/ml ATc using a GloMax plate reader (Promega). The mean values and standard deviations are shown for three independent experiments. (D) Selected images from light microscopy observation of 630/p, 630/pT and 630/pTA strains grown in TY medium at 37°C after 1 h of 250 ng/ml ATc addition.

based on the expression of an inducible antisense RNA targeting the 5' part of the coding region of the *hfq* gene leading to Hfq depletion in *C. difficile* (49). Here, we used a similar strategy for the construction of *C. difficile* strains in which we can deplete for RNase J, RNase Y and RNase III using an antisense RNA targeting each of these genes expressed under the control of a P_{tet} inducible promoter (Supplementary Table S1). The efficient depletion conditions in the presence of ATc were confirmed by western blotting (Supplementary Figure S6).

The half-lives of analyzed toxin and antitoxin RNAs were similar in the wild-type and in the Hfq depleted strain suggesting that neither of the two RNAs is stabilized by Hfq (Supplementary Figure S7). We observed a moderate stabilization of the CD2907.1/CD0956.2 toxin mRNA in the

strain depleted for RNase III but also, at a higher level, in the strains depleted for RNase J or RNase Y (Supplementary Figure S7A). These results suggest the possible contribution of the RNase J, RNase Y and in less extent of RNase III to the degradation of toxin mRNA. No major changes in the estimated half-life values were observed for the abundant transcript of antitoxin RNA RCd9/RCd10. However, RNase depletion resulted in observable changes in the degradation pattern of the antitoxin RNA RCd9/RCd10. We observed the appearance of a biphasic degradation curve and the accumulation of longer species especially for RNase Y depleted strain (Supplementary Figure S7B). Generally, the half-lives of longer transcripts increase under RNase depletion conditions. This re-

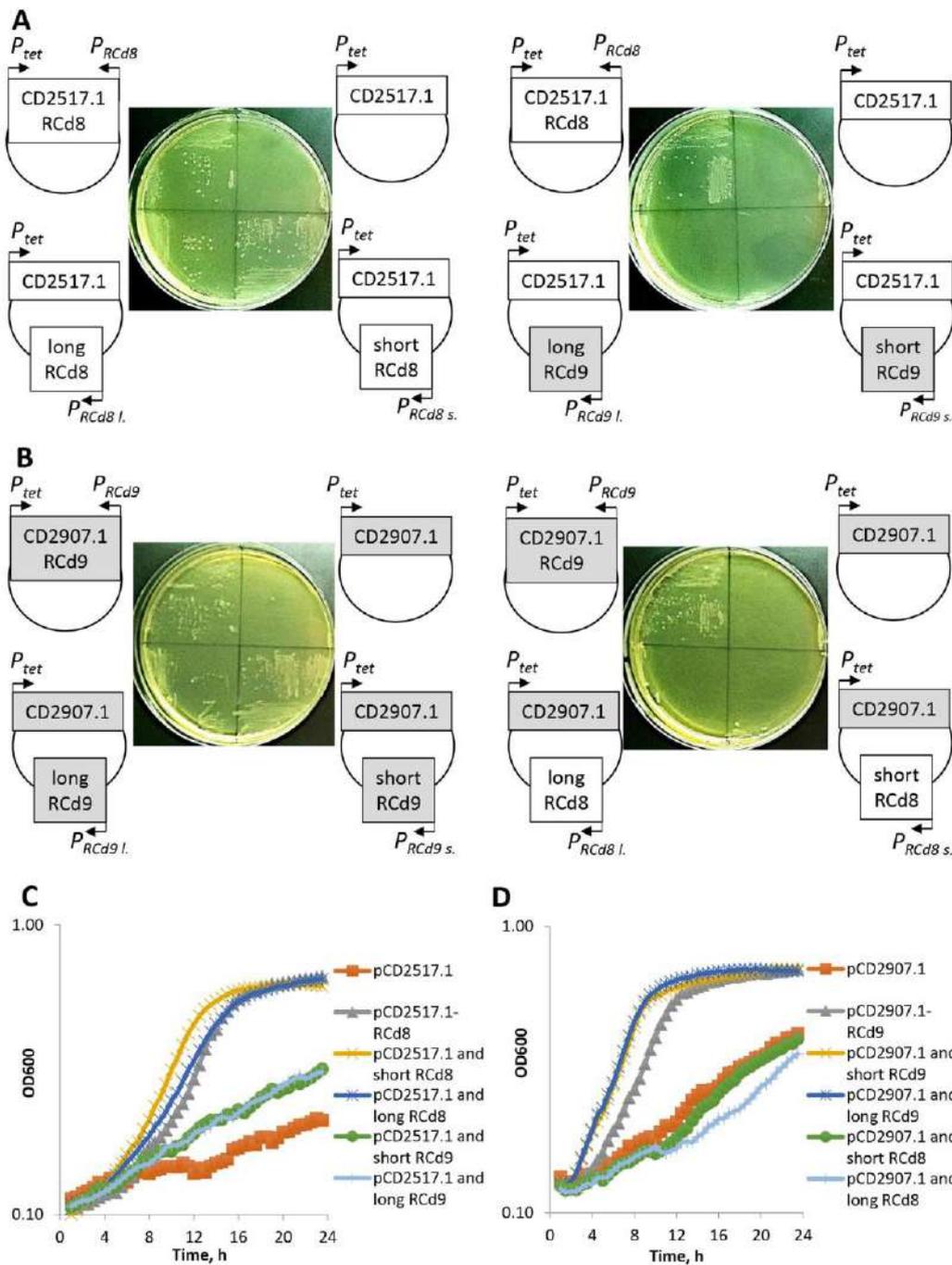


Figure 8. Effect of short and full-length antitoxin co-overexpression with native toxin *in trans* for CD2517.1-RCd8 and CD2907.1-RCd9/CD0956.2-RCd10 TA modules on growth in solid (A and B) and liquid (C and D) medium. (A) Growth of *C. difficile* strains CDIP1191 (630/pCD2517.1-RCd8), CDIP357 (630/pCD2517.1), CDIP1161 (630/pCD2517.1 and long RCd8), CDIP1130 (630/pCD2517.1 and short RCd8), CDIP1162 (630/pCD2517.1 and long RCd9) and CDIP1131 (630/pCD2517.1 and short RCd9) on BHI agar plates supplemented with Tm and 2.5 ng/ml of ATc inducer after 24 h of incubation at 37°C. (B) Growth of *C. difficile* strains CDIP1192 (630/pCD2907.1-RCd9), CDIP317 (630/pCD2907.1), CDIP1164 (630/pCD2907.1 and long RCd9), CDIP1133 (630/pCD2907.1 and short RCd9), CDIP1163 (630/pCD2907.1 and long RCd8) and CDIP1132 (630/pCD2907.1 and short RCd8) on BHI agar plates supplemented with Tm and 3 ng/ml of ATc inducer after 24 h of incubation at 37°C. (C) Growth of the same *C. difficile* strains as in (A) in TY medium at 37°C in the presence of 2.5 ng/ml ATc using a GloMax plate reader (Promega). The mean values are shown for three independent experiments. (D) Growth of the same *C. difficile* strains as in (B) in TY medium at 37°C in the presence of 3 ng/ml ATc using a GloMax plate reader (Promega). The mean values are shown for three independent experiments.

Downloaded from https://academic.oup.com/nar/article-abstract/46/9/4733/4909981 by guest on 21 September 2019

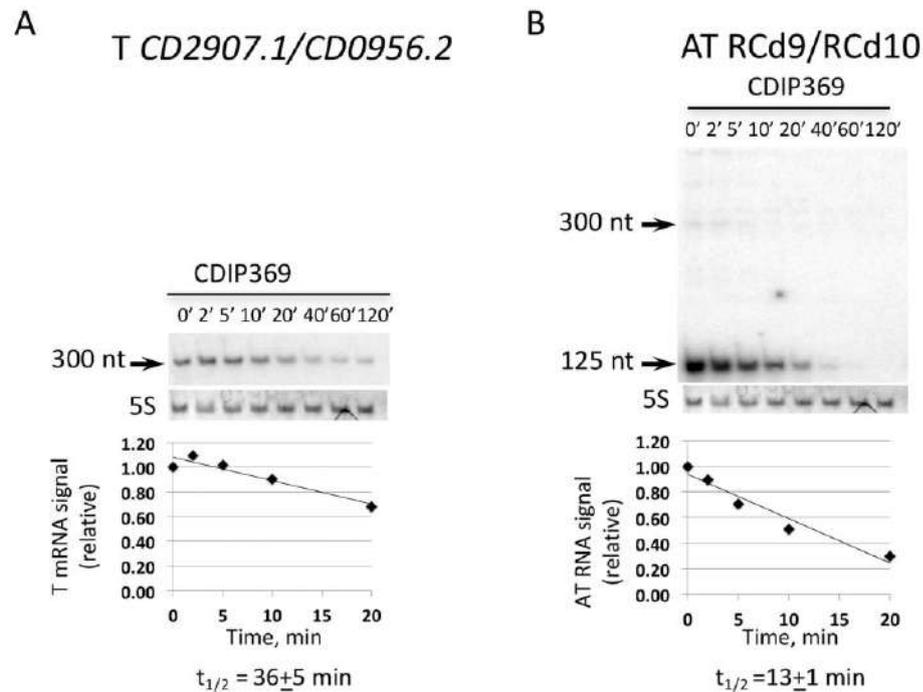


Figure 9. Expression and stability of CD2907.1/CD0956.2 toxin (A) and RCd9/RCd10 antitoxin (B) transcripts by northern blot. For determination of half-lives samples were taken at the indicated times after addition of 200 $\mu\text{g/ml}$ rifampicin. RNAs were extracted from strain CDIP369 (630/p). 5S RNA at the bottom of each northern blot autoradiogram serves as loading control. The same 5S control panel is shown when reprobing of the same membrane was performed. The relative intensities of the bands from northern blot analysis via autoradiography were quantified using ImageJ software.

sult suggests that at least RNase Y could in some way contribute to the antitoxin RNA degradation.

Expression analysis of TA and CRISPR-Cas systems

We wondered whether the chromosomal co-localization of CRISPR arrays and TA modules would imply the possible connection between these systems. As mentioned above, the alignment of CRISPR-associated TA module sequences strongly suggested the presence of both Sigma-A-dependent and Sigma-B-dependent promoters upstream of the TSS of the toxin and antitoxin genes for the six TA modules (Supplementary Figure S2). We have recently demonstrated the crucial role of the alternative Sigma B factor in the adaptive strategies of *C. difficile* inside the host (66). We then re-examined the transcriptome data for the *sigB* mutant as compared to the parental strain and observed up to 5-fold decrease in the expression of the entire gene sets for both the partial and complete *cas* operons (*CD2455* and *CD2982*) of type I-B *C. difficile* CRISPR-Cas system (Table 2). qRT-PCR analysis validated these transcriptome data (Table 2). In accordance, the search for Sigma-B-dependent promoter sequences revealed the presence of consensus elements GTTTTA-N12-GGGATTT and TTATAA-N12-GGGTTAA upstream of TSS for *cas* gene operons *CD2455* and *CD2982*, respectively. These promoter sequences are characterized by the presence of a conserved -10 promoter element associated and a less conserved -35 promoter element. Such a promoter structure suggests the possible im-

plication of other regulatory components controlling these operons together with the Sigma B factor. The high sequence conservation among direct repeats within multiple CRISPR arrays suggests that the same set of Cas proteins processes all expressed pre-crRNA in *C. difficile* strains (16). Thus, the induction of *cas* genes under stress conditions would allow the overall activation of CRISPR-Cas defense mechanisms. Transcriptome analysis of the *sigB* mutant also revealed differential expression of several newly identified TA genes and associated CRISPR arrays (Table 2). To confirm these data, we performed qRT-PCR analysis for selected TA gene pairs and CRISPR arrays (Table 2). In accordance with transcriptome data, we confirmed by qRT-PCR the downregulation of several CRISPR-associated TA genes in the *sigB* mutant strain as compared to the parental strain even without stress exposure (Table 2).

The induction of CRISPR-Cas-mediated defense capacities within biofilm community or more generally within the gut microbiota, which includes phages, could be important for bacterial survival under conditions promoting gene transfer. In *E. coli*, type I toxin *ralR* gene expression is induced during growth in biofilms (67). We thus compared the expression of *cas* genes and CRISPR-associated TA modules and CRISPR-Cas systems within biofilm and planktonic cultures and observed a strong, up to 20-fold, induction of expression of selected genes (Table 2). Overall these results suggest that the *cas* operons and the CRISPR arrays could be co-regulated with associated type I TA systems by stress- and biofilm-related factors.

Table 2. Differential expression of TA and CRISPR-Cas systems revealed by transcriptome and/or qRT-PCR analysis

Gene ID	Function	Ratio <i>sigB</i> /630Δ <i>erm</i> Microarray ^a	Ratio <i>sigB</i> /630Δ <i>erm</i> qRT-PCR	Ratio biofilm/plankton qRT-PCR
<i>CD2982</i> ^b	CRISPR-associated Cas6 family protein	0.19	0.22	14.7
<i>CD2981</i>	CRISPR-associated protein, CXXC-CXXC	0.21		
<i>CD2980</i>	CRISPR-associated autoregulator DevR family protein	0.26		
<i>CD2979</i>	CRISPR-associated Cas5 family protein	0.25		
<i>CD2978</i>	CRISPR-associated Cas3 family helicase	0.30		
<i>CD2977</i>	CRISPR-associated Cas4 family protein	0.33		
<i>CD2976</i>	CRISPR-associated Cas1 family protein	0.37		
<i>CD2975</i>	CRISPR-associated Cas2 family protein	0.37		
<i>CD2455</i> ^c	CRISPR-associated protein	0.55	0.61	9.4
<i>CD2454</i>	Conserved hypothetical protein	0.55		
<i>CD2453</i>	CRISPR-associated negative autoregulator	0.47		
<i>CD2452</i>	CRISPR-associated protein	0.53		
<i>CD1233.1</i>	Toxin of TA associated with CRISPR 6	0.51		
<i>CD2517.1</i>	Toxin of TA associated with CRISPR 12	0.25	0.42	26.0
RCd8	Antitoxin of TA associated with CRISPR 12		0.67	7.3
CD630_n00860	CRISPR 12			7.7
<i>CD2907.1</i>	Toxin of TA associated with CRISPR 16/15			1.8
RCd9/RCd10	Antitoxin of TA associated with CRISPR 16/15 /CRISPR 3/4	0.5	0.42	2.3
CD630_n00990	CRISPR 16/15		0.54	9.8

Gene names and functions correspond to those indicated in the MaGe database Clostriscope (<https://www.genoscope.ens.fr>).

^aA gene was considered as differentially expressed between the strain 630Δ*erm* and the *sigB* mutant when the *P*-value is < 0.05.

^bFirst gene of the complete *cas* operon *CD2982-CD2975*.

^cFirst gene of the partial *cas* operon *CD2455-CD2452*. SQ1781 corresponds to RCd8, CD630_n01000 to RCd9 and CD630_n00370 to RCd10.

Genomic analysis of TA and CRISPR arrays co-localization

We analyzed the extent of co-localization of potential type I TA with CRISPR arrays in available *C. difficile* sequences. From more than 2,500 *C. difficile* genome sequences assembled and automatically annotated, we first found that 98% contain CRISPR arrays (from 1 to 30). In these CRISPR-containing strains, we then searched for the presence, immediately adjacent to CRISPR loci, of open reading frames from 40 to 60 amino acids, as one of the characteristic features of type I toxins is their small size. This search resulted in about 7000 hits. The CRISPR-associated small proteins were only absent in 67 genomes of which 58 lacked *cas* gene homologs. Then, an orthology analysis identified 16 proteins present each in more than 25 strains (Supplementary Table S4). Figure 10 shows an alignment of these 16 representative small proteins adjacent to CRISPR arrays combined in five major groups (A–E) according to their homology.

The three small proteins characterized in this study (*CD2907.1*, *CD0956.2* and *CD2517.1*) belong to group A. This group is largely distributed in *C. difficile* as it is present in two-third of the analyzed strains (Figure 10). CRISPR 16/15 and CRISPR 3/4-associated toxins belong to the most represented subgroup, A1, found in 63% of strains, CRISPR 12 associated toxin belongs to subgroup A2, that is present in 20% of the analyzed strains. Other CRISPR-associated toxins of strain 630 are represented in less extent within the same group. Finally, as two of the characterized toxins are located within prophage regions in strain 630, we wondered whether prophage localisation could be a common feature of CRISPR-associated small proteins. In 13 from 22 known *C. difficile* phages, we found potential toxins all belonging to the group A that could be part of TA mod-

ules. However, the co-localization with CRISPR arrays is detected only in the phi027 prophage of the R20291 strain.

To provide an experimental confirmation of potential TA and CRISPR arrays co-expression in another *C. difficile* strain, we have looked at the RNA-seq data of the epidemic strain R20291 (68). In this strain, we detected three co-localized CRISPR and TA pairs (Supplementary Figure S8A). One pair was intact (TA and CRISPR), while the two others have mutation in the toxin genes and only antitoxin was detected (A and CRISPR). We confirmed their co-expression in the published R20291 RNA-seq data using the COV2HTML software for visualization (69) (Supplementary Figure S8B).

In summary, we found that (i) CRISPR-associated small proteins are present in the vast majority of *C. difficile* strains and (ii) their primary orthology group is homologous to newly identified type I TA toxins.

DISCUSSION

Here, we report the first identification of functional type I TA modules in *C. difficile* 630 chromosome. Deep-sequencing, northern blotting and 5'/3'RACE revealed the presence of overlapping transcripts for type I toxin gene and associated RNA antitoxin in several chromosomal loci. Comparison of the newly identified type I TA systems in *C. difficile* with previously studied TA systems in other bacteria revealed no sequence homology for small toxin proteins. However, we observed a conservation of their membrane association and the presence of charged amino acids in the C-terminal part (32,35,70). The inducible overexpression of toxin genes strongly impaired the *C. difficile* growth while co-expression of associated antitoxin RNA *in cis* or

majority of sequenced *C. difficile* strains. Initially, TA systems were shown to be important for maintenance of plasmids through a post-segregation killing mechanism (20,21,84). The role of numerous chromosomal TA systems remains largely enigmatic, even though their possible implication in stabilization of chromosomal regions has been emphasized. For example, a TA module has been shown to promote the maintenance of an integrative conjugative element STX in *V. cholerae* (85).

The co-localization of functional type I TA systems with CRISPR arrays that we observed on *C. difficile* chromosome has never been reported for any other bacterial genome. Nevertheless, several type I TA systems are located within prophage or prophage-like regions both in *C. difficile* and *B. subtilis* (29,57,59,64,82,86), even though *B. subtilis* genome lacks CRISPR arrays (2). The type I TA modules are present within the *skin* element, which is excised from the chromosome during sporulation, in *B. subtilis* and *C. difficile*. As for *B. subtilis* systems, a role in stabilisation of these chromosomal regions can be hypothesized for TA systems in *C. difficile*, which carries a high proportion of stable mobile genetic elements in its genome (87).

Based on the observations that prophage-located CRISPR arrays are often associated with type I TA modules in *C. difficile*, an interesting evolutionary aspect of the *C. difficile* CRISPR-Cas system can be underlined. Indeed, the TA systems could contribute to the stabilization of the chromosomal regions carrying CRISPR-Cas systems after acquisition of large defense capacities associated with CRISPR arrays. We can hypothesize that TA modules are implicated in maintaining of CRISPR regions, but also in stress response, prophage stability, sporulation control, biofilm formation and other community-associated processes important for this pathogen.

Possible connections between CRISPR and TA systems were highlighted by several recent studies focusing on type II TA (88). Bioinformatics search identified the so-called 'defense islands' in bacteria associating immunity and cell death or dormancy functions including CRISPR and type II TA systems (89,90). The original features of *C. difficile* are that type I toxins were not found in 'defense islands'. The role of this functional coupling 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 invaders. Thus, our findings are in line with recently emerged concept on a functional coupling between distinct defense strategies provided by immunity and cell dormancy systems in prokaryotes (88).

The co-regulation of CRISPR-Cas and newly described type I TA systems by the stress-specific factor, Sigma B and the biofilm-related stimuli further suggests the possible connections between these systems in *C. difficile*. Our findings emphasize additional original features of the recently characterized *C. difficile* CRISPR-Cas system including the link with community-behavior control, stress response and type I TA systems. Such control of CRISPR-Cas expression in response to stress-related factors could be relevant for the *C. difficile* infection cycle.

Together with alternative roles of CRISPR-Cas in the control of bacterial physiology and pathogenesis beyond the role in defense against foreign invaders (91,92), stimuli and mechanisms controlling CRISPR-Cas system expression just start to be uncovered. However, multiple connections between TA systems in bacteria and stress response have been reported (22,25). We provide here new data on the co-regulation of type I TA and CRISPR-Cas systems by the general stress response Sigma B factor in *C. difficile*. Sigma B likely plays a crucial role in the responses to stresses encountered by this pathogen inside the host. Interestingly, the MazEF type II TA module is encoded within the *sigB* operon in *S. aureus* with possible regulatory connections (93). Various environmental stimuli including metabolic and genotoxic stresses induce TA gene expression of type I TA systems in *B. subtilis*, *E. coli* and *S. aureus* (29,70,86,94–96). In a multi-stress responsive type I TA system *bsrE*/SR5 from *B. subtilis*, the control of antitoxin RNA SR5 by iron limitation stress has been reported to be dependent on the alternative Sigma B factor (82).

Key roles of both type II and type I TA systems have been suggested in bacterial pathogens where they can contribute to virulence, fitness inside the host, persistence, intracellular lifestyle, stress response and biofilm formation (26,27,97,98). More generally, biofilm formation process has been associated in previous studies with bacterial TA systems (26). Recent data suggest that the TxpA type I toxin from the *skin* element acts to eliminate defective cells and preserve symmetry in *B. subtilis* biofilms (99). We show here that both the expression of the CRISPR-Cas and the associated TA systems are induced in biofilm conditions in *C. difficile*. In general, TA systems including well-documented type II TA exist in surprisingly high numbers in all prokaryotes but clostridial TA modules have been only poorly characterized so far. Before this study, no data were available on TA modules in *C. difficile* with the exception of the recently identified MazEF, a type II TA system member (100). Possible implications of type II TA modules in recurrent *C. difficile* infection, sporulation and biofilm formation were recently discussed (101). Among the most challenging aspects of *C. difficile*-associated disease remain the high incidence of recurrent infections and the ability of transition from inert colonization to active infection (102,103). A comparative genomic study showed that the genomes of most dangerous epidemic bacteria are characterized by the accumulation of TA modules (97). Promising perspectives for the applications of TA and CRISPR as a basis for the development of new anti-bacterial strategies could be examined in the future (27,104).

In conclusion, this study provides the first characterization of type I TA modules in the emergent enteropathogen *C. difficile*. Intriguingly, these chromosomal TA pairs are co-localized with CRISPR array components of bacterial adaptive immunity defense system CRISPR-Cas in the majority of sequenced *C. difficile* strains. Further investigations will help to precise the biological functions of these widespread chromosomal TA loci for *C. difficile* physiology and its successful development inside the host, to uncover the molecular mechanisms involved in their regulation and the possible crosstalk between homologous systems, as well as to evaluate their potential for future thera-

peutic and biotechnological applications in pathogenic bacteria.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to J.-M. Ghigo for microfermentor biofilm experiment and L.-C. Fortier for helpful discussions. We thank H. Putzer and C. Condon for anti-RNase J, anti-RNase Y and anti-RNase III antibodies. A.M. research was performed in fulfilment of educational requirements of a dual degree Skoltech-University Paris-Diderot Ph.D. program supported by the Vernadski Fellowship from the French Embassy in Russia and is also partly supported by Skoltech Biomedical Initiative Program grant (SBI RF-0000000136) to Konstantin Severinov.

FUNDING

Institut Pasteur; University Paris Diderot; Agence Nationale de la Recherche [‘CloSTARn’, ANR-13-JSV3-0005-01 to O.S.]; Institut Universitaire de France (IUF to O.S.); Pasteur-Weizmann Council; University Paris-Sud; Institute for Integrative Biology of the Cell; Vernadski Fellowship (to A.M.); Skoltech Biomedical Initiative [SBI RF-0000000136]; Centre National de la Recherche Scientifique [UMR8261]; ‘Initiative d’Excellence’ Program from the French State [Grant ‘DYNAMO’, ANR-11-LABX-0011 to E.H.]. Funding for open access charge: IUF, Skoltech. *Conflict of interest statement.* None declared.

REFERENCES

- Marraffini, L.A. (2015) CRISPR-Cas immunity in prokaryotes. *Nature*, **526**, 55–61.
- Grissa, I., Vergnaud, G. and Pourcel, C. (2007) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics*, **8**, 172.
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A. and Horvath, P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, **315**, 1709–1712.
- Bhaya, D., Davison, M. and Barrangou, R. (2011) CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu. Rev. Genet.*, **45**, 273–297.
- Amitai, G. and Sorek, R. (2016) CRISPR-Cas adaptation: insights into the mechanism of action. *Nat. Rev. Microbiol.*, **14**, 67–76.
- Lopez-Sanchez, M.J., Sauvage, E., Da Cunha, V., Clermont, D., Ratsima Hariniaina, E., Gonzalez-Zorn, B., Poyart, C., Rosinski-Chupin, I. and Glaser, P. (2012) The highly dynamic CRISPR1 system of *Streptococcus agalactiae* controls the diversity of its mobilome. *Mol. Microbiol.*, **85**, 1057–1071.
- Mick, E., Stern, A. and Sorek, R. (2013) Holding a grudge: persisting anti-phage CRISPR immunity in multiple human gut microbiomes. *RNA Biol.*, **10**, 900–906.
- Stern, A., Mick, E., Tirosh, I., Sagy, O. and Sorek, R. (2012) CRISPR targeting reveals a reservoir of common phages associated with the human gut microbiome. *Genome Res.*, **22**, 1985–1994.
- Carroll, K.C. and Bartlett, J.G. (2011) Biology of *Clostridium difficile*: implications for epidemiology and diagnosis. *Annu. Rev. Microbiol.*, **65**, 501–521.
- Rupnik, M., Wilcox, M.H. and Gerding, D.N. (2009) *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. *Nat. Rev. Microbiol.*, **7**, 526–536.
- Seekatz, A.M. and Young, V.B. (2014) *Clostridium difficile* and the microbiota. *J. Clin. Investig.*, **124**, 4182–4189.
- Just, I., Selzer, J., Wilm, M., von Eichel-Streiber, C., Mann, M. and Aktories, K. (1995) Glucosylation of Rho proteins by *Clostridium difficile* toxin B. *Nature*, **375**, 500–503.
- Vedantam, G., Clark, A., Chu, M., McQuade, R., Mallozzi, M. and Viswanathan, V. (2012) *Clostridium difficile* infection: toxins and non-toxin virulence factors, and their contributions to disease establishment and host response. *Gut Microbes*, **3**, 121–134.
- Wagner, E.G.H. and Romby, P. (2015) Chapter three—small RNAs in bacteria and archaea: who they are, what they do, and how they do it. *Adv. Genet.*, **90**, 133–208.
- Soutourina, O.A., Monot, M., Boudry, P., Saujet, L., Pichon, C., Sismeyro, O., Semenova, E., Severinov, K., Le Bouguennec, C. and Coppée, J.-Y. (2013) Genome-wide identification of regulatory RNAs in the human pathogen *Clostridium difficile*. *PLoS Genet.*, **9**, e1003493.
- Boudry, P., Semenova, E., Monot, M., Datsenko, K.A., Lopatina, A., Sekulovic, O., Ospina-Bedoya, M., Fortier, L.C., Severinov, K., Dupuy, B. et al. (2015) Function of the CRISPR-Cas system of the human pathogen *Clostridium difficile*. *Mbio*, **6**, e01112–e01115.
- Andersen, J.M., Shoup, M., Robinson, C., Britton, R., Olsen, K.E. and Barrangou, R. (2016) CRISPR diversity and microevolution in *Clostridium difficile*. *Genome Biol. Evol.*, **8**, 2841–2855.
- Hargreaves, K.R., Flores, C.O., Lawley, T.D. and Clokie, M.R. (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**, e01045-01013.
- Sorek, R., Kunin, V. and Hugenholtz, P. (2008) CRISPR—a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat. Rev. Microbiol.*, **6**, 181–186.
- Page, R. and Peti, W. (2016) Toxin-antitoxin systems in bacterial growth arrest and persistence. *Nat. Chem. Biol.*, **12**, 208–214.
- Hayes, F. (2003) Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest. *Science*, **301**, 1496–1499.
- Gerdes, K., Christensen, S.K. and Lobner-Olesen, A. (2005) Prokaryotic toxin-antitoxin stress response loci. *Nat. Rev. Microbiol.*, **3**, 371–382.
- Gerdes, K. and Maisonneuve, E. (2012) Bacterial persistence and toxin-antitoxin loci. *Annu. Rev. Microbiol.*, **66**, 103–123.
- Maisonneuve, E., Castro-Camargo, M. and Gerdes, K. (2013) (p)ppGpp controls bacterial persistence by stochastic induction of toxin-antitoxin activity. *Cell*, **154**, 1140–1150.
- Wang, X. and Wood, T.K. (2011) Toxin-antitoxin systems influence biofilm and persist cell formation and the general stress response. *Appl. Environ. Microbiol.*, **77**, 5577–5583.
- Wen, Y., Behiels, E. and Devreese, B. (2014) Toxin-antitoxin systems: their role in persistence, biofilm formation, and pathogenicity. *Pathog. Dis.*, **70**, 240–249.
- Yamaguchi, Y. and Inouye, M. (2011) Regulation of growth and death in *Escherichia coli* by toxin-antitoxin systems. *Nat. Rev. Microbiol.*, **9**, 779–790.
- Brantl, S. (2012) Bacterial type I toxin-antitoxin systems. *RNA Biol.*, **9**, 1488–1490.
- Brantl, S. and Jahn, N. (2015) sRNAs in bacterial type I and type III toxin-antitoxin systems. *FEMS Microbiol. Rev.*, **39**, 413–427.
- Coray, D.S., Wheeler, N.E., Heinemann, J.A. and Gardner, P.P. (2017) Why so narrow: distribution of anti-sense regulated, type I toxin-antitoxin systems compared with type II and type III systems. *RNA Biol.*, **14**, 275–280.
- Nogueira, T., Rankin, D.J., Touchon, M., Taddei, F., Brown, S.P. and Rocha, E.P. (2009) Horizontal gene transfer of the secretome drives the evolution of bacterial cooperation and virulence. *Curr. Biol.*, **19**, 1683–1691.
- Fozo, E.M., Makarova, K.S., Shabalina, S.A., Yutin, N., Koonin, E.V. and Storz, G. (2010) Abundance of type I toxin-antitoxin systems in bacteria: searches for new candidates and discovery of novel families. *Nucleic Acids Res.*, **38**, 3743–3759.
- Brielle, R., Pinel-Marie, M.L. and Felden, B. (2016) Linking bacterial type I toxins with their actions. *Curr. Opin. Microbiol.*, **30**, 114–121.
- Guo, Y., Quiroga, C., Chen, Q., McAnulty, M.J., Benedik, M.J., Wood, T.K. and Wang, X. (2014) RalR (a DNase) and RalA (a small

- RNA) form a type I toxin-antitoxin system in *Escherichia coli*. *Nucleic Acids Res.*, **42**, 6448–6462.
35. Jahn, N., Brantl, S. and Strahl, H. (2015) Against the mainstream: the membrane-associated type I toxin BsrG from *Bacillus subtilis* interferes with cell envelope biosynthesis without increasing membrane permeability. *Mol. Microbiol.*, **98**, 651–666.
 36. Dupuy, B. and Sonenshein, A.L. (1998) Regulated transcription of *Clostridium difficile* toxin genes. *Mol. Microbiol.*, **27**, 107–120.
 37. Bertani, G. (1951) Studies on lysogeny. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J. Bacteriol.*, **62**, 293–300.
 38. Fagan, R.P. and Fairweather, N.F. (2011) *Clostridium difficile* has two parallel and essential Sec secretion systems. *J. Biol. Chem.*, **286**, 27483–27493.
 39. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual, Second Edition*. Cold Spring Harbor Laboratory, NY.
 40. O'Connor, J.R., Lyras, D., Farrow, K.A., Adams, V., Powell, D.R., Hinds, J., Cheung, J.K. and Rood, J.I. (2006) Construction and analysis of chromosomal *Clostridium difficile* mutants. *Mol. Microbiol.*, **61**, 1335–1351.
 41. Collins, T.J. (2007) ImageJ for microscopy. *Biotechniques*, **43**, 25–30.
 42. Andre, G., Even, S., Putzer, H., Burguiere, P., Croux, C., Danchin, A., Martin-Verstraete, I. and 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.
 43. Ghigo, J.M. (2001) Natural conjugative plasmids induce bacterial biofilm development. *Nature*, **412**, 442–445.
 44. Saujet, L., Monot, M., Dupuy, B., Soutourina, O. and 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.
 45. Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻(Delta Delta C(T)) Method. *Methods*, **25**, 402–408.
 46. Folichon, M., Allemand, F., Regnier, P. and Hajnsdorf, E. (2005) Stimulation of poly(A) synthesis by *Escherichia coli* poly(A) polymerase I is correlated with Hfq binding to poly(A) tails. *FEBS J.*, **272**, 454–463.
 47. Hajnsdorf, E. and Regnier, P. (2000) Host factor Hfq of *Escherichia coli* stimulates elongation of poly(A) tails by poly(A) polymerase I. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 1501–1505.
 48. Peltier, J., Shaw, H.A., Couchman, E.C., Dawson, L.F., Yu, L., Choudhary, J.S., Kaever, V., Wren, B.W. and Fairweather, N.F. (2015) Cyclic diGMP regulates production of sortase substrates of *Clostridium difficile* and their surface exposure through ZmpI protease-mediated cleavage. *J. Biol. Chem.*, **290**, 24453–24469.
 49. Boudry, P., Gracia, C., Monot, M., Caillet, J., Saujet, L., Hajnsdorf, E., Dupuy, B., Martin-Verstraete, I. and Soutourina, O. (2014) Pleiotropic role of the RNA chaperone protein Hfq in the human pathogen *Clostridium difficile*. *J. Bacteriol.*, **196**, 3234–3248.
 50. Cairns, M.D., Preston, M.D., Hall, C.L., Gerding, D.N., Hawkey, P.M., Kato, H., Kim, H., Kuijper, E.J., Lawley, T.D., Pituch, H. et al. (2017) Comparative genome analysis and global phylogeny of the toxin variant *Clostridium difficile* PCR ribotype 017 reveals the evolution of two independent sublineages. *J. Clin. Microbiol.*, **55**, 865–876.
 51. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D. et al. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.*, **19**, 455–477.
 52. Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, **30**, 2068–2069.
 53. Lechner, M., Findeiss, S., Steiner, L., Marz, M., Stadler, P.F. and Prohaska, S.J. (2011) Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics*, **12**, 124.
 54. Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R. et al. (2007) Clustal W and clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.
 55. Pinel-Marie, M.L., Brielle, R. and Felden, B. (2014) Dual toxic-peptide-coding *Staphylococcus aureus* RNA under antisense regulation targets host cells and bacterial rivals unequally. *Cell Rep.*, **7**, 424–435.
 56. Sayed, N., Jouselin, A. and Felden, B. (2011) A *cis*-antisense RNA acts in trans in *Staphylococcus aureus* to control translation of a human cytolytic peptide. *Nat. Struct. Mol. Biol.*, **19**, 105–112.
 57. Silvaggi, J.M., Perkins, J.B. and Losick, R. (2005) Small untranslated RNA antitoxin in *Bacillus subtilis*. *J. Bacteriol.*, **187**, 6641–6650.
 58. Serrano, M., Kint, N., Pereira, F.C., Saujet, L., Boudry, P., Dupuy, B., Henriques, A.O. and Martin-Verstraete, I. (2016) A recombination directionality factor controls the cell type-specific activation of sigmaK and the fidelity of spore development in *Clostridium difficile*. *PLoS Genet.*, **12**, e1006312.
 59. Durand, S., Jahn, N., Condon, C. and Brantl, S. (2012) Type I toxin-antitoxin systems in *Bacillus subtilis*. *RNA Biol.*, **9**, 1491–1497.
 60. Jahn, N. and Brantl, S. (2013) One antitoxin—two functions: SR4 controls toxin mRNA decay and translation. *Nucleic Acids Res.*, **41**, 9870–9880.
 61. Wen, J. and Fozo, E.M. (2014) sRNA antitoxins: more than one way to repress a toxin. *Toxins*, **6**, 2310–2335.
 62. Wagner, E.G., Blomberg, P. and Nordstrom, K. (1992) Replication control in plasmid R1: duplex formation between the antisense RNA, CopA, and its target, CopT, is not required for inhibition of RepA synthesis. *EMBO J.*, **11**, 1195–1203.
 63. Patel, S. and Weaver, K.E. (2006) Addiction toxin Fst has unique effects on chromosome segregation and cell division in *Enterococcus faecalis* and *Bacillus subtilis*. *J. Bacteriol.*, **188**, 5374–5384.
 64. Durand, S., Gilet, L. and Condon, C. (2012) The essential function of *B. subtilis* RNase III is to silence foreign toxin genes. *PLoS Genet.*, **8**, e1003181.
 65. Dembek, M., Barquist, L., Boinett, C.J., Cain, A.K., Mayho, M., Lawley, T.D., Fairweather, N.F. and Fagan, R.P. (2015) High-throughput analysis of gene essentiality and sporulation in *Clostridium difficile*. *Mbio*, **6**, e02383.
 66. Kint, N., Janoir, C., Monot, M., Hoys, S., Soutourina, O., Dupuy, B. and Martin-Verstraete, I. (2017) The alternative sigma factor sigmaB and a crucial role in adaptive strategies of *Clostridium difficile* during gut infection. *Environ. Microbiol.*, **19**, 1933–1958.
 67. Domka, J., Lee, J., Bansal, T. and Wood, T.K. (2007) Temporal gene-expression in *Escherichia coli* K-12 biofilms. *Environ. Microbiol.*, **9**, 332–346.
 68. Maldarelli, G.A., Piepenbrink, K.H., Scott, A.J., Freiberg, J.A., Song, Y., Achermann, Y., Ernst, R.K., Shirliff, M.E., Sundberg, E.J., Donnenberg, M.S. et al. (2016) Type IV pili promote early biofilm formation by *Clostridium difficile*. *Pathog. Dis.*, **74**, doi:10.1093/femsdp/ftw061.
 69. Monot, M., Orgeur, M., Camiade, E., Brehier, C. and Dupuy, B. (2014) COV2HTML: a visualization and analysis tool of bacterial next generation sequencing (NGS) data for postgenomics life scientists. *OMICS*, **18**, 184–195.
 70. Sayed, N., Nonin-Lecomte, S., Rety, S. and Felden, B. (2012) Functional and structural insights of a *Staphylococcus aureus* apoptotic-like membrane peptide from a toxin-antitoxin module. *J. Biol. Chem.*, **287**, 43454–43463.
 71. Fozo, E.M. (2012) New type I toxin-antitoxin families from 'wild' and laboratory strains of *E. coli*: Ibs-Sib, ShoB-OhsC and Zor-Orz. *RNA Biol.*, **9**, 1504–1512.
 72. Fozo, E.M., Kawano, M., Fontaine, F., Kaya, Y., Mendieta, K.S., Jones, K.L., Ocampo, A., Rudd, K.E. and Storz, G. (2008) Repression of small toxic protein synthesis by the Sib and OhsC small RNAs. *Mol. Microbiol.*, **70**, 1076–1093.
 73. Han, K., Kim, K.S., Bak, G., Park, H. and Lee, Y. (2010) Recognition and discrimination of target mRNAs by Sib RNAs, a *cis*-encoded sRNA family. *Nucleic Acids Res.*, **38**, 5851–5866.
 74. Wen, J., Won, D. and Fozo, E.M. (2014) The ZorO-OrzO type I toxin-antitoxin locus: repression by the OrzO antitoxin. *Nucleic Acids Res.*, **42**, 1930–1946.
 75. Iqbal, N., Guerout, A.M., Krin, E., Le Roux, F. and Mazel, D. (2015) Comprehensive functional analysis of the 18 *Vibrio cholerae* N16961 toxin-antitoxin systems substantiates their role in stabilizing the superintegron. *J. Bacteriol.*, **197**, 2150–2159.
 76. Arnion, H., Korkut, D.N., Masachs Gelo, S., Chabas, S., Reignier, J., Iost, I. and Darfeuille, F. (2017) Mechanistic insights into type I toxin antitoxin systems in *Helicobacter pylori*: the importance of mRNA folding in controlling toxin expression. *Nucleic Acids Res.*, **45**, 4782–4795.

77. Meissner, C., Jahn, N. and Brantl, S. (2016) In vitro characterization of the type I toxin-antitoxin system *bsrE*/SR5 from *Bacillus subtilis*. *J. Biol. Chem.*, **291**, 560–571.
78. Wen, J., Harp, J.R. and Fozo, E.M. (2017) The 5' UTR of the type I toxin ZorO can both inhibit and enhance translation. *Nucleic Acids Res.*, **45**, 4006–4020.
79. Berghoff, B.A., Hoekzema, M., Aulbach, L. and Wagner, E.G. (2017) Two regulatory RNA elements affect TisB-dependent depolarization and persister formation. *Mol. Microbiol.*, **103**, 1020–1033.
80. Darfeuille, F., Unoson, C., Vogel, J. and Wagner, E.G. (2007) An antisense RNA inhibits translation by competing with standby ribosomes. *Mol. Cell*, **26**, 381–392.
81. Gerdes, K., Gulyaev, A.P., Franch, T., Pedersen, K. and Mikkelsen, N.D. (1997) Antisense RNA-regulated programmed cell death. *Annu. Rev. Genet.*, **31**, 1–31.
82. Muller, P., Jahn, N., Ring, C., Maiwald, C., Neubert, R., Meissner, C. and Brantl, S. (2016) A multistress responsive type I toxin-antitoxin system: *bsrE*/SR5 from the *B. subtilis* chromosome. *RNA Biol.*, **13**, 511–523.
83. Dambach, M., Irnov, I. and Winkler, W.C. (2013) Association of RNAs with *Bacillus subtilis* Hfq. *PLoS One*, **8**, e51156.
84. Shen, Z., Patil, R.D., Sahin, O., Wu, Z., Pu, X.Y., Dai, L., Plummer, P.J., Yaeger, M.J. and Zhang, Q. (2016) Identification and functional analysis of two toxin-antitoxin systems in *Campylobacter jejuni*. *Mol. Microbiol.*, **101**, 909–923.
85. Wozniak, R.A., Fouts, D.E., Spagnoletti, M., Colombo, M.M., Ceccarelli, D., Garriss, G., Dery, C., Burrus, V. and Waldor, M.K. (2009) Comparative ICE genomics: insights into the evolution of the SXT/R391 family of ICEs. *PLoS Genet.*, **5**, e1000786.
86. Jahn, N., Preis, H., Wiedemann, C. and Brantl, S. (2012) BsrG/SR4 from *Bacillus subtilis*—the first temperature-dependent type I toxin-antitoxin system. *Mol. Microbiol.*, **83**, 579–598.
87. Sebahia, M., Wren, B.W., Mullany, P., Fairweather, N.F., Minton, N., Stabler, R., Thomson, N.R., Roberts, A.P., Cerdeno-Tarraga, A.M., Wang, H. et al. (2006) The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nat. Genet.*, **38**, 779–786.
88. Koonin, E.V. and Zhang, F. (2017) Coupling immunity and programmed cell suicide in prokaryotes: Life-or-death choices. *Bioessays*, **39**, 1–9.
89. Makarova, K.S., Wolf, Y.I. and Koonin, E.V. (2013) Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Res.*, **41**, 4360–4377.
90. Makarova, K.S., Wolf, Y.I., Snir, S. and Koonin, E.V. (2011) Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *J. Bacteriol.*, **193**, 6039–6056.
91. Richter, C., Chang, J.T. and Fineran, P.C. (2012) Function and regulation of clustered regularly interspaced short palindromic repeats (CRISPR) / CRISPR associated (Cas) systems. *Viruses*, **4**, 2291–2311.
92. Sampson, T.R. and Weiss, D.S. (2013) Alternative roles for CRISPR/Cas systems in bacterial pathogenesis. *PLoS Pathog.*, **9**, e1003621.
93. Donegan, N.P. and Cheung, A.L. (2009) Regulation of the mazEF toxin-antitoxin module in *Staphylococcus aureus* and its impact on *sigB* expression. *J. Bacteriol.*, **191**, 2795–2805.
94. Jahn, N. and Brantl, S. (2016) Heat-shock-induced refolding entails rapid degradation of *bsrG* toxin mRNA by RNases Y and J1. *Microbiology*, **162**, 590–599.
95. Kawano, M. (2012) Divergently overlapping cis-encoded antisense RNA regulating toxin-antitoxin systems from *E. coli*: *hok/sok*, *ldr/rdl*, *symE/symR*. *RNA Biol.*, **9**, 1520–1527.
96. Kawano, M., Aravind, L. and Storz, G. (2007) An antisense RNA controls synthesis of an SOS-induced toxin evolved from an antitoxin. *Mol. Microbiol.*, **64**, 738–754.
97. Georgiades, K. and Raouf, D. (2011) Genomes of the most dangerous epidemic bacteria have a virulence repertoire characterized by fewer genes but more toxin-antitoxin modules. *PLoS One*, **6**, e17962.
98. Lobato-Marquez, D., Moreno-Cordoba, I., Figueroa, V., Diaz-Orejas, R. and Garcia-del Portillo, F. (2015) Distinct type I and type II toxin-antitoxin modules control *Salmonella* lifestyle inside eukaryotic cells. *Sci. Rep.*, **5**, 9374.
99. Bloom-Ackermann, Z., Steinberg, N., Rosenberg, G., Oppenheimer-Shaanan, Y., Pollack, D., Ely, S., Storzi, N., Levy, A. and Kolodkin-Gal, I. (2016) Toxin-antitoxin systems eliminate defective cells and preserve symmetry in *Bacillus subtilis* biofilms. *Environ. Microbiol.*, **18**, 5032–5047.
100. Rothenbacher, F.P., Suzuki, M., Hurley, J.M., Montville, T.J., Kirn, T.J., Ouyang, M. and Woychik, N.A. (2012) *Clostridium difficile* MazF toxin exhibits selective, not global, mRNA cleavage. *J. Bacteriol.*, **194**, 3464–3474.
101. Gil, F., Pizarro-Guajardo, M., Alvarez, R., Garavaglia, M. and Paredes-Sabja, D. (2015) *Clostridium difficile* recurrent infection: possible implication of TA systems. *Future Microbiol.*, **10**, 1649–1657.
102. Shields, K., Araujo-Castillo, R.V., Theethira, T.G., Alonso, C.D. and Kelly, C.P. (2015) Recurrent *Clostridium difficile* infection: From colonization to cure. *Anaerobe*, **34**, 59–73.
103. Smits, W.K., Lyras, D., Lacy, D.B., Wilcox, M.H. and Kuijper, E.J. (2016) *Clostridium difficile* infection. *Nat. Rev. Dis. Primers*, **2**, 16020.
104. Lee, K.Y. and Lee, B.J. (2016) Structure, biology, and therapeutic application of toxin-antitoxin systems in pathogenic bacteria. *Toxins*, **8**, 305.

Annex. Article 2



New Insights Into Functions and Possible Applications of *Clostridium difficile* CRISPR-Cas System

Anna Maikova^{1,2,3,4}, Konstantin Severinov^{1,4,5} and Olga Soutourina^{3,6*}

¹ Center for Life Sciences, Skolkovo Institute of Science and Technology, Moscow, Russia, ² Université Paris Diderot, Sorbonne Paris Cité, Paris, France, ³ Microbiology, Institute for Integrative Biology of the Cell, Commissariat à l'Energie Atomique et aux Energies Alternatives, Centre National de la Recherche Scientifique, Université Paris-Sud, Université Paris-Saclay, Gif-sur-Yvette, France, ⁴ Peter the Great St. Petersburg Polytechnic University, Saint Petersburg, Russia, ⁵ Waksman Institute for Microbiology, Rutgers, The State University of New Jersey, Piscataway, NJ, United States, ⁶ Institut Pasteur, Paris, France

OPEN ACCESS

Edited by:

Meina Neumann-Schaal,
Deutsche Sammlung von
Mikroorganismen und Zellkulturen
(DSMZ), Germany

Reviewed by:

Konstantinos Papadimitriou,
Agricultural University of Athens,
Greece

Meera Unnikrishnan,
University of Warwick,
United Kingdom

*Correspondence:

Olga Soutourina
olga.soutourina@i2bc.paris-saclay.fr

Specialty section:

This article was submitted to
Infectious Diseases,
a section of the journal
Frontiers in Microbiology

Received: 30 May 2018

Accepted: 12 July 2018

Published: 31 July 2018

Citation:

Maikova A, Severinov K and
Soutourina O (2018) New Insights Into
Functions and Possible Applications
of *Clostridium difficile* CRISPR-Cas
System. *Front. Microbiol.* 9:1740.
doi: 10.3389/fmicb.2018.01740

Over the last decades the enteric bacterium *Clostridium difficile* (novel name *Clostridioides difficile*) – has emerged as an important human nosocomial pathogen. It is a leading cause of hospital-acquired diarrhea and represents a major challenge for healthcare providers. Many aspects of *C. difficile* pathogenesis and its evolution remain poorly understood. Efficient defense systems against phages and other genetic elements could have contributed to the success of this enteropathogen in the phage-rich gut communities. Recent studies demonstrated the presence of an active CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR-associated) subtype I-B system in *C. difficile*. In this mini-review, we will discuss the recent advances in characterization of original features of the *C. difficile* CRISPR-Cas system in laboratory and clinical strains, as well as interesting perspectives for our understanding of this defense system function and regulation in this important enteropathogen. This knowledge will pave the way for the development of promising biotechnological and therapeutic tools in the future. Possible applications for the *C. difficile* strain monitoring and genotyping, as well as for CRISPR-based genome editing and antimicrobials are also discussed.

Keywords: CRISPR, *C. difficile*, I-B subtype CRISPR-Cas system, prophage, CRISPR regulation, stress, antimicrobials, genome editing

CRISPR-Cas SYSTEMS: GENERAL FUNCTIONAL ASPECTS AND CLASSIFICATION

The CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR-associated) systems are adaptive immune systems protecting prokaryotes against phages and other mobile genetic elements (Sorek et al., 2013). These defensive systems are found in almost all sequenced archaeal and in about half of bacterial genomes (Grissa et al., 2007). CRISPR-Cas systems are composed of CRISPR arrays and *cas* operons. CRISPR arrays in turn consist of short direct repeats (20–40 bp) separated by variable spacers. Some spacers are complementary to mobile genetic elements sequences (Shmakov et al., 2017). CRISPR arrays also contain leader regions carrying promoters from which their transcription initiates.

CRISPR-based defensive functions include two major processes: immunity (interference) and immunization (adaptation) (for general review, see Marraffini, 2015). CRISPR interference itself can be divided into two phases: the biogenesis of CRISPR RNAs and the targeting phase. At the first phase a CRISPR array is transcribed into a long RNA transcript (pre-crRNA), which is processed into small CRISPR RNAs (crRNAs), each consisting of one spacer and flanking repeat sequences. Individual crRNAs bind to Cas proteins forming a nucleoprotein effector complex, which is necessary for the second, targeting phase. The crRNAs serve as guides for recognizing nucleic acids by complementary base pairing. In this way, crRNAs direct recognition and, ultimately, cleavage of genetic elements by the Cas nucleases (Garneau et al., 2010). Spacers are incorporated into CRISPR arrays in the process of adaptation (Jackson et al., 2017). Cas1 and Cas2 proteins, found in almost all investigated CRISPR-Cas systems, are essential for this process (Koonin et al., 2017). A very important aspect of CRISPR-based immunity is the ability to distinguish host DNA from the foreign one. Protospacer-adjacent motifs (PAMs) are short sequences situated on the 3' or 5' end of the protospacer (foreign DNA region corresponding to a CRISPR spacer) and required for protospacer recognition. PAMs are absent from CRISPR arrays, which prevents autoimmunity (Sorek et al., 2013). PAMs are essential during spacer selection at the adaptation stage, which ensures that acquired spacers are functional in interference. Previous studies in type I CRISPR-Cas systems identified the sequence requirements for the CRISPR targeting that includes a perfect match between the 5' end of the spacer and the protospacer within up to a 10-nt "seed" sequence (Semenova et al., 2011; Wiedenheft et al., 2011; Maier et al., 2013a).

CRISPR-Cas systems are highly diverse. This is reflected in both CRISPR array architectures and cas genes composition (Takeuchi et al., 2012). The variability of cas gene sets has formed the basis of CRISPR-Cas systems classification (Makarova et al., 2011). All investigated CRISPR-Cas systems are divided into two classes, characterized by the composition of cas genes involved in interference module (Koonin et al., 2017). These classes in turn are divided into six types and 33 subtypes (see Table 1 for examples). The Class 1 comprises the most abundant and diverse type I and type III CRISPR-Cas systems as well as rare type IV. These types of CRISPR-Cas systems are found in both archaeal and bacterial genomes. Effector complexes of the type I and type III include Cas5, Cas7, Cas8 (in type I), and Cas10 (in type III) proteins (Koonin et al., 2017). For crRNA processing Cas6 family proteins are necessary in these CRISPR-Cas systems (Charpentier et al., 2015). Type I systems are also characterized by the presence of Cas3 proteins responsible for degradation of DNA recognized by effector complexes (Brouns et al., 2008). The Class 2 includes type II, type V and type VI CRISPR-Cas systems. These systems possess effector modules consisting of only one multi-domain protein. The most characterized is the type II Cas9 protein widely used in genome editing (Wang et al., 2016).

The type I CRISPR-Cas systems are highly diverse and subdivided into seven subtypes (I-A, I-B, I-C, I-U, I-D, I-E, I-F) (Makarova et al., 2015). The subtypes I-C, I-D, I-E, I-F are encoded by a single operon in CRISPR loci, whereas

subtype I-A and I-B are often encoded by several operons. I-C, I-E, and I-F subtypes are mostly present in bacteria, while I-A, I-B, and I-D are common in archaea (Makarova et al., 2011) (Table 1). The subtype I-B, characterized by a specific Cas8b protein, is present in methanogenic and halophilic archaea and in clostridia. Studies of the I-B CRISPR-Cas systems in haloarchaea showed some interesting features such as multiple PAMs and 9-nucleotide non-contiguous seed region (Maier et al., 2015). Although the subtype I-B was found in clostridial species it has not been well studied there yet. It is suggested that I-B CRISPR-Cas system possibly had been acquired by clostridia from archaea via horizontal gene transfer and afterward evolved independently (Peng et al., 2014). Other CRISPR-Cas systems subtypes, including I-A, I-C, III-A, III-B, and II-C, are also present in some clostridial species (Table 1).

CHARACTERIZATION OF *Clostridium difficile* CRISPR-Cas SYSTEM

Clostridium difficile (novel name *Clostridioides difficile*) is an anaerobic spore-forming bacterium, one of the major clostridial pathogens and the major cause of nosocomial infections associated with antibiotic therapy (Abt et al., 2016). During its infection cycle, this enteropathogen must cope with the presence of foreign DNA elements, including bacteriophages, in the crowded environment of the gut, and is thus expected to rely on efficient defense systems such as CRISPR-Cas to control genetic exchanges favored in its complex niche.

The first evidence suggesting the presence of active CRISPR-Cas system in *C. difficile* was obtained during deep-sequencing of regulatory RNAs in *C. difficile* (Soutourina et al., 2013). This study revealed abundant and diverse crRNAs. Active expression and processing of CRISPR loci was detected in this and a subsequent study (Soutourina et al., 2013; Boudry et al., 2015).

Bioinformatics analysis of more than 200 *C. difficile* genomes (Hargreaves et al., 2014; Andersen et al., 2016) demonstrated that *C. difficile* CRISPR-Cas system belongs to I-B subtype (Koonin et al., 2017). *C. difficile* CRISPR-Cas system possesses several original features (Figure 1). CRISPR-Cas system of this enteropathogen is characterized by an unusual large set of CRISPR arrays. For example, reference 630 and hypervirulent R20291 *C. difficile* strains contain 12 and 9 CRISPR arrays, respectively (Soutourina et al., 2013; Boudry et al., 2015). These CRISPR arrays are orientated in the direction of chromosome replication, as observed for highly expressed bacterial genes and presumably ensuring their optimal transcription (Arakawa and Tomita, 2007; Boudry et al., 2015). On average, known *C. difficile* genomes contain 8.5 arrays (Andersen et al., 2016). In most sequenced *C. difficile* strains several CRISPR arrays are located in prophages (Hargreaves et al., 2014; Boudry et al., 2015). The crRNAs originating from arrays located in prophages were found to be the most expressed in 630 and R20291 strains. Prophage localization of actively expressed CRISPR arrays may play a role in preventing infection by related competing phages by targeting their DNA (Sorek et al., 2008).

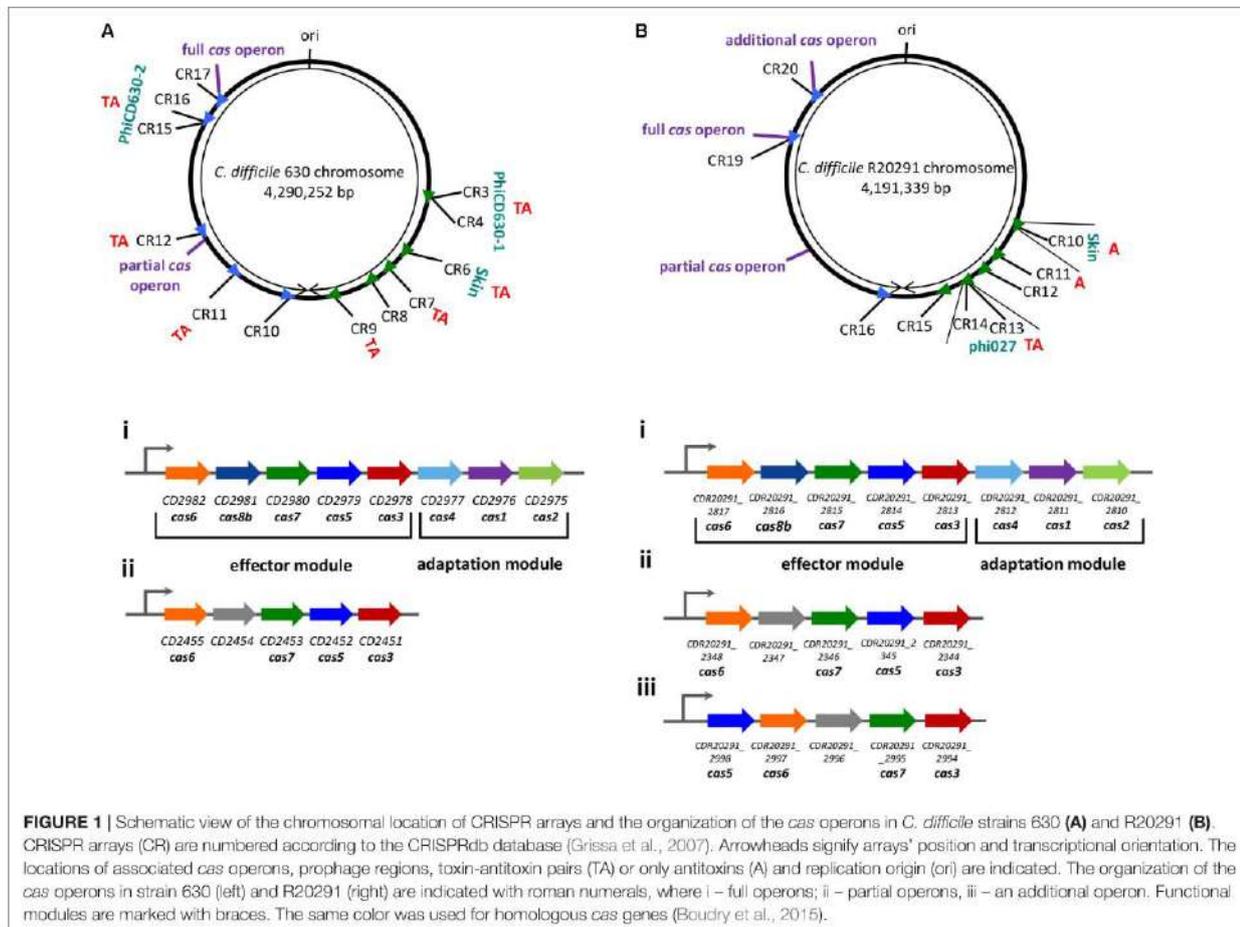
Another unusual feature of *C. difficile* CRISPR-Cas system is the presence of two or three (in 027 ribotype strains) *cas* gene sets in the majority of sequenced strains (Boudry et al., 2015) (Figure 1). The full *cas* operon encodes all necessary genes for CRISPR interference (*cas6*, *cas8b*, *cas7*, *cas5*, *cas3*) as well as *cas1*, *cas2*, *cas4* genes essential for spacer acquisition (Amitai and Sorek, 2016; Kieper et al., 2018; Lee et al., 2018). The additional *cas* operons lack the adaptation module. While the complete *cas* gene operons were found in ~90% of sequenced *C. difficile* strains, the additional partial *cas* gene sets are present in almost all strains (Boudry et al., 2015). Thus, some *C. difficile* strains could

have lost the ability to adapt to new genetic elements through their CRISPR-Cas systems. The *cas* operon occurrence correlates with evolutionary relationships of *C. difficile* strains reflecting their epidemiological context and, possibly, the intensity of interactions with foreign DNA elements (Boudry et al., 2015). When present, complete *cas* gene operons are usually associated with longest CRISPR arrays, which is indicative of active new spacer acquisition (or slower spacer loss) and hints to an existence of some still unknown *in cis* mechanisms responsible for different dynamics of *cas* proximal arrays. The conservation of CRISPR array structure and sequences of all CRISPR repeats in *C. difficile*

TABLE 1 | Main CRISPR-Cas systems subtypes and examples of system-harboring microorganisms and clostridial species.

Class	Subtype	<i>cas</i> operon composition	Example	Examples of clostridial species and strains	
Class 1	I-A	<i>cas6</i> , <i>cas11(csa5)</i> , <i>cas7</i> , <i>cas5</i> , <i>cas8a1</i> , <i>cas3*</i> , <i>cas3*</i> , <i>cas2</i> , <i>cas4</i> , <i>cas1</i> , <i>cas4</i>	<i>Listeria monocytogenes</i> L99 (Sesto et al., 2014)	<i>C. stercoarium</i> subsp. <i>stercoarium</i> DSM 8532 (Poehlein et al., 2013); <i>C. tetani</i> ATCC 9441 (Cohen et al., 2017)	
	I-B	<i>cas6</i> , <i>cas8b1</i> , <i>cas7</i> , <i>cas5</i> , <i>cas3</i> , <i>cas4</i> , <i>cas1</i> , <i>cas2</i>	<i>Haloflex volcanii</i> H119 (Maier et al., 2013b)	<i>C. difficile</i> 630, <i>C. difficile</i> R20291 (Boudry et al., 2015); <i>C. pasteurianum</i> BC1 (Pyne et al., 2016); <i>C. acetobutylicum</i> GXAS18-1 (Peng et al., 2014); <i>C. tetani</i> ATCC 9441 (Cohen et al., 2017)	
	I-C	<i>cas3</i> , <i>cas5</i> , <i>cas8c</i> , <i>cas7</i> , <i>cas4</i> , <i>cas1</i> , <i>cas2</i>	<i>Desulfovibrio vulgaris</i> str. <i>Hildenborough</i> (Hochstrasser et al., 2016)	<i>C. cellulolyticum</i> H10 (Brown et al., 2014)	
	I-U	<i>cas3</i> , <i>cas8u2</i> , <i>cas7</i> , <i>cas5-cas6</i> , <i>cas4-cas1</i> , <i>cas2</i>	<i>Geobacter sulfurreducens</i> (Koonin et al., 2017)	–	
	I-D	<i>cas3*</i> , <i>cas3*</i> , <i>cas10d</i> , <i>cas7(csc2)</i> , <i>cas5(csc1)</i> , <i>cas6</i> , <i>cas4</i> , <i>cas1</i> , <i>cas2</i>	<i>Cyanotheca</i> sp. 8802 (Koonin et al., 2017)	–	
	I-E	<i>cas3</i> , <i>cas8e(cse1)</i> , <i>cas11(cse2)</i> , <i>cas7</i> , <i>cas5</i> , <i>cas6</i> , <i>cas1</i> , <i>cas2</i>	<i>Escherichia coli</i> K12 (Koonin et al., 2017)	–	
	I-F	<i>cas1</i> , <i>cas2-cas3</i> , <i>cas8f(csy1)</i> , <i>cas5(csy2)</i> , <i>cas7(csy3)</i> , <i>cas6f</i>	<i>Pseudomonas aeruginosa</i> PA14 (Wiedenheft et al., 2011)	–	
	III-A	<i>cas6</i> , <i>cas10</i> , <i>cas11(csm2)</i> , <i>cas7(csm3)</i> , <i>cas5(csm4)</i> , <i>cas7(csm5)</i> , <i>csm6</i> , <i>cas1</i> , <i>cas2</i>	<i>Staphylococcus epidermidis</i> (Koonin et al., 2017)	<i>C. tetani</i> ATCC 453 (Cohen et al., 2017)	
	III-B	<i>cas7(cmr1)</i> , <i>cas10</i> , <i>cas5(cmr3)</i> , <i>cas7(cmr4)</i> , <i>cas11(cmr5)</i> , <i>cas6</i> , <i>cas7(cmr6)</i>	<i>Pyrococcus furiosus</i> (Koonin et al., 2017)	<i>C. botulinum</i> ATCC 3502 (Negahdaripour et al., 2017)	
	III-C	<i>cas7(cmr1)</i> , <i>cas7(cmr6)</i> , <i>cas10</i> , <i>cas7(cmr4)</i> , <i>cas11(cmr5)</i> , <i>cas5(cmr3)</i>	<i>Methanothermobacter thermoautotrophicus</i> (Koonin et al., 2017)	–	
	III-D	<i>cas10</i> , <i>cas7(csm3)</i> , <i>cas5(csx10)</i> , <i>cas11(csm2)</i> , <i>cas7(csm7)</i> , <i>cas7(csm5)</i> , <i>all1473</i> , <i>cas7(csm5)</i>	<i>Synechocystis</i> sp. 6803 (Makarova et al., 2015)	–	
	Class 2	II-A	<i>cas9</i> , <i>cas1</i> , <i>cas2</i> , <i>csn2</i>	<i>Enterococcus faecalis</i> OG1RF (Bourgogne et al., 2008)	–
		II-B	<i>cas9</i> , <i>cas1</i> , <i>cas2</i> , <i>cas4</i>	<i>Legionella pneumophila</i> str. <i>Paris</i> (Koonin et al., 2017)	–
		II-C	<i>cas9</i> , <i>cas1</i> , <i>cas2</i>	<i>Neisseria lactamica</i> 020-06 (Koonin et al., 2017)	<i>C. perfringens</i> JGS1495 (Pearson et al., 2015)
V-A		<i>cas12a(cpf1)</i> , <i>cas4</i> , <i>cas1</i> , <i>cas2</i>	<i>Francisella cf. novicida</i> Fx1 (Koonin et al., 2017)	–	
V-B		<i>cas12b(c2c1)</i> , <i>cas4</i> , <i>cas1</i> , <i>cas2</i>	<i>Alicyclobacillus acidoterrestris</i> (Koonin et al., 2017)	–	
V-C		<i>cas1</i> , <i>cas12c(c2c3)</i>	<i>Oleiphilus</i> sp. (Koonin et al., 2017)	–	
V-D		<i>cas1</i> , <i>cas12d(casY)</i>	<i>Bacterium</i> CG09_39_24 (Koonin et al., 2017)	–	
V-E	<i>cas12e(casX)</i> , <i>cas4</i> , <i>cas1</i> , <i>cas2</i>	<i>Delta proteobacteria bacterium</i> (Koonin et al., 2017)	–		

CRISPR-Cas systems subtypes and the composition of *cas* operons are shown according to classification of Koonin et al. (2017). Fused *cas* genes in operons are marked with a dash.



genomes suggests that CRISPR arrays located far from *cas* operons use the same set of Cas proteins for their function.

An interesting evolutionary aspect of *C. difficile* CRISPR-Cas system has been recently reported (Maikova et al., 2018). Analysis of about 2,500 *C. difficile* genomes revealed co-localization of type I toxin-antitoxin (TA) modules with CRISPR arrays (Figure 1). TA – CRISPR array co-localization has never been reported for other bacteria and its significance remains unclear. CRISPR-arrays localized in prophage regions are in particular prone to be associated with type I TA modules, which may contribute to the stabilization of these chromosomal regions by the mechanism similar to plasmid maintenance through post-segregation killing.

The function of CRISPR-Cas system is to provide defense against viruses and other mobile genetic elements. Recent bioinformatics analysis of *C. difficile* CRISPR spacers matching known sequences showed that most of them target clostridial phages and prophage regions (Hargreaves et al., 2014; Boudry et al., 2015). This suggests that this enteropathogen actively interacts with phages, and that CRISPR-Cas actively modulates this interaction. Identification of protospacers allowed to deduce PAM sequences. While 3-nucleotide 5'-motifs CCA or CCT were most common, alternative sequences CCC, CCG, and TCA

were also frequently found. Multiple PAMs were also observed in other type I-B systems (Shah et al., 2013). Conjugation efficiency experiments with plasmid vectors containing CCA and CCT PAMs and protospacers corresponding to the first spacers from actively expressed *C. difficile* 630 CRISPR arrays showed active CRISPR interference in *C. difficile* cells thus validating bioinformatically predicted PAMs and showing that *C. difficile* CRISPR-Cas system is naturally capable of defensive function (Boudry et al., 2015). Phage infection assays in 630 and R20291 strains revealed the correlation between the presence of CRISPR spacer-targeting phage sequences and phage susceptibility. Experiments using a heterologous *E. coli* host system showed that both *cas* operons of *C. difficile* 630 strain are capable of interference.

REGULATION AND POTENTIAL ALTERNATIVE FUNCTIONS OF *C. difficile* CRISPR-Cas SYSTEM

During its infection cycle *C. difficile* faces with different stress conditions and changing environments inside the host. To

survive in phage-rich gut community while relying on the CRISPR-Cas defense, *C. difficile* needs to regulate CRISPR-Cas expression in response to different environmental signals. A study by Boudry et al. (2015) revealed that all the CRISPR arrays and *cas* genes are expressed under standard laboratory conditions. The level of CRISPR-Cas expression could be modulated by specific regulatory mechanisms.

Bacterial pathogens often form biofilms, which help them resist different threats inside the host. It was shown that *C. difficile* actively forms biofilms (Dapa et al., 2013; Nale et al., 2016; Soavelomandroso et al., 2017) during its infection cycle. Biofilm conditions are characterized by high cell densities, which increase the possibility of horizontal gene transfer (Babic et al., 2011; Abedon, 2012). Quorum sensing is one of bacterial mechanisms that regulates gene expression depending on the density of the population (Miller and Bassler, 2001). Recent studies showed that *cas* gene expression is induced by quorum sensing signals in *Serratia* sp. (I-E, I-F, and III-A subtypes) (Patterson et al., 2016) and *Pseudomonas aeruginosa* (I-F subtype) (Høyland-Kroghsbo et al., 2016). Moreover, CRISPR-Cas systems may play a role in biofilm formation and colonization of the host. For instance, CRISPR-Cas (II-A subtype) harboring *Enterococcus faecalis* strain has increased biofilm formation (Bourgogne et al., 2008). Furthermore, CRISPR-Cas-mediated gene regulation of the ability to swarm and form biofilms was revealed in *P. aeruginosa* (Zegans et al., 2009). In *C. difficile* strain 630, a recent study revealed up to 20-fold induction of *cas* genes expression in biofilms (Maikova et al., 2018), suggesting the regulation of *C. difficile* CRISPR-Cas system activity by biofilm-related factors. During infection, the complex interactions with different microbiota members within gut communities should be considered. More studies are thus needed to assess the possible link between biofilm-related signals and the regulation of CRISPR-Cas expression during the *C. difficile* infection cycle.

The obvious stress to induce CRISPR-Cas system is phage infection. At the earliest stages of attachment to the cell surface, it is often accompanied by the envelope stress (Ratner et al., 2015). The induction of the CRISPR-Cas system expression in response to this type of stress was found in different bacteria (Westra et al., 2014). Bacterial pathogens and commensals always combat with the host's immune response, which results in a wide range of stressful effects. Several studies reported the changes of *cas* gene transcription in *Desulfovibrio vulgaris* (Mukhopadhyay et al., 2007), *Streptococcus sanguinis* (Rodriguez et al., 2011), *Pasteurella multocida* (Melnikow et al., 2008), and *Lactobacillus rhamnosus* (Koskenniemi et al., 2011) in response to different stresses such as changes in growth rate, bile, oxidative, nitrosative stresses and exposure to antibiotics. Virulence is a specific response of pathogens to different stress factors inside the host (Louwen et al., 2014). The regulation of CRISPR-Cas systems during the infection cycle may indicate an important role of these systems in pathogens. Recently, a role of an alternative SigB factor in stress response was investigated in *C. difficile* (Kint et al., 2017). Interestingly, SigB-dependent promoters were found upstream of both *cas* operons in *C. difficile* strain 630 (Maikova et al., 2018) and fivefold decrease in expression levels of both *cas* operons was observed in *sigB* mutant strain. This suggests regulation of

C. difficile CRISPR-Cas system via stress-related signals and a potential role of this system in the survival of *C. difficile* inside the host.

Besides the adaptive immunity, multiple alternative functions of CRISPR-Cas systems have been revealed (Louwen et al., 2014; Westra et al., 2014). These functions occur through targeting bacterium's own genes by partially or fully matching crRNAs. For instance, in *Listeria monocytogenes* a specific long type I-A CRISPR array transcript *rliB* processed by polynucleotide phosphorylase (PNPase) controls the expression of the *feoAB* genes important for virulence (Mandin et al., 2007; Sesto et al., 2014). An *rliB* mutant colonizes its host more effectively than the wild type strain. Bioinformatics analysis of *C. difficile* CRISPR spacers showed that all investigated strains carry genome-targeting spacers (Boudry et al., 2015). It may thus be speculated that *C. difficile* CRISPR-Cas system might also have functions in the regulation of the endogenous gene expression. The possible role of CRISPR-Cas systems in genome evolution via self-targeting is actively discussed (Westra et al., 2014).

POTENTIAL APPLICATIONS OF *C. difficile* CRISPR-Cas SYSTEM

During the last decade, discoveries in the CRISPR field led to rapid development of revolutionary biotechnological applications especially in genome editing by CRISPR-Cas9 technology (Hsu et al., 2014). Different CRISPR-based tools have proved to be effective both in prokaryotes and eukaryotes (Hsu et al., 2014; Barrangou and Horvath, 2017).

Since spacers are acquired in an orderly manner, with more recently acquired spacer being closer to the leader sequence (Barrangou et al., 2007; Nuñez et al., 2015) the order of spacers within an array reflects phage invasions in different populations of the same bacterial species. This feature can reveal phylogenetic relations between strains and can be used in genotyping techniques (Louwen et al., 2014; Andersen et al., 2016). Such "CRISPR-typing" has been already applied for outbreak tracking of *Yersinia pestis* (Cui et al., 2008; Barros et al., 2014) and *Salmonella enterica* (Timme et al., 2013; Pettengill et al., 2014). Moreover, CRISPR typing is capable to reveal antibiotic-resistant phenotypes (Palmer and Gilmore, 2010) or prophages (Nozawa et al., 2011). These correlations can be explained by the influence of active CRISPR-Cas systems on the horizontal gene transfer, which plays important role in the acquisition of new genes and operons, essential for bacterial pathogenesis and adaptation (Louwen et al., 2014). CRISPR-typing approach based on spacer content and polymorphism can be successfully applied to *C. difficile* with correlation between CRISPR-groups and toxin groups (Andersen et al., 2016).

CRISPR-Cas systems can be applied for development of new antimicrobials based on the self-targeting (Bikard et al., 2012). The general strategy is the use of phage particles and phagemids as vectors to deliver auto-targeting CRISPR-Cas components inside a pathogenic cell (Bikard and Barrangou, 2017). Many pathogens possess endogenous active CRISPR-Cas systems, which can be repurposed for self-targeting. Since *C. difficile*

contains a naturally active CRISPR-Cas system, such a strategy could be promising for control and even treatment of *C. difficile* infection (CDI), in the context of recent worldwide emergence of antibiotic-resistant *C. difficile* strains (Banawas, 2018). Phage therapy of CDI has proved to be another promising alternative, but it faces some difficulties including lack of appropriate phages (Hargreaves and Clokie, 2014; Sekulovic et al., 2014). The presence of active CRISPR-Cas system should effectively prevent infection by at least some phages complicating matters further.

The most popular biotechnological application of CRISPR-Cas systems is genome editing (Barrangou and Horvath, 2017). In prokaryotes, the most interesting is the application of endogenous CRISPR-Cas systems since it requires the introduction of less additional components for the editing process. Several works showing the applications of endogenous I-B subtype systems for genome editing were recently published. The first one, by Pyne et al. (2016) describes this approach in *Clostridium pasteurianum*. In this study, a plasmid vector containing an artificial CRISPR array with a protospacer targeting the gene of interest and arms for homologous recombination was used to delete the *cpaAIR* gene encoding a restriction enzyme (Pyne et al., 2016). This approach allows fast and markless deletion or modification of the genes of interest in bacteria. Later, other studies confirmed the efficiency of this method in other I-B subtype-carrying organisms: archaeon *Haloarcula hispanica* (Cheng et al., 2017) and butanol producing *Clostridium tyrobutyricum* (Zhang et al., 2018). Another study revealed that *Haloferax volcanii* CRISPR-Cas system with deletions of *cas3* and *cas6* genes can be used for programmable repression of genes in this archaeon (Stachler and Marchfelder, 2016). Many efficient approaches for *C. difficile* genome manipulation exist to date. Clostron is a method based on altered type II intron, which is able to insert in almost every region of the chromosome (Kuehne et al., 2011). Another method is CodA allele exchange

technique based on semi-suicidal vector carrying the *E. coli codA* gene as a counter-selectable marker (Cartman et al., 2012). Successful application of CRISPR-Cas9 (McAllister et al., 2017; Wang et al., 2018) and Cpf1 (Hong et al., 2018) systems for genome editing in *C. difficile* was recently reported and may further extend our ability to manipulate the genome of this pathogen.

Despite the recent insights, many aspects of *C. difficile* CRISPR-Cas system remain to be characterized. We hope that future studies will shed new light on the secrets of *C. difficile* success within host environments relying on effective defense systems and will lead to promising medical and biotechnological applications.

AUTHOR CONTRIBUTIONS

AM wrote the draft of the paper. OS and KS designed the project and performed critical revisions of the manuscript.

FUNDING

This work was supported by grants from the Paris Diderot University, Agence Nationale de la Recherche ("CloSTARn," ANR-13-JSV3-0005-01 to OS), the Institut Universitaire de France (to OS), the Pasteur-Weizmann Council, the University Paris-Sud, the Institute for Integrative Biology of the Cell, Vernadski fellowship to AM and the Skoltech Biomedical Initiative grant (SBI RF-0000000136) to KS.

ACKNOWLEDGMENTS

We are grateful to E. Semenova for helpful discussions.

REFERENCES

- Abedon, S. T. (2012). Spatial vulnerability: bacterial arrangements, microcolonies, and biofilms as responses to low rather than high phage densities. *Viruses* 4, 663–687. doi: 10.3390/v4050663
- Abt, M. C., McKenney, P. T., and Pamer, E. G. (2016). *Clostridium difficile* colitis: pathogenesis and host defence. *Nat. Rev. Microbiol.* 14, 609–620. doi: 10.1038/nrmicro.2016.108
- Amitai, G., and Sorek, R. (2016). CRISPR-Cas adaptation: insights into the mechanism of action. *Nat. Rev. Microbiol.* 14, 67–76. doi: 10.1038/nrmicro.2015.14
- Andersen, J. M., Shoup, M., Robinson, C., Britton, R., Olsen, K. E. P., and Barrangou, R. (2016). CRISPR Diversity and microevolution in *Clostridium difficile*. *Genome Biol. Evol.* 8, 2841–2855. doi: 10.1093/gbe/evw203
- Arakawa, K., and Tomita, M. (2007). Selection effects on the positioning of genes and gene structures from the interplay of replication and transcription in bacterial genomes. *Evol. Bioinform. Online* 3, 279–286. doi: 10.1177/1176934307003000005
- Babic, A., Berkmen, M. B., Lee, C. A., and Grossman, A. D. (2011). Efficient gene transfer in bacterial cell chains. *mBio* 2:e00027-11. doi: 10.1128/mBio.00027-11
- Banawas, S. S. (2018). *Clostridium difficile* infections: a global overview of drug sensitivity and resistance mechanisms. *Biomed Res. Int.* 2018:8414257. doi: 10.1155/2018/8414257
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., et al. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709–1712. doi: 10.1126/science.1138140
- Barrangou, R., and Horvath, P. (2017). A decade of discovery: CRISPR functions and applications. *Nat. Microbiol.* 2:17092. doi: 10.1038/nmicrobiol.2017.92
- Barros, M. P., França, C. T., Lins, R. H., Santos, M. D., Silva, E. J., Oliveira, M. B., et al. (2014). Dynamics of CRISPR loci in microevolutionary process of *Yersinia pestis* strains. *PLoS One* 9:e108353. doi: 10.1371/journal.pone.0108353
- Bikard, D., and Barrangou, R. (2017). Using CRISPR-Cas systems as antimicrobials. *Curr. Opin. Microbiol.* 37, 155–160. doi: 10.1016/j.mib.2017.08.005
- Bikard, D., Hatoum-Aslan, A., Mucida, D., and Marraffini, L. A. (2012). CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. *Cell Host Microbe* 12, 177–186. doi: 10.1016/j.chom.2012.06.003
- Boudry, P., Semenova, E., Monot, M., Datsenko, K. A., Lopatina, A., Sekulovic, O., et al. (2015). Function of the CRISPR-Cas system of the human pathogen *Clostridium difficile*. *mBio* 6:e01112-15. doi: 10.1128/mBio.01112-15
- Bourgogne, A., Garsin, D. A., Qin, X., Singh, K. V., Sillanpaa, J., Yerrapragada, S., et al. (2008). Large scale variation in *Enterococcus faecalis* illustrated by the genome analysis of strain OG1RF. *Genome Biol.* 9:R110. doi: 10.1186/gb-2008-9-7-r110
- Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuys, R. J., Snijders, A. P., et al. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321, 960–964. doi: 10.1126/science.1159689

- Brown, S. D., Nagaraju, S., Utturkar, S., De Tissera, S., Segovia, S., Mitchell, W., et al. (2014). Comparison of single-molecule sequencing and hybrid approaches for finishing the genome of *Clostridium autoethanogenum* and analysis of CRISPR systems in industrial relevant Clostridia. *Biotechnol. Biofuels* 7:40. doi: 10.1186/1754-6834-7-40
- Cartman, S. T., Kelly, M. L., Heeg, D., Heap, J. T., and Minton, N. P. (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. doi: 10.1128/AEM.00249-12
- Charpentier, E., Richter, H., van der Oost, J., and White, M. F. (2015). Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. *FEMS Microbiol. Rev.* 428–441. doi: 10.1093/femsre/fuv023
- Cheng, F., Gong, L., Zhao, D., Yang, H., Zhou, J., Li, M., et al. (2017). Harnessing the native type I-B CRISPR-Cas for genome editing in a polyploid archaeon. *J. Genet. Genomics* 44, 541–548. doi: 10.1016/j.jgg.2017.09.010
- Cohen, J. E., Wang, R., Shen, R.-F., Wu, W. W., and Keller, J. E. (2017). Comparative pathogenomics of *Clostridium tetani*. *PLoS One* 12:e0182909. doi: 10.1371/journal.pone.0182909
- Cui, Y., Li, Y., Gorgé, O., Platonov, M. E., Yan, Y., Guo, Z., et al. (2008). Insight into microevolution of *Yersinia pestis* by clustered regularly interspaced short palindromic repeats. *PLoS One* 3:e2652. doi: 10.1371/journal.pone.0002652
- Dapa, T., Leuzzi, R., Ng, Y. K., Baban, S. T., Adamo, R., Kuehne, S. A., et al. (2013). Multiple factors modulate biofilm formation by the anaerobic pathogen *Clostridium difficile*. *J. Bacteriol.* 195, 545–555. doi: 10.1128/JB.01980-1912
- Garneau, J. E., Dupuis, M.-V., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., et al. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468, 67–71. doi: 10.1038/nature09523
- Grissa, I., Vergnaud, G., and Pourcel, C. (2007). The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* 8:172. doi: 10.1186/1471-2105-8-172
- Hargreaves, K. R., and Clokie, M. R. J. (2014). *Clostridium difficile* phages: Still difficult? *Front. Microbiol.* 5:184. doi: 10.3389/fmicb.2014.00184
- Hargreaves, K. R., Flores, C. O., Lawley, T. D., Clokie, R. J., Trevor, D., Hargreaves, K. R., et al. (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. doi: 10.1128/mBio.01045-13
- Hochstrasser, M. L., Taylor, D. W., Kornfeld, J. E., Nogales, E., and Doudna, J. A. (2016). DNA targeting by a minimal CRISPR RNA-guided cascade. *Mol. Cell* 63, 840–851. doi: 10.1016/j.molcel.2016.07.027
- Hong, W., Zhang, J., Cui, G., Wang, L., and Wang, Y. (2018). Multiplexed CRISPR-CpfI-mediated genome editing in *Clostridium difficile* toward the understanding of pathogenesis of *C. difficile* infection. *ACS Synth. Biol.* 7, 1588–1600. doi: 10.1021/acssynbio.8b00087
- Høyland-Kroghsbo, N. M., Paczkowski, J., Mukherjee, S., Broniewski, J., Westra, E., Bondy-Denomy, J., et al. (2016). Quorum sensing controls the *Pseudomonas aeruginosa* CRISPR-Cas adaptive immune system. *Proc. Natl. Acad. Sci. U.S.A.* 114, 131–135. doi: 10.1073/pnas.1617415113
- Hsu, P. D., Lander, E. S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157, 1262–1278. doi: 10.1016/j.cell.2014.05.010
- Jackson, S. A., McKenzie, R. E., Fagerlund, R. D., Kieper, S. N., Fineran, P. C., and Brouns, S. J. J. (2017). CRISPR-Cas: adapting to change. *Science* 356:eaal5056. doi: 10.1126/science.aal5056
- Kieper, S. N., Almendros, C., Behler, J., McKenzie, R. E., Nobrega, F. L., Haagsma, A. C., et al. (2018). Cas4 facilitates PAM-compatible spacer selection during CRISPR adaptation. *Cell Rep.* 22, 3377–3384. doi: 10.1016/j.celrep.2018.02.103
- Kint, N., Janoir, C., Monot, M., Hoys, S., Soutourina, O., Dupuy, B., et al. (2017). The alternative sigma factor σ B plays a crucial role in adaptive strategies of *Clostridium difficile* during gut infection. *Environ. Microbiol.* 19, 1933–1958. doi: 10.1111/1462-2920.13696
- Koonin, E. V., Makarova, K. S., and Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.* 37, 67–78. doi: 10.1016/j.mib.2017.05.008
- Koskeniemi, K., Laakso, K., Koponen, J., Kankainen, M., Greco, D., Auvinen, P., et al. (2011). roteomics and transcriptomics characterization of bile stress response in probiotic *Lactobacillus rhamnosus* GG. *Mol. Cell. Proteomics* 10, M110.002741. doi: 10.1074/mcp.M110.002741
- Kuehne, S. A., Heap, J. T., Cooksley, C. M., Cartman, S. T., and Minton, N. P. (2011). ClosTron-mediated engineering of *Clostridium*. *Methods Mol. Biol.* 765, 389–407. doi: 10.1007/978-1-61779-197-0_23
- Lee, H., Zhou, Y., Taylor, D. W., and Sashital, D. G. (2018). Cas4-dependent prespacer processing ensures high-fidelity programming of CRISPR arrays. *Mol. Cell* 70, 48.e5–59.e5. doi: 10.1016/j.molcel.2018.03.003
- Louwen, R., Staals, R. H. J., Endtz, H. P., van Baarlen, P., and van der Oost, J. (2014). The role of CRISPR-Cas systems in virulence of pathogenic bacteria. *Microbiol. Mol. Biol. Rev.* 78, 74–88. doi: 10.1128/MMBR.00039-13
- Maier, L.-K., Dyll-Smith, M., and Marchfelder, A. (2015). The adaptive immune system of *Haloflex volcanii*. *Life* 5, 521–537. doi: 10.3390/life5010521
- Maier, L.-K., Lange, S. J., Stoll, B., Haas, K. A., Fischer, S. M., Fischer, E., et al. (2013a). Essential requirements for the detection and degradation of invaders by the *Haloflex volcanii* CRISPR/Cas system I-B. *RNA Biol.* 10, 865–874. doi: 10.4161/rna.24282
- Maier, L.-K., Stoll, B., Brendel, J., Fischer, S., Pfeiffer, F., Dyll-Smith, M., et al. (2013b). The ring of confidence: a haloarchaeal CRISPR/Cas system. *Biochem. Soc. Trans.* 41, 374–378. doi: 10.1042/BST20120263
- Maikova, A., Peltier, J., Boudry, P., Hajnsdorf, E., Kint, N., Monot, M., et al. (2018). Discovery of new type I toxin-antitoxin systems adjacent to CRISPR arrays in *Clostridium difficile*. *Nucleic Acids Res.* 46, 4733–4751. doi: 10.1093/nar/gky124
- Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J. J., Charpentier, E., Horvath, P., et al. (2011). Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* 9, 467–477. doi: 10.1038/nrmicro2577
- Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., et al. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* 13, 722–736. doi: 10.1038/nrmicro3569
- Mandin, P., Repoila, F., Vergassola, M., Geissmann, T., and Cossart, P. (2007). Identification of new noncoding RNAs in *Listeria monocytogenes* and prediction of mRNA targets. *Nucleic Acids Res.* 35, 962–974. doi: 10.1093/nar/gkl1096
- Marraffini, L. A. (2015). CRISPR-Cas immunity in prokaryotes. *Nature* 526, 55–61. doi: 10.1038/nature15386
- McAllister, K. N., Bouillaut, L., Kahn, J. N., Self, W. T., and Sorg, J. A. (2017). Using CRISPR-Cas9-mediated genome editing to generate *C. difficile* mutants defective in selenoproteins synthesis. *Sci. Rep.* 7:14672. doi: 10.1038/s41598-017-15236-5
- Melnikow, E., Schoenfeld, C., Spehr, V., Warrass, R., Gunkel, N., Duszenko, M., et al. (2008). A compendium of antibiotic-induced transcription profiles reveals broad regulation of *Pasteurella multocida* virulence genes. *Vet. Microbiol.* 131, 277–292. doi: 10.1016/j.vetmic.2008.03.007
- Miller, M. B., and Bassler, B. L. (2001). Quorum sensing in bacteria. *Annu. Rev. Microbiol.* 55, 165–199. doi: 10.1146/annurev.micro.55.1.165
- Mukhopadhyay, A., Redding, A. M., Joachimiak, M. P., Arkin, A. P., Borglin, S. E., Dehal, P. S., et al. (2007). Cell-wide responses to low-oxygen exposure in *Desulfovibrio vulgaris* Hildenborough. *J. Bacteriol.* 189, 5996–6010. doi: 10.1128/JB.00368-07
- Nale, J. Y., Chutia, M., Carr, P., Hickenbotham, P. T., Clokie, M. R. J., and Abedon, S. T. (2016). 'Get in Early': biofilm and Wax Moth (*Galleria mellonella*) Models reveal new insights into the therapeutic potential of *Clostridium difficile* bacteriophages. 7, 1–16. doi: 10.3389/fmicb.2016.01383
- Negahdaripour, M., Nezafat, N., Hajigharamani, N., Rahmatabadi, S. S., and Ghasemi, Y. (2017). Investigating CRISPR-Cas systems in *Clostridium botulinum* via bioinformatics tools. *Infect. Genet. Evol.* 54, 355–373. doi: 10.1016/j.meegid.2017.06.027
- Nozawa, T., Furukawa, N., Aikawa, C., Watanabe, T., Haobam, B., Kurokawa, K., et al. (2011). CRISPR inhibition of prophage acquisition in *Streptococcus pyogenes*. *PLoS One* 6:e19543. doi: 10.1371/journal.pone.0019543
- Núñez, J. K., Lee, A. S. Y., Engelman, A., and Doudna, J. A. (2015). Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* 519, 193–198. doi: 10.1038/nature14237
- Palmer, K. L., and Gilmore, M. S. (2010). Multidrug-resistant enterococci lack CRISPR-cas. *mBio* 1:e00227-10. doi: 10.1128/mBio.00227-10
- Patterson, A. G., Jackson, S. A., Taylor, C., Evans, G. B., Salmond, G. P. C., Przybilski, R., et al. (2016). Quorum sensing controls adaptive immunity through the regulation of multiple CRISPR-Cas systems. *Mol. Cell* 64, 1102–1108. doi: 10.1016/j.molcel.2016.11.012

- Pearson, B. M., Louwen, R., van Baarlen, P., and van Vliet, A. H. M. (2015). Differential distribution of type II CRISPR-Cas systems in agricultural and nonagricultural *Campylobacter coli* and *Campylobacter jejuni* isolates correlates with lack of shared environments. *Genome Biol. Evol.* 7, 2663–2679. doi: 10.1093/gbe/evv174
- Peng, L., Pei, J., Pang, H., Guo, Y., Lin, L., and Huang, R. (2014). Whole genome sequencing reveals a novel CRISPR system in industrial *Clostridium acetobutylicum*. *J. Ind. Microbiol. Biotechnol.* 41, 1677–1685. doi: 10.1007/s10295-014-1507-3
- Pettengill, J. B., Timme, R. E., Barrangou, R., Toro, M., Allard, M. W., Strain, E., et al. (2014). The evolutionary history and diagnostic utility of the CRISPR-Cas system within *Salmonella enterica* ssp. *enterica*. *PeerJ* 2:e340. doi: 10.7717/peerj.340
- Poehlein, A., Zverlov, V. V., Daniel, R., Schwarz, W. H., and Liebl, W. (2013). Complete genome sequence of *Clostridium stercoarium* subsp. *stercoarium* strain DSM 8532, a thermophilic degrader of plant cell wall fibers. *Genome Announc.* 1:e0007313. doi: 10.1128/genomeA.00073-13
- Pyne, M. E., Bruder, M. R., Moo-Young, M., Chung, D. A., and Chou, C. P. (2016). Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in *Clostridium*. *Sci. Rep.* 6:25666. doi: 10.1038/srep25666
- Ratner, H. K., Sampson, T. R., and Weiss, D. S. (2015). I can see CRISPR now, even when phage are gone. *Curr. Opin. Infect. Dis.* 28, 267–274. doi: 10.1097/QCO.000000000000154
- Rodriguez, A. M., Callahan, J. E., Fawcett, P., Ge, X., Xu, P., and Kitten, T. (2011). Physiological and molecular characterization of genetic competence in *Streptococcus sanguinis*. *Mol. Oral Microbiol.* 26, 99–116. doi: 10.1111/j.2041-1014.2011.00606.x
- Sekulovic, O., Garneau, J. R., Néron, A., and Fortier, L. C. (2014). Characterization of temperate phages infecting *Clostridium difficile* isolates of human and animal origins. *Appl. Environ. Microbiol.* 80, 2555–2563. doi: 10.1128/AEM.00237-14
- Semenova, E., Jore, M. M., Datsenko, K. A., Semenova, A., Westra, E. R., Wanner, B., et al. (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. doi: 10.1073/pnas.1104144108
- Sesto, N., Touchon, M., Andrade, J. M., Kondo, J., Rocha, E. P. C., Arraiano, C. M., et al. (2014). A PNPase dependent CRISPR system in *Listeria*. *PLoS Genet.* 10:e1004065. doi: 10.1371/journal.pgen.1004065
- Shah, S. A., Erdmann, S., Mojica, F. J. M., and Garrett, R. A. (2013). Protospacer recognition motifs. *RNA Biol.* 10, 891–899. doi: 10.4161/rna.23764
- Shmakov, S. A., Sitnik, V., Makarova, K. S., Wolf, Y. I., Severinov, K. V., and Koonin, E. V. (2017). The CRISPR spacer space is dominated by sequences from species-specific mobilomes. *mBio* 8:e01397-17. doi: 10.1128/mBio.01397-17
- Soavelomandroso, A. P., Gaudin, F., Hoys, S., Nicolas, V., Vedantam, G., Janoir, C., et al. (2017). Biofilm structures in a mono-associated mouse model of *Clostridium difficile* infection. *Front. Microbiol.* 8:2086. doi: 10.3389/fmicb.2017.02086
- Sorek, R., Kunin, V., and Hugenholz, P. (2008). CRISPR — a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat. Rev. Microbiol.* 6, 181–186. doi: 10.1038/nrmicro1793
- Sorek, R., Lawrence, C. M., and Wiedenheft, B. (2013). CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annu. Rev. Biochem.* 82, 237–266. doi: 10.1146/annurev-biochem-072911-172315
- Soutourina, O. A., Monot, M., Boudry, P., Saujet, L., Pichon, C., Sismeiro, O., et al. (2013). Genome-wide identification of regulatory RNAs in the human pathogen *Clostridium difficile*. *PLoS Genet.* 9:e1003493. doi: 10.1371/journal.pgen.1003493
- Stachler, A.-E., and Marchfelder, A. (2016). Gene repression in Haloarchaea using the CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas 1-B system. *J. Biol. Chem.* 291, 15226–15242. doi: 10.1074/jbc.M116.724062
- Takeuchi, N., Wolf, Y. I., Makarova, K. S., and Koonin, E. V. (2012). Nature and intensity of selection pressure on CRISPR-associated genes. *J. Bacteriol.* 194, 1216–1225. doi: 10.1128/JB.06521-11
- Timme, R. E., Pettengill, J. B., Allard, M. W., Strain, E., Barrangou, R., Wehnes, C., et al. (2013). Phylogenetic diversity of the enteric pathogen *Salmonella enterica* subsp. *enterica* inferred from genome-wide reference-free SNP characters. *Genome Biol. Evol.* 5, 2109–2123. doi: 10.1093/gbe/evt159
- Wang, H., La Russa, M., and Qi, L. S. (2016). CRISPR/Cas9 in genome editing and beyond. *Annu. Rev. Biochem.* 85, 227–264. doi: 10.1146/annurev-biochem-060815-014607
- Wang, S., Hong, W., Dong, S., Zhang, Z.-T., Zhang, J., Wang, L., et al. (2018). Genome engineering of *Clostridium difficile* using the CRISPR-Cas9 system. *Clin. Microbiol. Infect.* doi: 10.1016/j.cmi.2018.03.026 [Epub ahead of print].
- Westra, E. R., Buckling, A., and Fineran, P. C. (2014). CRISPR-Cas systems: beyond adaptive immunity. *Nat. Rev. Microbiol.* 12, 317–326. doi: 10.1038/nrmicro3241
- Wiedenheft, B., van Duijn, E., Bultema, J. B., Waghmare, S. P., Zhou, K., Barendregt, A., et al. (2011). RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10092–10097. doi: 10.1073/pnas.1102716108
- Zegans, M. E., Wagner, J. C., Cady, K. C., Murphy, D. M., Hammond, J. H., and O'Toole, G. A. (2009). Interaction between bacteriophage DMS3 and host CRISPR region inhibits group behaviors of *Pseudomonas aeruginosa*. *J. Bacteriol.* 191, 210–219. doi: 10.1128/JB.00797-08
- Zhang, J., Zong, W., Hong, W., Zhang, Z.-T., and 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. doi: 10.1016/j.mben.2018.03.007

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Maikova, Severinov and Soutourina. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Annex. Article 3



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

AQ: au 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^fInstitute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris-Sud, Université Paris-Saclay, Gif-sur-Yvette, France^gWaksman 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

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-

FI

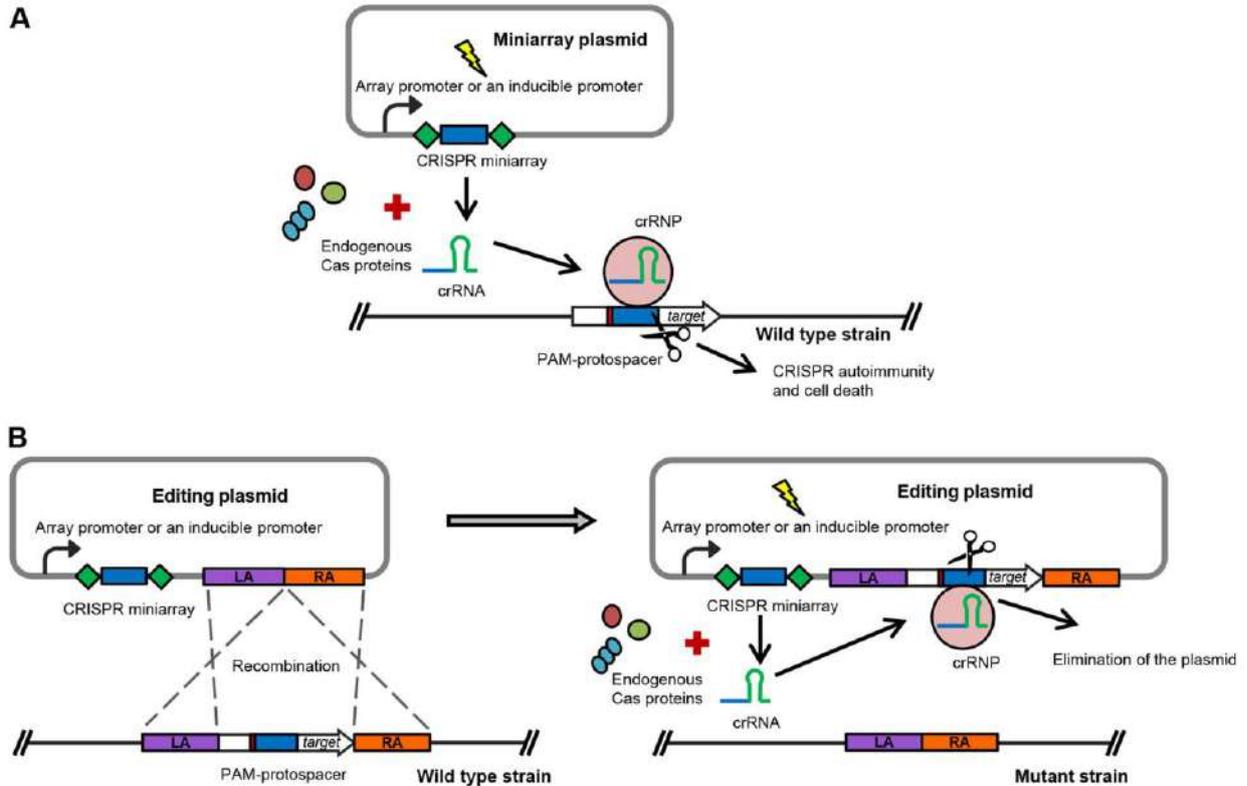


FIG 1 General scheme of using endogenous CRISPR-Cas systems for genome editing in bacteria and archaea. (A) The crRNA is expressed from a vector-borne CRISPR miniarray under the control of native or inducible promoters. The crRNA forms a ribonucleoprotein (crRNP) complex with endogenous Cas proteins, which recognizes and directs the cleavage of the PAM-associated protospacer, localized at the target chromosome 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,

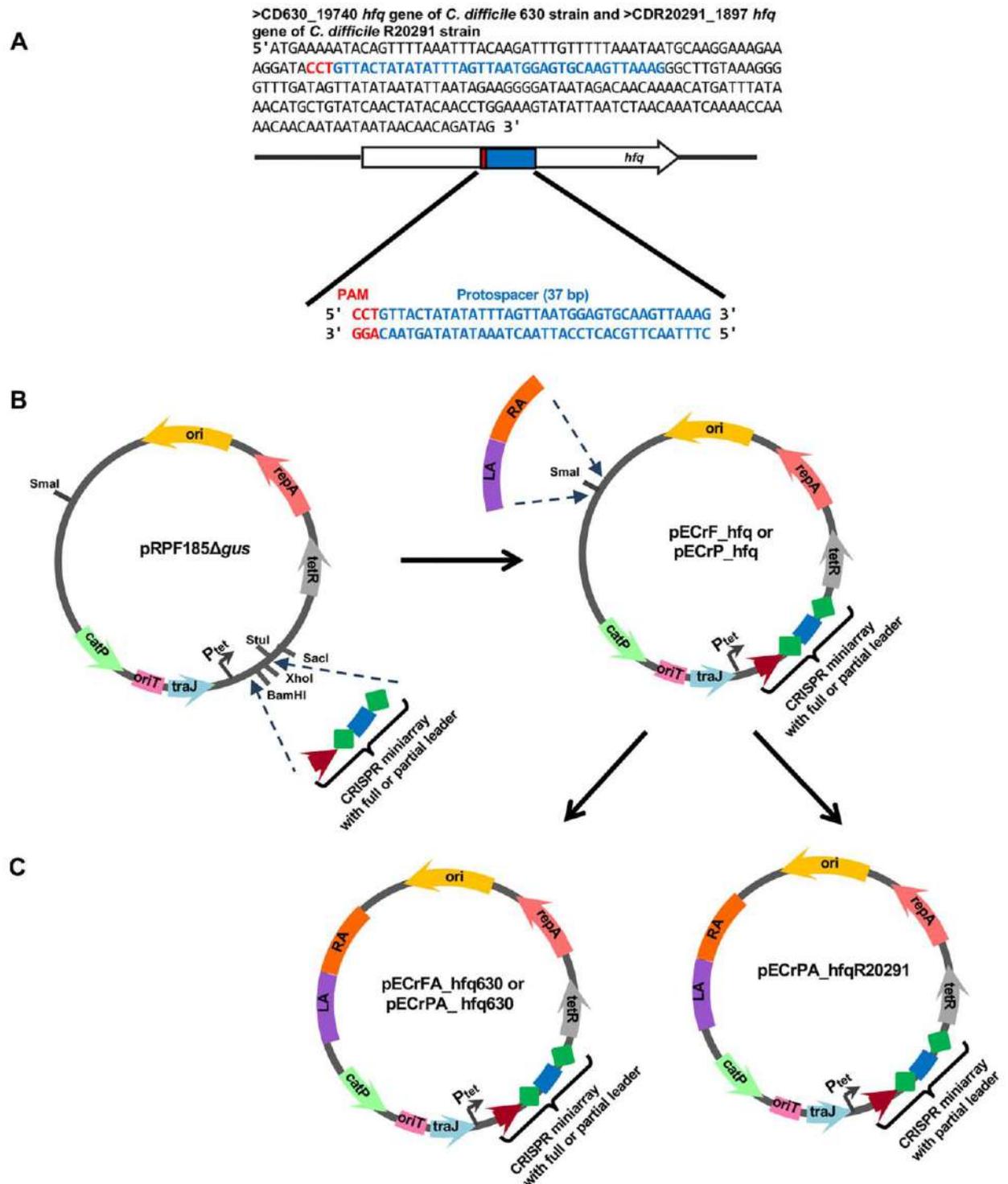


FIG 2 Strategy for the design of the editing plasmids to delete the *hfq* gene in the *C. difficile* 630 Δ erm and R20291 strains. (A) The coding sequences of the *C. difficile* 630 and R20291 *hfq* genes and a 37-bp sequence associated with the 5' CCT PAM, selected as a protospacer for the miniarray. (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)

COLOR

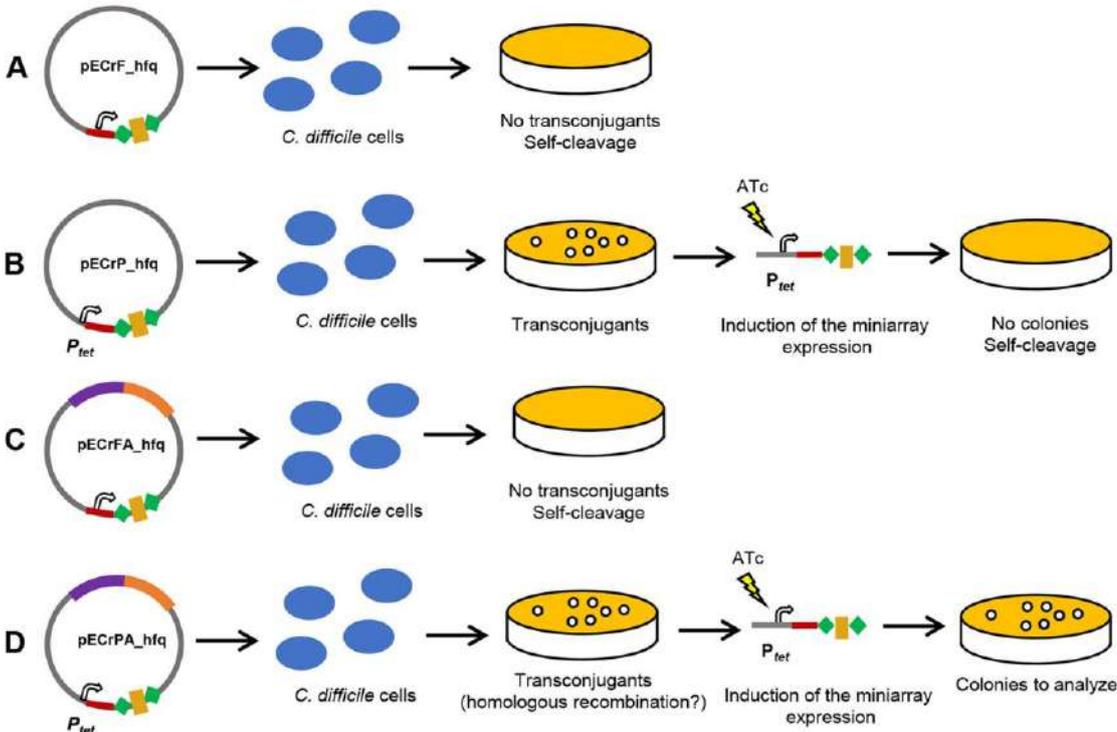


FIG 3 Different effects of the conjugation of the miniarray and editing plasmids into *C. difficile* cells. (A) CRISPR self-cleavage induced by immediate expression of the miniarray from the plasmid pECrF_hfq after conjugation. (B) CRISPR self-cleavage resulted from the ATc-induced expression of the 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 630Δerm strain (B and C), and the effects were tested and gene deletion was performed in both the 630Δerm 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* 630Δerm 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* 630Δerm, 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 pRPF185Δgus 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 630Δerm 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 630Δerm strain; in contrast, pECrPA_hfq630 and pECrPA_hfqR20291 were efficiently used for *hfq* gene deletion in the 630Δerm 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).

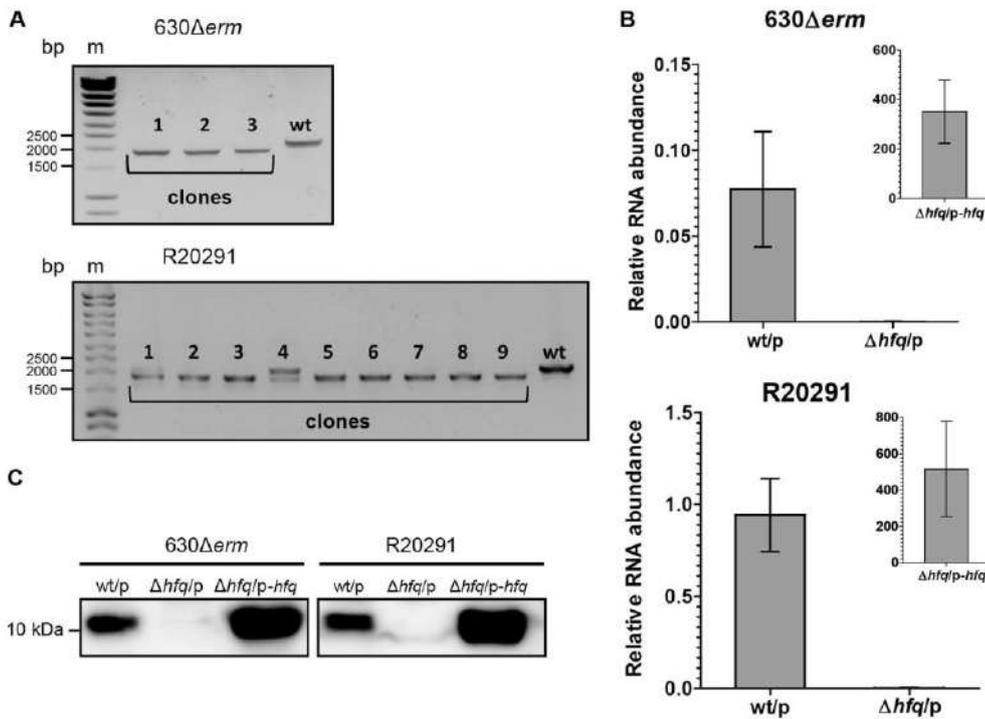


FIG 4 Validation of *hfq* deletion mutants. (A) PCR analysis of the *C. difficile* clones which lost the plasmid after genome editing. The 2,151-bp PCR bands correspond to the wild-type genotype; the 1,893-bp PCR bands correspond to the mutant genotype. For the R20291 strain, both the wild-type and mutant 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 Δ*hfq* mutant strains carrying an empty pRPF185Δ*gus* (the wt/p and Δ*hfq*/p strains, respectively) and the complemented Δ*hfq* *C. difficile* strain (the Δ*hfq*/p-*hfq* strain). mRNA levels are relative to those of 16S rRNA. (C) Western blot analysis of the wt/p, Δ*hfq*-p, and Δ*hfq*/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* Δ*hfq* 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, Δ*hfq*/p, and Δ*hfq*/p-*hfq* strains. After 24 h and 48 h in BHIS medium supplemented with Tm and ATc, the mutant strain (the Δ*hfq*/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 Δ*hfq*/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

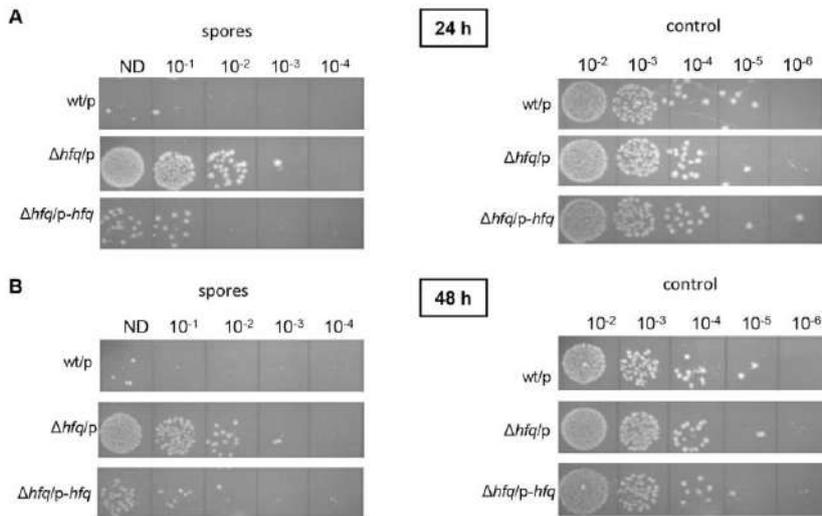


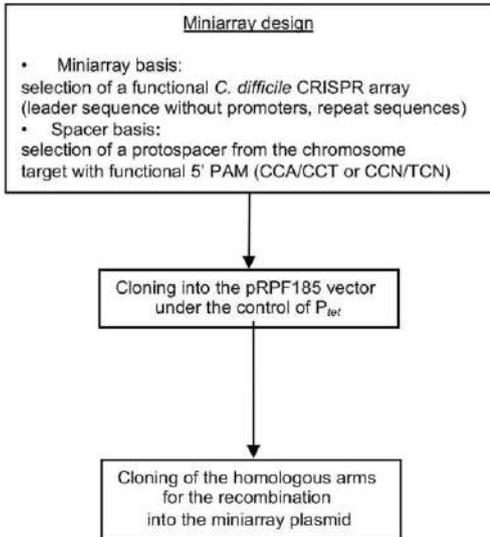
FIG 5 Sporulation levels in the *C. difficile* 630 Δerm wt/p, $\Delta hfq/p$, and $\Delta hfq/p-hfq$ strains (as numbers of spores) and the total amount of bacteria (as numbers of CFU) (control) after 24 h (A) and 48 h (B) of growth in BHI5 supplemented with Tm and ATc. The serial dilutions of the cultures 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* 630 Δerm 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* 630 Δerm and epidemic R20291 strains. The CRISPR repeats in the 630 Δerm 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 630 Δerm -based miniarray. Thus, the artificial miniarray designed from the *C. difficile* 630 Δerm 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,

I. Plasmid construction



II. Genome editing

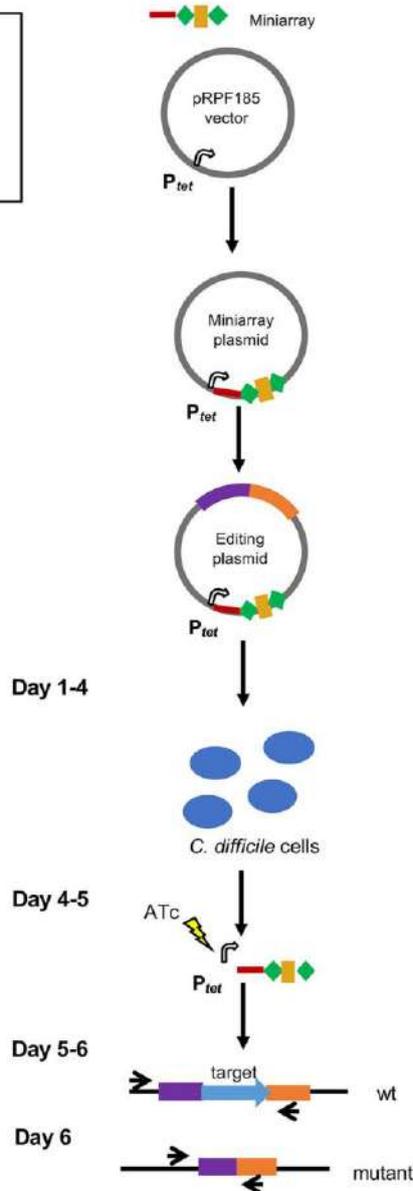
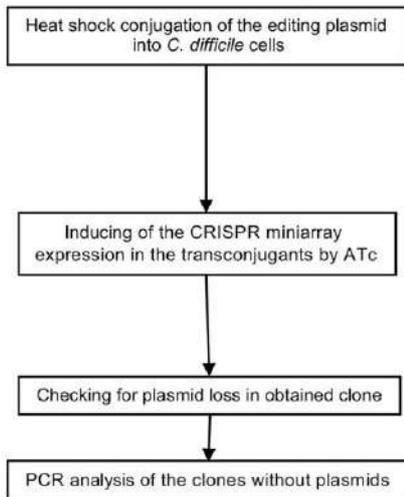


FIG 6 General work flow for application of endogenous CRISPR-Cas-based genome editing method in *C. difficile*.

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).

AQ: F

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.

AQ: G **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	$\Delta(\text{ara-leu})7697 \text{ araD139 fhuA } \Delta\text{lacX74 galK16 galE15 e14 mutant } \phi 80\text{dlacZ}\Delta\text{M15}$	New England Biolabs
HB101(RP4)	$\text{recA1 relA1 endA1 nupG rpsL (Str}^r\text{) rph spoT1 } \Delta(\text{mrr hsdRMS-mcrBC})$ $\text{supE44 aa14 galK2 lacY1 } \Delta(\text{gpt-proA})62 \text{ rpsL20 (Str}^r\text{) xyl-5 mtl-1 recA13}$ $\Delta(\text{mcrC-mrr}) \text{ hsdS}_{\text{B}}(\text{r}_{\text{B}}^- \text{ m}_{\text{B}}^-)$ RP4 (Tra ⁺ IncP Ap ^r Km ^r Tc ^r)	Laboratory stock
<i>C. difficile</i>		
630 Δ erm	Sequenced reference strain, Δ ermB	Laboratory stock (52)
R20291	PCR ribotype 027 epidemic strain	Laboratory stock
wt/p	630 Δ erm or R20291 carrying the pRPF Δ gus plasmid	This work
Δ hfq/p	630 Δ erm Δ hfq or R20291 Δ hfq carrying the pRPF Δ gus plasmid	This work
Δ hfq/p-hfq	630 Δ erm Δ hfq or R20291 Δ hfq carrying the p-hfq plasmid	This work
Plasmid		
pRPF185 Δ gus	pRPF185 Δ gus vector derivative	14, 47
pECrF_hfq	pRPF185 Δ gus carrying the hfq gene targeting the CRISPR miniarray with the full leader sequence	This work
pECrP_hfq	pRPF185 Δ gus carrying the hfq gene targeting the CRISPR miniarray with the partial leader sequence under the control of the P _{ter} promoter	This work
pECrFA_hfq630	pECrF_hfq carrying arms for recombination in the 630 Δ erm strain	This work
pECrPA_hfq630	pECrP_hfq carrying arms for recombination in the 630 Δ erm strain	This work
pECrPA_hfqR20291	pECrP_hfq carrying arms for recombination in the R20291 strain	This work
p-hfq	pRPF185 Δ gus carrying the hfq gene under the control of the P _{ter} 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_{ter} 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 Δ erm 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 Δ gus 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 Δ erm 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 Δ gus under the control of the ATc-inducible P_{ter} promoter, giving the p-hfq plasmid.

DNA sequencing was performed to verify the plasmid constructs. pRPF185 Δ gus 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 cefoxitin (8 μg/ml).

Deletion of the hfq gene and validation of Δ hfq mutants. To induce the expression of the CRISPR miniarrays under the control of the P_{ter} promoter, *C. difficile* transconjugants containing the pECrP_hfq, pECrPA_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	TAACAGATCTGAGCTCCAGGCCCTTCAATTATATGATAGGTTTTTTAATGACTACTAGCTGGTGTATATC	Full leader CR16 F
AM82	GTAAATCTAAAACCCCAAATAAACTTAGTATTTCCAATATCTACACATACAC	Leader CR16-R
AM83	TAACAGATCTGAGCTCCAGGCCCTCTGAGCAATATTTGCGATAAATGAAGTTTAAACAATTG	Partial leader CR16-F
AM91	GTTTTAGATTAACATATATGGAATGTAATGTTACTATATATTTAGTTAATGGAGTGCAAGTTAAAGGTT TTAGATTAACATATATGGAATGTAAT	<i>hfq</i> repeat-spacer-repeat motif
AM92	AGTTTTATTTGGGGTTTTAGATTAACATATATGGAATGTAATGTTACTATATATTTAGTTAATGGAGTG	Repeat-spacer-repeat F
AM93	TTTAAAGTTTTATTAACCTTATAGATTTACATCCATATAGTTAATCTAAAACCTTAACTTGAC	Repeat-spacer-repeat R
Construction of editing plasmids		
AM158	GAACACTTGCCGAAAAAGAAAACTGCCGGGTACGTACCCCGATATTGAAATAAAAAAGTTATTG	Left arm 630 and R20291 F
AM159	TCTTAAATTAATAATTATTAGATTTGTACCCCTCCAAAG	Left arm 630 and R20291 R
AM160	CTTGGGAGGGTACAAATCTAATAATTAATTTAATTAAGATGATTGAG	Right arm 630 and R20291 F
AM161	GAGCGAGGAAGCGGAAGCGCTCGCGGGGATCGATCCCGGAACAGGTTTTACATAAGAATC	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		
AQ: I	HFQ1: GAAGGCCTGGTAGGAATATTTTGAAGT HFQ2: GGGGATCCCATTAAGCATTATATCACCTGTC	5' <i>hfq</i> StuI 3' <i>hfq</i> BamHI
qRT-PCR		
AQ: J	QRTBD37: GGGAGACTTGAGTGCAGGAG QRTBD38: GTCCTCAGCGTCAGTTACA IMV447: AGGGCTTGTAAGGGGTTTG IMV448: TTGTTGTTTGGTTTGATTGTT	16S RNA F 16S RNA R qRT-PCR <i>hfq</i> F 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-0000000136) 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, Phillip S, Collyer 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 deficient 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, Cailliet J, Saujet L, Hajsnsdorf 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 (630Δerm) and demonstration that the conjugative transposon Tn916ΔE enters the genome of this strain at multiple sites. *J Med Microbiol* 54:137–141. <https://doi.org/10.1099/jmm.0.45790-0>.

AQ: M

