

Thesis, COLLÉGIALITÉ

Auteur : p211718

Promoteur(s) : Taminiau, Bernard

Faculté : Faculté de Médecine

Diplôme : Master en sciences biomédicales, à finalité approfondie

Année académique : 2022-2023

URI/URL : <http://hdl.handle.net/2268.2/17839>

Avertissement à l'attention des usagers :

Tous les documents placés en accès ouvert sur le site le site MatheO sont protégés par le droit d'auteur. Conformément aux principes énoncés par la "Budapest Open Access Initiative"(BOAI, 2002), l'utilisateur du site peut lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces documents, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale (ou prévue par la réglementation relative au droit d'auteur). Toute utilisation du document à des fins commerciales est strictement interdite.

Par ailleurs, l'utilisateur s'engage à respecter les droits moraux de l'auteur, principalement le droit à l'intégrité de l'oeuvre et le droit de paternité et ce dans toute utilisation que l'utilisateur entreprend. Ainsi, à titre d'exemple, lorsqu'il reproduira un document par extrait ou dans son intégralité, l'utilisateur citera de manière complète les sources telles que mentionnées ci-dessus. Toute utilisation non explicitement autorisée ci-avant (telle que par exemple, la modification du document ou son résumé) nécessite l'autorisation préalable et expresse des auteurs ou de leurs ayants droit.

ANALYSE TRANSCRIPTOMIQUE DU COMPORTEMENT DE *CLOSTRIDIODES DIFFICILE* EN PRÉSENCE D'UN VOISINAGE BACTÉRIEN



Berg Noémie

Promoteur: Bernard Taminiau

Co-Promoteur : Elisa Martinez

SBIM master 2 – 2022-2023 | noemie.berg@student.uliege.be

Remerciements :

J'adresse ma profonde gratitude à mes promoteurs, Dr Bernard Taminiau et Madame Elisa Martinez ainsi que le Pr Georges Daube pour leur guidance experte et leur soutien constant tout au long de cette recherche. Je tiens également à exprimer ma gratitude envers mes deux collègues stagiaires, Emma Bonnesire et Julien Hardy dont l'aide précieuse a grandement contribué à la réalisation de ce mémoire.

Pour continuer, je souhaite exprimer ma reconnaissance envers mes collègues de laboratoire et de bureau, Houyem, Pierre, Benjamin, Raphaëlle, Sébastien, Lara, Céline et Papa-Abdoulaye pour leur assistance et leur enthousiasme. Je remercie particulièrement Patricia pour ses conseils éclairés et ses encouragements.

Un merci à ma famille et amis pour leur amour, leur soutien et leurs encouragements constants. Leur présence a été une source de motivation et de réconfort tout au long de ce travail.

Résumé français :

Une majorité des cas de diarrhées nosocomiales infectieuses est due à *Clostridioides difficile*, une bactérie anaérobie, à Gram positif, sporulé, responsable de colites pseudomembraneuses. Cependant, par un phénomène encore peu compris, certaines personnes ne développent pas de symptômes malgré la présence de la bactérie dans leurs selles. Les facteurs de risques liés à l'infection sont premièrement la dysbiose qui permettrait une meilleure colonisation des intestins par *C. difficile*, celle-ci pourrait être due à la prise de médicaments, d'inhibiteur de protons, l'âge et une hospitalisation récente qui facilitent également la colonisation de *C. difficile* (Anjewierden et al. 2021). Ces observations ont permis d'émettre l'hypothèse que le microbiote pourrait exercer une influence sur le comportement de *C. difficile* (Martinez, Rodriguez, et al. 2022) par un mécanisme encore flou.

L'objectif de ce mémoire est de mettre en évidence un changement d'expression des gènes de *C. difficile* en présence ou non d'un voisinage bactérien à 12h par analyse transcriptomique. Pour ce faire, des expériences de croissance en modèle « transwell » ont été menées à différents temps (1 h, 6 h, 12 h et 24 h) durant lesquels *C. difficile* se retrouve soit sans contexte microbien, soit entouré d'un microbiote. L'ARN bactérien total a été extrait et l'ADNc a été séquencé. Les données transcriptomiques récoltées au point 12 h, en fin de croissance exponentielle, ont été analysées afin de mettre en évidence des changements au niveau de processus impliqués dans la virulence ou le métabolisme.

Le génome complet de notre souche a été séquencé et annoté. Les données de l'analyse transcriptomique ont été soumises à une analyse d'abondance différentielle et à un outil d'enrichissement de fonction créé spécifiquement. La comparaison pairée entre les conditions de croissance après 12 h a montré que 12,45% et 13,63% des gènes de *C. difficile* voient leur taux d'expression significativement varier en présence de deux microbiotes différents par rapport à la condition sans matière fécale. L'analyse d'enrichissement a montré que *C. difficile* modifie significativement la voie de Stickland et qu'elle surexprime certains processus liés à la virulence en présence d'un voisinage microbien. L'expression de certains gènes sélectionnés a été validée par RT-PCR et la technique du delta-delta Ct.

Abstract:

Most cases of nosocomial infectious diarrhea are due to *Clostridioides difficile*, an anaerobic, Gram-positive spore-forming bacterium responsible of pseudomembranous colitis. However, some people do not develop symptoms despite the presence of the bacterium in their feces. Risk factors associated with infection are firstly a dysbiosis that would allow a better gut colonization by *C. difficile*, moreover, this dysbiosis could be enhanced by drugs, proton inhibitors as well as age and recent hospitalization also enhance *C. difficile* colonization (Anjewierden et al. 2021). Those observations lead to the hypothesis that their microbiome could influence *Clostridioides difficile* behaviour by undetermined mechanism (Martinez, Rodriguez, et al. 2022).

The purpose of this master thesis is to highlight changes of *C. difficile* gene expression with and without bacterial neighbours at 12 hours by transcriptomic analysis. The experimental design was based on transwell insert experiments where total mRNA from *C. difficile* in different growth conditions (with and without microbiome) have been collected at different times (12 h and 24 h). The collected mRNA has been converted into cDNA before sequencing. Transcriptomic data at 12hours, corresponding to the end of the exponential growth, was analysed in order to highlight virulent and metabolic changes.

Our reference strain genome has been assembled and annotated. Transcriptomic results were then analysed through an in-house enrichment tool. A paired comparison between condition showed that 12.45% and 13.63% of *C. difficile* genes had a significant change of expression in presence of two different microbiome at 12 hours. By the enrichment analysis, the Stickland metabolism as well as virulent process showed a significant expression change in presence of bacterial neighbourhoods. Those results have been validated by RT-PCR with some selected genes by the delta-delta Ct to demonstrate a similar expression change trend with transcriptomic analysis.

Table des matières

I.	Introduction :	3
A.	Clostridioides difficile :	3
B.	Facteurs de risque :	4
1.	Propre au patient :	4
C.	Cycle d'infection par Clostridioides difficile :	6
1.	Entrée dans le tube digestif :	6
2.	Germination :	7
3.	Colonisation :	7
D.	Sporulation :	9
E.	Facteurs de virulence :	10
F.	Métabolisme :	12
G.	Microbiote intestinal et la dysbiose :	15
H.	Whole genome sequencing and transcriptomic analysis :	16
II.	Hypothèses, objectifs et stratégies :	18
III.	Matériel et méthodes :	20
A.	Analyse génomique de la référence :	20
1.	Extraction ADNg :	20
2.	Séquençage MinION :	20
3.	Assemblage :	20
4.	Annotation :	22
B.	Analyse transcriptomique :	23
1.	Extraction ARN total :	23
2.	Séquençage RNAseq par Illumina :	23
3.	Premières étapes de nettoyage de bibliothèques :	24
4.	Analyse transcriptomique :	24
5.	Analyse par enrichissement :	25

C.	Validation par qPCR :	26
1.	Choix des cibles et Design PCR :	26
2.	Extraction ADNg :	26
3.	Validation design :	27
4.	qPCR :	29
IV.	Résultats :	30
A.	Assemblage du génome :	30
B.	Analyse transcriptomique :	31
1.	Préambule :	31
2.	Traitements des librairies :	31
3.	Obtention des tables de comptages avec featureCounts :	31
4.	Analyse DESEQ2 :	32
5.	Comparaison du taux d'expression :	32
C.	Enrichissement par FeedMaurice :	33
1.	Métabolisme :	34
2.	Virulence :	36
D.	Validation par qPCR :	37
V.	Discussion :	38
VI.	Conclusion	44
VII.	Bibliographie :	45

I. Introduction :

A. Clostridioides difficile :

C. difficile est un Bacille bactérien, anaérobie, Gram-positive, capable de former des spores. Anciennement appelée *Clostridium difficile*, elle ne fait plus partie de la famille des *Clostridiaceae* en raison de sa récente reclassification dans la famille des *Bacillus* suite à l'analyse de son génome, la différenciant du genre *Clostridium* (Lawson et al. 2016).

Il s'agit d'une bactérie ubiquitaire retrouvée à la fois dans l'environnement naturel, mais également dans les selles de nombreux mammifères ainsi que dans les lieux d'hospitalisation (Rodriguez et al. 2019). L'infection par *C. difficile* (ICD) est définie comme étant la présence de la bactérie dans les selles de l'individu accompagnée soit d'une colite pseudomembraneuse, soit de la présence de toxines dans les selles. Alors que la colonisation asymptomatique est définie par la présence de *C. difficile* ou des toxines dans les selles sans symptôme de diarrhée (Martinez, Taminiau, et al. 2022).

C'est en 1935 qu'elle a été découverte pour la première fois dans l'intestin d'enfants (Hall et O'Toole, 1935). Chez ceux-ci, aucune manifestation clinique n'est observée malgré un taux de toxines important (Martinez, Taminiau, et al. 2022). A l'inverse, les adultes peuvent souffrir de douleurs abdominales accompagnées de diarrhées. Dans les cas plus graves, l'infection peut provoquer des colites pseudomembraneuses (Borriello 1998). *C. difficile* est reconnue, depuis les années 70, comme une des causes principales du développement de colites pseudomembraneuses. Elles se manifestent par l'apparition de nodules jaunâtres et la formation d'une plaque pseudomembraneuse qui peut engendrer des nécroses. L'apparition de la pseudomembrane fait suite à une infiltration des neutrophiles dans l'épithélium intestinal (Farooq et al. 2015).

L'infection par *Clostridioides difficile* est considérée comme une menace importante pour l'US Center for Disease Control. En effet, il s'agit d'une des causes majeures d'infection nosocomiales résistantes aux antibiotiques dans le monde (Gupta et Ananthakrishnan 2021). En 2016-2017, l'« European Centre for Disease Prevention and Control » rapporte 3,48 cas pour 10 000 patients par jour et le taux augmente aux États-Unis avec 10 cas pour 10 000 patients par jour (« Clostridioides (*Clostridium*) Difficile Infections - Annual Epidemiological Report for 2016–2017 » 2022). De manière générale, il existe deux lieux d'infection par *C. difficile*, les infections acquises à l'hôpital ou les infections acquises en communauté. A noter que les infections acquises en communauté correspondent à seulement 31,6% du taux des infections nosocomiales. En Europe, on rapporte que

58,8% des personnes infectées ont été en contact avec des lieux de soins dans les trois mois précédant l'infection, dont 81,3% ont été hospitalisées. De plus, l'infection par *C. difficile* peut induire des infections récurrentes dans 10 à 20% des cas d'infection (« Clostridioides (Clostridium) Difficile Infections - Annual Epidemiological Report for 2016–2017 » 2022). Les infections récurrentes sont définies par une rechute moins de 8 semaines post-infection (Gupta et Ananthakrishnan 2021). Les infections à *C. difficile* récurrentes sont associées avec deux fois plus de risques de complications et de décès que les infections primaires (« Clostridioides (Clostridium) Difficile Infections - Annual Epidemiological Report for 2016–2017 » 2022).

En plus d'être un risque sanitaire important, cela engendre également une charge économique considérable puisque les infections par *C. difficile* ont été dans le top 5 des infections les plus coûteuses de 1986 à 2013 (Gupta et Ananthakrishnan 2021).

B. Facteurs de risque :

1. Propre au patient :

Certains facteurs de risques à l'infection par *Clostridioides difficile* sont intrinsèques au patient.

Premièrement, le sexe de l'individu peut jouer un rôle dans la prévalence de l'infection. Le rapport de l'ECDC a rapporté qu'en Europe, les femmes sont généralement plus souvent infectées par *C. difficile* (58,8% des patients étaient des femmes) (« Clostridioides (Clostridium) Difficile Infections - Annual Epidemiological Report for 2016–2017 » 2022).

Pour continuer, l'âge du patient est un des principaux facteurs de risque. En effet, les enfants ne développent pas de maladie malgré le fait que l'on retrouve *C. difficile* dans 60 à 80% des cas chez les enfants de moins de 3 ans (Pahud et al. 2020). Une des hypothèses serait notamment dû à l'absence de récepteurs aux toxines de *C. difficile* (Schutze et al. 2013). En revanche, les personnes de plus de 65 ans sont 5 à 10 fois plus susceptibles de développer une infection à *C. difficile* (Czepiel et al. 2019). Ils représentaient 72% des cas en 2016-2017 (« Clostridioides (Clostridium) Difficile Infections - Annual Epidemiological Report for 2016–2017 » 2022).

Les patients immunodéprimés sont également sujets à un risque plus élevé d'ICD. Dans une étude palestinienne, 63% de leurs patients atteints étaient également immunodéprimés. Ceci fait suite à la prise d'immunosuppresseur lors d'une transplantation ou une immunodépression causée par un néoplasme malin (Abukhalil et al. 2021).

Enfin, l'infection à *C. difficile* est associée à plusieurs comorbidités. Entre autres, la maladie de Crohn et les chirurgies gastro-intestinales, mais aussi les maladies pulmonaires et rénales sont associées à l'ICD (Chitnis et al. 2013; Czepiel et al. 2019).

a) *Hospitalisation récente :*

L'hôpital est le lieu de contamination prédominant pour l'infection à *C. difficile* et le risque d'infection augmente avec le temps de l'hospitalisation. Au 1^{er} jour, le risque est de 2,1% jusqu'à 20%, ce chiffre augmente jusqu'à 45,4% après un mois d'hospitalisation. Cela serait notamment dû au fait que les hôpitaux sont un réservoir considérable pour les spores. Elles peuvent se retrouver dans les chambres des patients, sur le matériel médical ainsi que sur les mains des soignants (Weber et al. 2013).

b) *Prise d'antibiotiques :*

En outre l'âge du patient et l'hospitalisation récente, la prise d'antibiotique est liée de manière significative à l'infection par *C. difficile*. Le risque d'infection en communauté est 3 fois plus important chez les patients ayant reçu des antibiotiques. Ce constat est également présent pour les infections acquises à l'hôpital, où le risque augmente de 12,8% à chaque prise d'antibiotique. L'effet plus prononcé de certains antibiotiques par rapport à d'autres comme les fluoroquinolones et la clindamycine a pu être mis en évidence (Brown et al. 2013; Webb et al. 2020). Cela est dû à leur effet modulateur sur le microbiote intestinal. En effet, la perturbation du microbiote par les antibiotiques permet un environnement propice à la colonisation des intestins par *C. difficile* (Lewis et al. 2015).

c) *Diminution de l'acidité gastrique :*

L'acidité du tube digestif est également un acteur influençant l'infection par *C. difficile*. Une étude de cohorte et de case-control a permis de conclure que la prise d'inhibiteur de proton (IPP) serait responsable d'une augmentation du facteur de risque par 3 suivant les 90 jours précédant l'infection (Dial et al. 2005). D'Silva et al (2021) ont montré que le risque d'ICD récurrente est 69% plus élevé chez les personnes ayant reçues des IPP (D'Silva et al. 2021).

La fonction de ces IPP est de diminuer l'acidité dans le tube digestif au niveau des pompes à protons. Or, l'acidité gastrique est reconnue comme une barrière de protection face aux infections. Même si le mécanisme reste encore inconnu, plusieurs hypothèses existent quant au rôle de

l'inhibition de l'acidité : mauvaise stérilisation des aliments ingérés, diminution de la diversité microbienne, environnement permissif à la prolifération des spores et à la colonisation de celles-ci sous forme végétative (Dial et al. 2005; D'Silva et al. 2021). De plus, ce type de traitement est également reconnu comme un facteur de risque pour d'autres infections entériques (Dial et al. 2005).

La diminution de l'acidité est corrélée avec la diminution des acides biliaires secondaires. Ces sels biliaires secondaires permettent d'inhiber la germination de *C. difficile*. En cas d'un traitement PPI, il n'y a pas formation de ces acides secondaires. Il existe également d'autres facteurs ayant pour conséquence une baisse des acides secondaires comme l'obésité ainsi que le type de régime alimentaire qui influence à la fois la diminution d'acides secondaires, mais également le pH de l'intestin (Baloh et Sorg 2022).

C. Cycle d'infection par Clostridioides difficile :

1. Entrée dans le tube digestif :

Clostridioides difficile est transmis par voie féco-orale à partir des spores résistantes à l'oxygène et aux dommages de l'environnement.

Les spores de *Clostridioides difficile* ont une conformation classique. Elles contiennent en leur centre un cœur déshydraté rempli de calcium dipicolinic (Ca^{2+} DPA) qui participe à la dormance de la spore par déshydratation. Afin de garantir l'intégrité de l'ADN, il est hyperenroulé à l'aide de protéines acidosolubles qui le protègent et empêchent la translation (Nerber et Sorg 2021; Donnelly, Fimlaid, et Shen 2016)

Le cœur de la spore est protégé des toxines et des dommages chimiques par une membrane interne. C'est elle qui formera la membrane de la future cellule végétative (Kochan et al. 2018) (figure 1). Cette membrane est elle-même entourée par le cortex formé de peptidoglycanes modifiés (Meador-Parton et Popham 2000). Cela permet aux enzymes lytiques du cortex CLE de le lyser de façon spécifique lors de la germination.

La couche la plus externe de la spore (figure 1), l'exosporium, va permettre l'adhérence au tissu solide avant de démarrer la germination. Cela fait suite à l'expression du second messager la diguanosine monophosphate cyclique (C-di-GMP), également appelée diguanylate cyclique (Baloh et Sorg 2022).

2. Germination :

Pour démarrer la germination, la spore doit recevoir un signal de germination, le germinant. Ce signal va interagir avec un récepteur qui enclenche une cascade de réaction dans le but de réhydrater la spore (Sorg et Sonenshein 2008).

Chez *C. difficile*, le glycocholate va être clivé en glycine et cholate. Ils jouent le rôle de germinant en enclenchant la germination grâce à leur fixation sur un récepteur pseudoprotéase CspAB clivé en CspA et CspB (Sorg et Sonenshein 2008). Ce clivage catalyse une cascade d'activation qui commence par l'activation de CspC et puis de CspB pour enclencher in fine l'hydrolyse du cortex par activation du propeptide SleC (Kochan et al. 2018).

En parallèle de la dégradation du cortex, le Ca²⁺-DPA est libéré en échange d'eau, ce qui va permettre la réhydratation de la cellule (Baloh et Sorg 2022). C'est grâce à la protéine mécanosensible SpoVAC que le DPA est échangé avec de l'eau. Il s'agit d'un canal qui permet lors de la sporulation de concentrer le DPA mais également de le libérer à la suite du signal de dégradation du cortex. Après la réhydratation du cœur, le métabolisme de *C. difficile* se réveille et permet la transformation en cellule végétative capable de coloniser l'intestin (Baloh et Sorg 2022).

3. Colonisation :

Une fois la germination effectuée, *C. difficile* va pouvoir se multiplier au sein de l'intestin avant d'atteindre la phase stationnaire. Cette prolifération va engendrer la production de toxines responsables de l'apparition des symptômes cliniques suite aux dommages sur la barrière épithéliale et l'activation du système immunitaire. Ces mécanismes peuvent ainsi provoquer la mort cellulaire et la formation d'une pseudomembrane.

a) Production de toxines :

Clostridioides difficile exprime principalement trois toxines : TcdA et TcdB qui sont identiques sur une grande partie de leur séquence dans le locus Paloc mais également CDT, la toxine binaire.

L'expression des toxines TcdA et TcdB est régulée sur le locus Paloc par TcdR. Il s'agit d'une protéine sigma agissant directement sur les promoteurs de l'opéron et dont l'expression augmente avec le manque de nutriments (Ransom et al. 2018). A l'inverse, ces deux toxines sont régulées négativement par TcdC par déstabilisation de TcdR (Matamouros, England, et Dupuy 2007).

Le mécanisme de ces toxines se déroule en quatre étapes en commençant par la liaison à un récepteur membranaire responsable de l'endocytose de la toxine. Ensuite, l'endosome va s'acidifier provoquant la liaison des toxines à sa membrane pour permettre aux domaines GT et AP d'être transloqués dans le cytosol. *In fine*, les GTPases de l'hôte seront glycosylées par fixation des toxines.

L'inactivation des GTPases induit des dommages considérables à la cellule hôte menant jusqu'à la mort cellulaire (Chan et al. 2018). En effet, les GTPases jouent un rôle important dans le cycle cellulaire, leur inhibition est donc corrélée avec un arrêt du cycle et donc à la mort cellulaire. De plus, TcdA et TcdB induisent un changement du cytosquelette. C'est suite à cette perturbation que les cellules du tube digestif vont perdre les jonctions serrées augmentant ainsi la perméabilité de la barrière épithéliale (Hecht et al. 1988). Cela a pour conséquence d'induire des diarrhées, mais également de permettre le passage des toxines. Celles-ci sont responsables d'activation des chémokines responsables de l'activation du système immunitaire.

La toxine CDT est une « ADP-ribosylating clostridial toxins ». Elle entraîne une élongation des microtubules en empêchant la polymérisation de l'actine. Ce mécanisme va garantir une meilleure adhésion de *C. difficile* aux cellules épithéliales. Pour ce faire, CDTb va se lier avec un récepteur LSR des cellules de l'hôte pour former un pore. Par la suite, la fixation de CDTa permet l'endocytose de ce pore à la membrane d'un endosome. Après acidification, CDTa est libéré dans le cytosol pour se fixer à l'actine et inhiber sa polymérisation (Kordus, Thomas, et Lacy 2022). Pour terminer, comme TcdA et TcdB, CDT va également activer le système immunitaire en activant le facteur NF-Kb. Pour ce faire, CDT se fixe sur un récepteur « toll-like » qui permet l'activation de l'inflammasome et la cascade d'activation NF-Kb (Simpson et al. 2022).

b) Activation du système immunitaire :

Le système immunitaire est activé en 1^{er} suite à la reconnaissance de quatre PAMPs dont les toxines TcdA et TcdB par activation de l'inflammasome NLRP3.

Pour continuer, les toll-like receptor 5 et 4 (TLR) vont être respectivement activés par la flagelline et la protéine S-layer A. La protéine MyD88, un adaptateur au TLR va ensuite médier le recrutement des neutrophiles par la production de chémokines CXCL1. En outre, CXCL1 est également médié par l'activation des « nucleotide-binding oligomerization domain 1 » (NOD1) par reconnaissance de fragment de peptidoglycanes.

Ces cascades d'activation se traduisent toutes par une activation de la voie NF-Kb. C'est par cette voie que les cytokines pro-inflammatoires et de chémokines sont produites afin de recruter les cellules immunitaires au site d'infection (Cowardin et Petri 2014).

En outre, les toxines activent non-seulement l'inflammasome mais elles activent également l'apoptose des cellules épithéliales. Ces effets ont pour conséquence la perturbation de la barrière épithéliale et augmentation de l'activation de la réponse inflammatoire par la translocation des bactéries ainsi que la mort cellulaire des cellules épithéliales induite par la cascade P2Y6 et IL-8. Ces mécanismes sont responsables de la nécrose des cellules et la formation de la pseudomembrane (Linevsky et al. 1997; Cowardin et Petri 2014; Nibbering et al. 2021).

D. Sporulation :

Puisque *C. difficile* est une bactérie anaérobie, elle a dû trouver une stratégie pour survivre dans un environnement aérobie. Pour cela, *C. difficile* forme des spores aérotoles et dormantes pour survivre en dehors de l'hôte en absence de nutriment (Shen et al. 2019).

La sporulation se déroule en plusieurs étapes (figure 1) initiées par le niveau de phosphorylation de SpoOA, un régulateur de la transcription des gènes liés à la sporulation. Le taux de phosphorylation de SpoOA dépend des signaux perçus dans l'environnement notamment en fonction du taux de nutriments (DiCandia et al. 2022).

La sporulation débute par une division polaire de la cellule mère en deux cellules identiques génétiquement, mais de tailles différentes. La cellule mère va ensuite englober la pré-spore plus petite. La pré-spore sera alors entourée de deux membranes avant la formation du cortex (formé de peptidoglycane modifiés) situé entre les deux membranes. Ce cortex permet de survivre face à la chaleur et à l'éthanol. La membrane externe est également entourée d'une couche de protéine, le manteau. Cette couche permet la protection du génome. Par ailleurs, la membrane externe est également recouverte de l'exosporium, une couche externe typique de *C. difficile* (Zhu, Sorg, et Sun 2018; Shen et al. 2019).

En parallèle, la cellule va produire l'acide dipicolinique avec du Ca-DPA pour déshydrater la cellule par un transporteur de Ca-DPA. L'ADN sera également protégé et enroulé par des SASps. La sporulation se termine par la lyse de la cellule mère pour libérer la spore (Kochan et al. 2018).

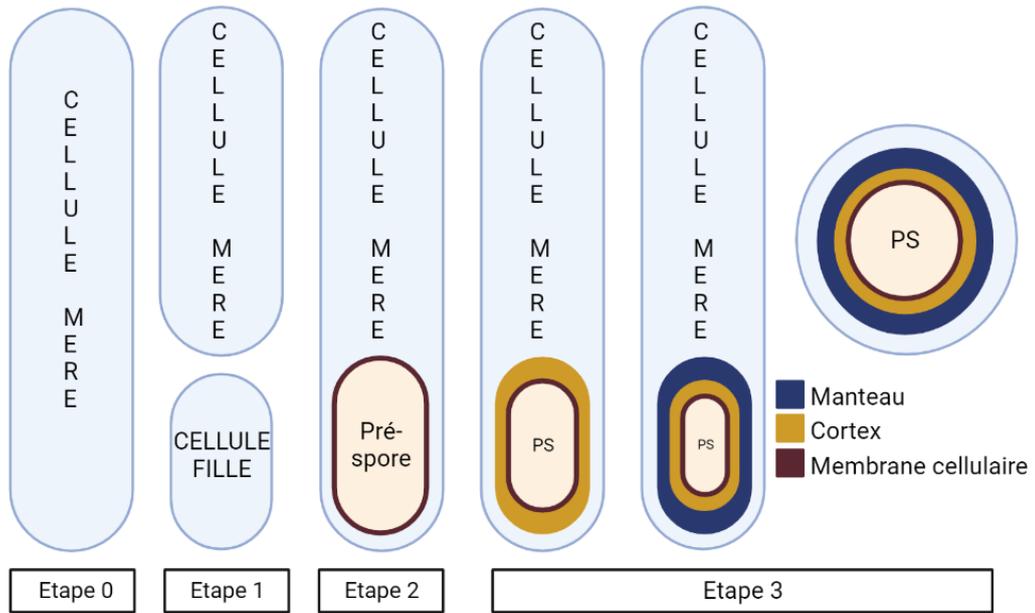


Figure 1 mécanisme de sporulation ainsi que les différentes couches d'une spore – inspiré de (Zhu, Sorg, et Sun 2018) -créé sur biorender.com

La capacité de produire des spores fait partie de la virulence de *C. difficile*. Effectivement, les spores aérotolérantes permettent la survie et la protection de son génome dans des environnements hostiles. Les spores contribuent notamment à la récurrence des infections, mais aussi au caractère ubiquitaire de la bactérie dans l'environnement (Shen et al. 2019).

E. Facteurs de virulence :

En plus de la production de toxines et sa capacité à former des spores, *C. difficile* possède plusieurs autres facteurs de virulence.

Pour commencer, *C. difficile* produit des protéines de surface ubiquitaires au genre. Ensembles, elles constituent la couche-S par assemblage de deux sous-unités. Pour se former, la protéine SlpA doit être clivée avant de pouvoir assembler les deux sous-unités ensemble. SlpA est utile dans l'adhérence lors de la colonisation, mais également pour l'activation de l'immunité. Effectivement, la bactérie peut se fixer aux parois de l'épithélium intestinal (Calabi et al. 2002). Les protéines Slp peuvent se lier de façons spécifiques à la paroi intestinale de manière morphologique et biochimique par liaison à un polypeptide spécifique de l'intestin. En outre, l'adhésion à la cellule de l'hôte va activer une cascade de réaction qui se termine par un réarrangement du cytosquelette comme c'est le cas pour les toxines. La couche-S permettrait également de diminuer la diffusion des toxines suite à la fixation de la bactérie à la bordure en brosse de l'intestin (Calabi et al. 2002). De plus, en absence de cette couche-S, la production de toxines diminue rendant les souches non virulentes (Kirk et al. 2017).

La membrane de *C. difficile* est composée d'autres facteurs comme des protéines de membranes cellulaires (Cwp) et des polysaccharides. Plusieurs protéines composent cette famille, entre autres Cwp84 et Cwp66. Les Cwp jouent un rôle fondamental dans la virulence, malgré leur propriétés immunoréactives (Wright et al. 2008). En effet, Cwp84 est un facteur important lors de la maturation de la couche-S. Les mutants « Knock-out » ne sont pas capables de cliver la protéine SlpA empêchant alors la formation d'une couche-S mature (Kirby et al. 2009). En outre, il a été montré que Cwp84 augmente l'adhésion à la paroi intestinale en présence d'antibiotiques. Cwp84 participe également à la propagation de l'infection par son activité de protéase par son action sur les composants de la lame basale (Denève et al. 2008). Concernant Cwp66, il s'agit d'une protéine choc thermique que l'on retrouve à la surface de *C. difficile*. Elle est également impliquée dans la virulence pour son rôle dans l'adhésion. C'est après le clivage d'une séquence hydrophobe et son transport à la surface de la bactérie qu'elle exerce son rôle d'une adhésine (Waligora et al. 2001).

Pour continuer, *C. difficile* possède des protéines capables de se lier à la fibronectine (Fbp). Ces protéines offrent un avantage à *C. difficile* puisque la fibronectine est retrouvée dans la matrice des tissus, mais également dans les fluides corporels. Les Fbp sont des adhésines que nous retrouvons chez de nombreuses bactéries pathogènes telles que *Streptococci* et *Staphylococci*. Chez *C. difficile*, la Fbp68 est la seule Fbp exprimée. Son gène possède une séquence hautement conservée. Par ailleurs, la Fbp68 a montré sa capacité à se fixer autant sur la fibronectine des tissus que sur la fibronectine libre. Cette force permet à *C. difficile* de se disséminer dans l'intestin, mais également de pouvoir échapper au système immunitaire (Hennequin et al. 2003).

En outre la capacité d'adhérence, la motilité de *C. difficile* est également un facteur de virulence important pour l'invasion et la colonisation des intestins. Pour ce faire, elle possède des flagelles formées de plusieurs protéines comme FliC et FliD. Il semblerait également que les flagelles ne jouent pas seulement un rôle dans la motilité, mais aussi dans la formation du biofilm (Awad et al. 2014).

Dans la même idée que les spores, *C. difficile* peut être protégée de l'environnement extérieur par un biofilm. Le biofilm est défini comme une matrice composée principalement de polysaccharides intrinsèquement liés avec un ensemble de bactéries devenues sessiles (Donlan 2002; Taggart et al. 2021). Un des avantages du biofilm est la formation d'un micro-environnement permettant les échanges de nutriments par des réactions rédox, mais aussi l'activation du « quorum sensing » suite au contact cellule-cellule (Donlan 2002).

La formation du biofilm participe à la virulence de nombreuses bactéries. En effet, il offre de nombreux avantages de survie telle que la formation d'une barrière contre le stress comme l'exposition à l'oxygène et d'autres toxines, mais il permet aussi d'éviter le système immunitaire ainsi que l'effet des antibiotiques tout en offrant un lieu de concentration en nutriments. Il pourrait également servir de réservoir dans les cas d'infections récurrentes à *C. difficile* (Dawson et al. 2012; Dapa et al. 2013).

Chez *C. difficile*, le biofilm se développe après 16 à 24 h de sédimentation par activation de plusieurs facteurs dont le c-di-GMP et Spo0A, deux régulateurs centraux (Tremblay et Dupuy 2022). Spo0A est également un régulateur de la sporulation lorsque celui-ci est phosphorylé, cependant, à un taux intermédiaire de phosphorylation, ce gène entraîne la formation du biofilm (Dawson et al. 2012; Tremblay et Dupuy 2022). Quant à c-di-GMP, il régule de nombreux gènes impliqués dans l'agrégation des cellules comme les flagelles dont l'expression est diminuée pour rendre la cellule immobile et économiser son énergie. Il régule également les protéines de surface comme Cwp84 qui

joue un rôle dans la dégradation de la matrice favorisant la formation du biofilm (Buddle et Fagan, s. d.; Taggart et al. 2021).

Afin de pouvoir se repérer dans l'environnement par rapport à la densité cellulaire, les bactéries utilisent le « quorum sensing ». Il s'agit d'un système de communication cellule-cellule permettant aux bactéries de modifier leur comportement. Pour ce faire, les bactéries produisent des « auto-inducers » (AI). En fonction de la concentration bactérienne, les AI seront être plus ou moins dilués dans l'environnement. Lorsque les « auto-inducers » atteignent une concentration seuil, ils seront détectés par des récepteurs membranaires responsables de l'activation cascade de réaction régulant l'expression de certains gènes et changeant le comportement de la bactérie (Rutherford et Bassler 2012).

C. difficile possède également des gènes lui permettant de répondre au stress oxydatif mais également lorsqu'il est dépourvu de fer ou de zinc. C'est dans l'environnement que *C. difficile* trouve la spermidine. Pour ce faire *C. difficile* possède des protéines de transports spécifique à l'entrée des polyamines. En effet, *C. difficile* possède des transporteurs de polyamines comme la spermidine et la putriscine. Les polyamines sont des molécules polycationiques (plusieurs charge positive) composées de groupes aminés. La plupart des bactéries entériques utilisent ces molécules lors de leur croissance. Elles jouent un rôle dans la réponse au stress oxydatif et lors de limitation en Zinc et en nutriments. Ces gènes offrent à *C. difficile* une grande capacité d'adaptation à son environnement (Shah et Swiatlo 2008).

Pour terminer, *C. difficile* est capable de détenir des résistances aux antibiotiques. C'est ainsi le cas pour la ciproflaxine et la famille des fluoroquinolones en général. Ces gènes de résistances aux antibiotiques permettent ainsi la récurrence et la survie de *Clostridioides difficile* dans un environnement hostile (Buddle et Fagan, s. d.).

F. Métabolisme :

C. difficile est une bactérie anaérobie, elle produit de l'énergie en l'absence d'oxygène. De plus, elle doit être capable d'utiliser différents nutriments qu'elle est capable de trouver dans l'intestin. Pour ce faire, elle doit utiliser différents chemins métaboliques spécifiques à ce type de bactérie. Une respiration anaérobie et des processus de fermentations permettent de produire de l'énergie sous forme d'ATP et de gradient d'électron en l'absence d'O₂. Le métabolisme de fermentation est composé de différents chemins métaboliques notamment Stickland et Wood-

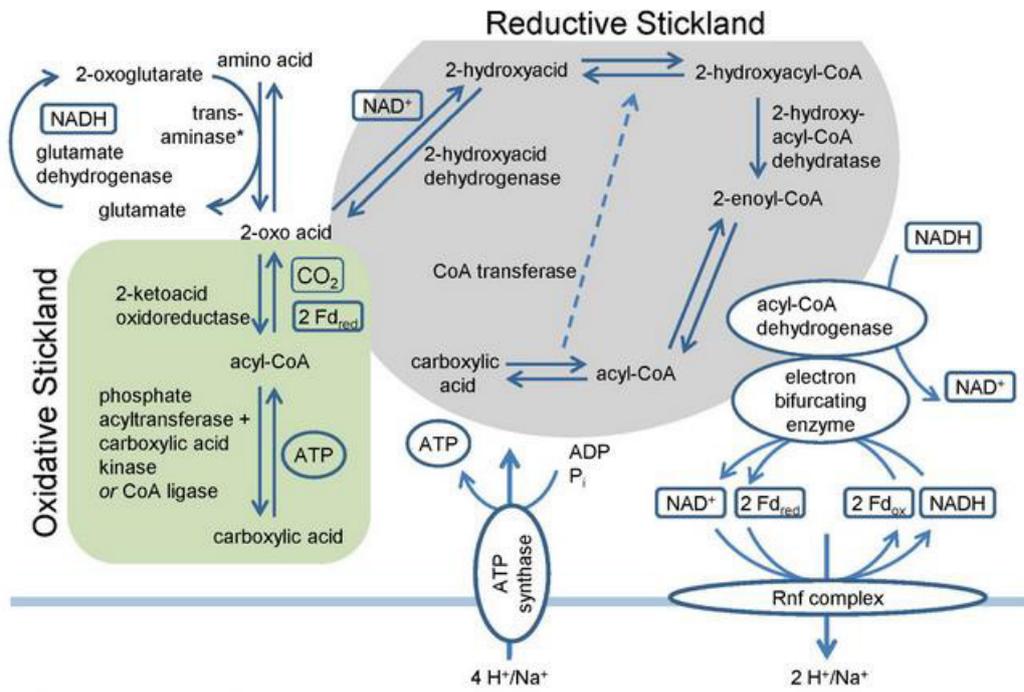


Figure 1 Métabolisme de Stickland - (Neumann-Schaal, Jahn, et Schmidt-Hohagen 2019)

Ljungdahl associés à la respiration anaérobique par le métabolisme central du carbone (Neumann-Schaal, Jahn, et Schmidt-Hohagen 2019).

Le métabolisme de Stickland est basé sur une réaction couplée d'oxydo-réduction de deux acides aminés (Stickland 1934) (figure 2). Les acides aminés ont chacun un potentiel réducteur ou oxydatif qui détermine leur fonction dans la voie de Stickland. Ce chemin métabolique est divisé en 3 étapes pour l'oxydation et la réduction des acides aminés. Il commence par une étape commune où les deux acides aminés sont transformés en 2-oxo-acide. Cette étape va notamment permettre la formation de NADH. Le NADH est ensuite couplé avec une réaction de glutamate déshydrogénase. Ce couplage a pour but de produire du 2-oxoglutarate précurseur du métabolisme du glutamate. En parallèle, les 2-oxo-acides seront soit oxydés soit réduits en fonction de leur potentiel rédox. L'oxydation commence par la formation de thioester CoA avec l'aide de ferrédoxine. Cela engendre une libération de CO₂ et la formation d'ATP suite au clivage du thioester CoA. Pour la réduction, le 2-oxo-acide est réduit avec du NADH pour former du 2-hydroxyacide. Successivement, l'enoyle-CoA est déshydraté pour former de l'acyl-coa par une réaction de réduction catalysée par une CoA-transférase du complexe de bifurcation d'électron (Stickland 1934; Neumann-Schaal, Jahn, et Schmidt-Hohagen 2019; Marshall et al. 2023).

Il s'agit d'un complexe qui va permettre de produire un gradient de sodium/proton à partir de la fermentation de Stickland et du complexe RNF. Cette association engendre l'oxydation de NADH et de ses deux électrons grâce à l'enoyle-CoA venant de Stickland et la ferrédoxine du complexe RNF. Cela va permettre la formation d'un potentiel rédox et d'un gradient de sodium/proton au travers de la membrane. Ce gradient va servir *in fine* à la formation d'ATP par le passage de proton au travers d'une ATP synthase (Neumann-Schaal, Jahn, et Schmidt-Hohagen 2019).

Pour continuer, *C. difficile* est également capable d'utiliser des carbohydrates comme source d'énergie en passant par le métabolisme central du carbone (figure 3). Il comprend la glycolyse, la glycogénèse mais également le cycle de Krebs et permet de fermenter les carbohydrates comme le glucose, le fructose et en faible quantité, le mannose. Il existe un 2^e chemin métabolique spécifique aux bactéries anaérobies, le chemin métabolique de Wood-Ljungdahl (WLP) faisant également partie du métabolisme central du carbone. *C. difficile* est une des seules bactéries pathogènes à posséder des gènes de ce métabolisme. Ce chemin métabolique utilise 2 CO₂ comme accepteur final d'électron pour produire du NADH (Neumann-Schaal, Jahn, et Schmidt-Hohagen 2019). *C. difficile* est une des seules bactéries pathogènes à posséder des gènes de ce métabolisme. Le WLP offre un avantage de croissance à *C. difficile* grâce à son activation lorsque la bactérie ne trouve plus de proline et de leucine

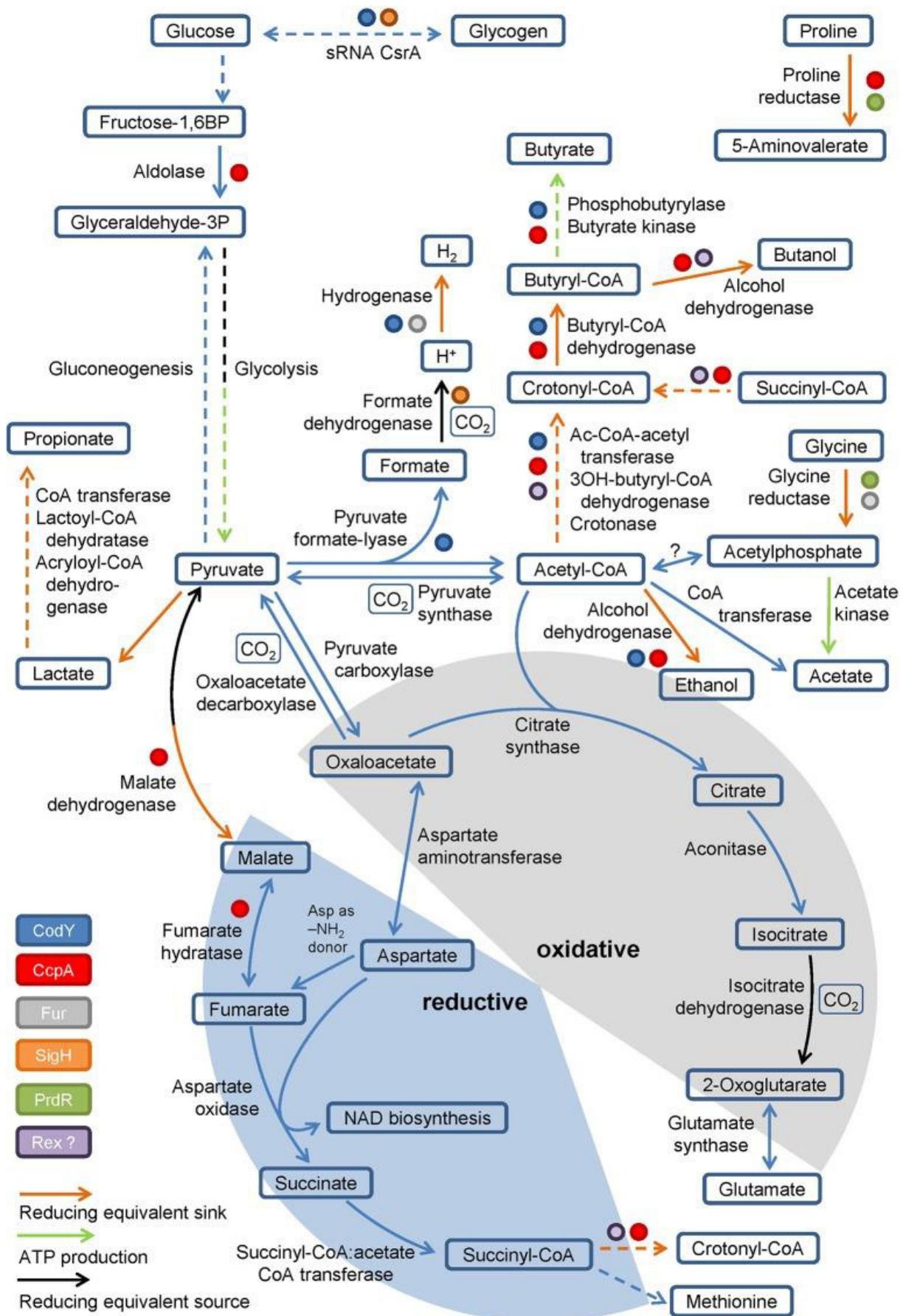


Figure 1 Métabolisme de *C. difficile* -(Neumann-Schaal, Jahn, et Schmidt-Hohagen 2019)

dans son milieu. En effet, c'est à partir de 2 molécules de CO₂ et d'H₂ que la bactérie peut former du NAD⁺ par oxydation suite à la formation notamment de butyrate, mais aussi d'éthanol et d'acétate (Gencic et Grahame 2020). Ce 2^e chemin métabolique offre une grande flexibilité métabolique qui serait responsable de la récurrence de *C. difficile* grâce à la possibilité d'utiliser d'autres sources de nutriments que ceux consommés par les bactéries compétitrices. Cela permet à *C. difficile* de fonctionner de façon autotrophe et de soutenir la fermentation d'une grande variété de carbohydrates (Köpke, Straub, et Dürre 2013).

En outre, *C. difficile* possède plusieurs gènes impliqués dans le transport des carbohydrates notamment les systèmes de phosphotransférases dépendantes du phosphoenolpyruvate (PTS). Il s'agit d'un transporteur qui lors du transport, catabolise la conversion des carbohydrates en phosphoesters qui pourront être utilisés dans le métabolisme central. A partir de la glycolyse et de la glycogénèse, *C. difficile* va réaliser la formation de pyruvate et de l'acétyl-Coenzyme A par dégradation d'acides aminés et de carbohydrates. Le pyruvate est un élément essentiel dans le métabolisme central, il est notamment utilisé dans le cycle de Krebs, mais également par la pyruvate-formate-lyase. Cette dernière va permettre la formation d'acétyl-CoA et de formate à l'aide de la formate déshydrogénase, est oxydé en CO₂ et en un électron (Marshall et al. 2023).

Par ailleurs, le métabolisme de *C. difficile* dépend de plusieurs ferrédoxines et d'enzyme Fe-S notamment dans le métabolisme central. En cas de déficience en fer, le métabolisme de *C. difficile* change. L'expression des ferrédoxines est fortement réduite au profit de flavodoxine. Lors d'une limitation en fer, les gènes du complexe RNF et de la proline sont également augmentés. Ce changement métabolique démontre la capacité d'adaptation de *C. difficile* (Berges et al. 2018).

Pour clôturer, comme d'autres bactéries entériques, *C. difficile* est capable d'utiliser l'éthanolamine comme une source intermédiaire de carbone et de nitrogène. L'éthanolamine forme les phosphatidyléthanolamines présentes dans les membranes des bactéries, mais également dans celles des cellules hôtes. En conséquence de la production de toxines, les membranes des cellules vont être dégradées, permettant la libération des phosphatidyléthanolamines. Ils pourront alors être transformés en glycérol et éthanolamine grâce à une phospholipase de *C. difficile*. L'éthanolamine est par la suite transportée dans un micro-compartiment pour être métabolisé en ammoniac, source de nitrogène et en acétaldéhyde, source d'acétyl-CoA et d'éthanol (Roof et Roth 1989).

G. Microbiote intestinal et la dysbiose :

Le microbiote intestinal est défini comme une collection de bactéries, les Archées et les Eucaryotes, colonisant l'intestin pour former un superorganisme grâce à une relation symbiotique avec l'hôte. Il offre de nombreux bénéfices à l'hôte jouant un rôle dans l'intégrité de l'épithélium intestinal, la production d'énergie, la protection contre les pathogènes ainsi que dans la régulation du système immunitaire (Thursby et Juge 2017). Le microbiote intestinal est un équilibre pouvant fluctuer au cours du temps. Chez les enfants, sa composition varie en fonction du mode d'accouchement, mais il est composé principalement de *Bifidobactérium*. Chez l'Homme adulte sain, les Firmicutes et Bacteroidetes constituent 90 % du microbiote intestinal tandis que le reste est composé d'Actinobacteria, Proteobacteria, Fusobacteria et de Verrucomicrobia. Il est influencé par de nombreux facteurs comme l'âge, le régime alimentaire, la fréquence d'activité physique et la prise d'antibiotique (Rinninella et al. 2019).

Ces facteurs peuvent entraîner une dysbiose correspondant à une diminution du taux de bactéries anaérobies obligatoires et une augmentation de l'abondance de certaines bactéries anaérobies non obligatoires entraînant une diminution globale de la diversité bactérienne (Martinez, Taminiau, et al. 2022). Les patients infectés par *C. difficile* montrent une diminution de la diversité microbienne. En effet, leur microbiote est majoritairement composé de Proteobacteria avec une diminution importante de l'abondance en Firmicutes. Des changements dans les proportions des genres de Bacteroidetes sont également observés (Martinez, Taminiau, et al. 2022).

Lors de la dysbiose, les changements d'équilibre vont permettre à *C. difficile* d'entamer la germination et de coloniser l'intestin. Pour commencer, la germination des spores va être favorisée grâce aux sels biliaires primaires produits dans le foie et excrétés dans l'intestin grêle. En temps normal, ils sont réabsorbés majoritairement et les résidus sont métabolisés en sels secondaires dans le colon à l'aide de bactéries faisant partie des Firmicutes. Cependant, en cas de dysbiose et de perte de la diversité, les sels primaires ne sont plus aussi bien métabolisés, ce qui conduit à une augmentation de leur concentration dans le colon, favorisant ainsi la germination des bactéries pathogènes opportunistes comme *C. difficile*.

Afin de pouvoir germer et coloniser l'intestin, *C. difficile* doit également pouvoir passer outre les barrières de défenses du tube digestif, notamment celle du mucus. Les bactéries productrices de mucus sont principalement des Firmicutes. C'est pourquoi, chez les patients infectés, il y a un changement de composition de ce mucus intestinal avec une augmentation des MUC1, tandis que chez les individus au microbiote sain, ce sont les MUC2 qui prédominent (Martinez, Taminiau, et al.

2022). Cela se traduit par un déséquilibre de la barrière de protection favorisant la colonisation (Engevik et al. 2015).

Pour continuer, les changements d'abondance engendrent également une diminution de la production des acides gras à chaîne courte par les Firmicutes. Les acides gras à chaîne courte jouent un rôle important dans la protection du colon, mais également dans la production d'énergie. Dans l'ICD, leur diminution va permettre de créer un environnement favorable à *C. difficile* en augmentant le pH tout en diminuant l'intégrité de l'épithélium (Martinez, Rodriguez, et al. 2022).

Pour terminer, les personnes souffrant d'un déséquilibre de leur microbiote ont tendance à avoir un déséquilibre du biofilm commensal. Lorsque la dysbiose s'installe, la concentration des bactéries capables de dégrader le biofilm commensal va augmenter. La disparition du biofilm a pour conséquence de rendre accessible les cellules eucaryotes aux agressions et aux pathogènes. En effet, le biofilm délétère va soit se disperser laissant les cellules eucaryotes à nu, soit devenir invasif et entraver la couche stérile entourant les cellules eucaryotes. *C. difficile* est une bactérie capable de produire un biofilm, particulièrement en présence de *Fusobacterium*, une bactérie prédominante du microbiote des personnes infectées par *C. difficile* (Meza-Torres et al. 2021).

H. Whole genome sequencing et transcriptomic analysis :

Le séquençage de première génération existe depuis 1977 avec la méthode Sanger. Cette technique utilisait des ddNTPS radioactifs puis des ddNTP fluorescents pour pouvoir déterminer la position des acides nucléiques. Il s'agit de nucléotides dépourvus du groupe 3' hydroxyle nécessaire à la polymérase à l'élongation de la séquence. Cependant, cette 1^{re} technique ne permettait pas d'obtenir des séquences de plus de 1kb même en utilisant des séquence overlaps grâce à la technique du « shot-gun ».

Par la suite, des techniques dites de seconde génération ont été mises en place, notamment Illumina seq. Ces techniques permettent le séquençage d'un groupe hétérogène d'ADN. Pour ce faire, des adaptateurs sont ajoutés aux brins d'ADN et sont amplifiés par PCR afin d'obtenir des séquences short-read avec peu d'erreur. Enfin, les techniques de 3^e génération permettent d'obtenir rapidement des longues séquences sans amplifications grâce à la détermination de plusieurs nucléotides simultanément. Il s'agit du séquençage par nanopore. La séquence est déterminée en fonction de la perturbation du champ électrique au passage des acides nucléiques au travers d'un pore. C'est ainsi que l'on peut obtenir de longues séquences, mais qui ont cependant une faible profondeur (Heather et Chain 2016).

Le séquençage a notamment permis d'obtenir des séquences complètes de génome qui peuvent servir de référence. On retrouve notamment MinION nanopore qui, suite à une préparation de la librairie déposée sur une « flowcell » et l'assemblage des longs read, permet d'obtenir un génome complet (Quick, Quinlan, et Loman 2014).

Le génome complet de différentes souches de *Clostridioides difficile* a déjà été séquencé, notamment celui des ribotype 078 et 027 qui correspondent aux souches les plus communes.

Les analyses transcriptomiques permettent d'étudier les ARN produits par un organisme rendant possible la quantification de l'expression des gènes. Ce type d'analyse aide à la compréhension des interactions entre gènes, mais également à la mise en évidence des fonctions exprimées. Ces informations permettent d'avoir une idée globale du comportement de croissance et des systèmes biologiques utilisés.

Comme pour l'étude du génome, il existe plusieurs plateformes de séquençage pour la transcriptomique dont Illumina Seq.

Les analyses transcriptomiques se concluent en général par une analyse différentielle de l'abondance à l'aide d'outils bio-informatiques pour calculer les logarithmes² des deltas du taux d'expression entre la condition testée et la condition contrôle.

II. Hypothèses, objectifs et stratégies :

L'infection par *Clostridioides difficile* ne se manifeste pas chez tous les individus atteints. Chez l'adulte, la bactérie doit être capable de produire des toxines pour potentiellement engendrer le développement de la maladie. Cette production de toxines est enclenchée durant la phase stationnaire qui a lieu suite à la réussite de la colonisation des intestins. Cela fait donc de la colonisation une des étapes importantes pour le développement de la maladie.

De plus, il semblerait que le microbiote soit un facteur influençant de façon significative la colonisation. Effectivement, la majorité des facteurs de risques sont responsables de l'induction d'une dysbiose mais également de l'activation du système immunitaire.

Par conséquent, l'hypothèse est de savoir quel est l'influence du microbiote de l'hôte sur le comportement de *C. difficile* dans le but de comprendre pourquoi certaines personnes dont l'intestin est colonisé par *C. difficile* sont asymptomatiques et d'autres non.

Ce mémoire s'ancre dans un projet doctoral visant à étudier le comportement de *C. difficile* en fonction de son environnement bactérien et du temps par analyse transcriptomique. L'objectif est de mettre en évidence des motifs d'expressions de *C. difficile* lorsque celle-ci se trouve en présence de bactéries compétitives.

Pour ce faire, des expériences *in vitro* sur des « transwells » ont été préalablement réalisées. Il s'agit d'une « well-plate » dont les puits sont remplis soit par des matières fécales provenant de deux individus différents (F03 et F04) soit remplis par un bouillon de culture Feed and PJ (WOF). Dans chaque puits se trouve également un insert contenant *C. difficile* dans le même bouillon de culture. Puisque l'étude se base sur une analyse transcriptomique, l'intérêt des « transwells » est que le pool d'ARNm produit par *C. difficile* ne soit pas perdu dans celui des ARNm produit par d'autres bactéries. En effet, le filtre de l'insert (0.4 mm) limite le passage des bactéries tout en permettant une interaction avec *C. difficile* par le passage des métabolites.

Afin de réaliser des analyses transcriptomiques, les ARN exprimés par *C. difficile* ont été récoltés. Plusieurs temps d'analyse ont pu être testés : 1h, 6h, 12h, 24h, 48h. Cependant, durant ce mémoire, nous nous sommes uniquement concentrés sur le temps 12h.

C'est à partir des résultats transcriptomiques que nous avons pu chercher des motifs d'expression entre les trois différents environnements à l'aide d'un programme/script qui a permis

l'enrichissement de données. L'enrichissement des données transcriptomiques sert à faciliter la gestion des données transcriptomiques. En effet, les données transcriptomiques vont être classées en fonction de différents paramètres grâce à « FeedMaurice », un script Perl qui utilise le génome de référence et l'annotation de celui-ci par deux encyclopédies génomiques NCBI et KEGG pour identifier les gènes.

Dans un second temps, la validation des résultats obtenus en transcriptomique s'est faite par qPCR SyberGreen. L'analyse transcriptomique a permis d'obtenir une analyse globale de l'expression des gènes, ce qui n'est pas possible en qPCR. Par cette raison, il a été nécessaire de faire une sélection de gènes à analyser en qPCR. Ce sont des gènes caractéristiques de *Clostridioides difficile* qui ont été choisis, ainsi que certains gènes transcrits en quantité suffisante et dont le changement d'expression s'est montré significatif en transcriptomique.

III. Matériel et méthodes :

A. Analyse génomique de la référence :

Afin de circulariser notre génome, l'ADNg de souche S0756-078 de *C. difficile* a été séquencé dans le but de reconstruire le génome complet sous forme circulaire après séquençage par Illumina (Martinez, Taminiau, et al. 2022) et MinION. Le génome circularisé nous permet d'obtenir un génome complet rendant possible le mapping de nos séquences RNAseq.

1. Extraction ADNg :

Nous avons utilisé l'ADNg de la souche S0756 - ribotype 078 de *Clostridioides difficile* extrait d'un isolat à l'aide du kit Blood and Tissue (Qiagen). Il a été séquencé préalablement séquencé sur Illumina (Martinez, Taminiau, et al. 2022) puis MinION.

2. Séquençage MinION :

Le séquenceur MinION est un séquenceur Oxford Nanopore, il a été utilisé pour séquencer notre génome de référence à l'aide du kit « ligation sequencing » (Oxford Nanopore) et d'une « flow cell » R9.4.1. (Oxford Nanopore).

3. Assemblage :

a) *Unicycler* :

Unicycler v0.5.0 (Wick et al. 2017) est un assembleur hybride utilisant les reads obtenus par séquençage Illumina et par séquençage MinION. Unicycler fonctionne comme un optimiseur de SPAdes, un autre assembleur pour permettre l'assemblage en passant par différentes étapes. Cet assembleur a été choisi car il est capable d'assembler un génome bactérien à partir de « short reads » Illumina qui sera ensuite poli par « long reads » de qualité variable tel que ceux obtenus par MinION.

Unicycler passe par 7 étapes dans le but d'obtenir un génome circularisé. Il démarre par la formation d'un graphe de De Bruijn avec SPAdes. Il s'agit d'un graphe représentant l'assemblage de différentes séquences définies par une longueur en k-mer pour essayer de former le plus long contig. SPAdes utilise le « paired-end information » pour former différents graphes de De Bruijn avec différentes tailles de k-mer. Le meilleur graphe reprendra une longueur de k-mer intermédiaire puisque, plus le « reads » est long, plus le contig que l'on va former sera long lui aussi. Cependant, plus le « reads » est long, plus le risque de « dead end » est élevé. Pour ce faire, un score de pénalité est attribué à chaque graphe de De Bruijn hypothétique et celui avec le meilleur score est conservé pour la suite. Pour continuer, Unicycler va supprimer les contigs de qualité moindre, c'est-à-dire les « reads » avec une profondeur plus petite que la moitié de profondeur médiane du graphe.

L'étape suivante est la « multiplicity », où un algorithme va déterminer le nombre de répétitions de chaque contig. En effet, il va distinguer les « single-copy » contigs qui ne sont représentés qu'une seule fois, des « repeat » contigs que l'on retrouve plusieurs fois dans le génome. Ces 2 types de contigs sont différenciés en fonction de leur profondeur et du nombre de connexion. Pour continuer, le nombre de connexion permet de repérer les séquences répétées puisque, plus il y a de connexions avec d'autres séquences, plus la probabilité d'avoir une séquence répétée est grande. L'algorithme de « multiplicity » va prendre en compte ces informations pour chaque séquence et accorder un « copy number ». Un single contig aura un copy number égal à un. Ils correspondent aux « anchors » contigs qui serviront de base/de squelette à l'assemblage. De plus, Unicycler va également supprimer les séquences « overlaps » pour permettre la circularisation du génome.

Lors de la dernière étape, Unicycler va former des ponts pour lier les différents « anchor » contigs à l'aide des « repeat » contigs ainsi que les long « reads » obtenus par MinION. Il existe plusieurs possibilités pour lier les contigs ensemble, un score est accordé aux différents ponts et le pont qui a le meilleur score est conservé pour essayer de former un génome circulaire.

b) Trycycler :

Trycycler v0.5.0 (Wick et al. 2021) est un outil bio-informatique qui utilise 3 assembleurs hybrides différents. Cela permet de palier aux erreurs propres à l'outil d'assemblage. Pour utiliser Trycycler, il faut des « reads » d'une certaine qualité. En effet, ils doivent être suffisamment longs avec une profondeur de minimum 100, comme ceux obtenus par séquençage au MinION. Trycycler a la capacité d'assembler des « reads » obtenus par deux types de séquenceurs puisqu'il va utiliser les long « reads » de MinION pour produire un assemblage qui sera poli par les « reads » d'Illumina.

Avant de commencer, une étape de nettoyage est recommandée, mais non obligatoire, pour nettoyer les « reads » afin d'éliminer les moins bons ainsi que les séquences dont la longueur ne dépasse pas 1 kpb. Dans notre cas, nous avons passé cette étape afin de garantir une couverture suffisante de 100x.

La 1^{ère} étape consiste en la formation d'un sous-sampling/sous-échantillonnage à partir de la librairie MinION, c'est ce qu'on appelle le « subsampling ». Cette étape permet d'obtenir des « reads » indépendants les uns des autres pour pouvoir générer plusieurs assemblages « input » du même génome. C'est grâce à ces 12 sous-assemblages que les erreurs propres aux assembleurs sont éliminées. Cela permet *in fine* d'augmenter la robustesse de l'assemblage final.

Pour continuer, Trycycler va sous-échantillonner les « reads » de manière aléatoire, ce qui va également améliorer la robustesse de l'assemblage final. Effectivement, il faut savoir que les runs de

Assembleur	Unicycler	Tricycler	SPADES v. 3.10.0
Longueur	4,071,259 bp	4,071,782 bp	4,042,996 bp
# contig	1	1	258
Profondeur	1.00x	1.00x	1.00x
% GC	29,05%	29,05%	29,01%
RNA _t	89	89	81
RNA _r	35	35	28
CDS	3666	3685	3741
CDS hypothétiques	960	964	1009
N50	4 071 259	4 071 782	129 124
L50	1	1	11

Table 1 Comparaison des trois assemblages produits par Unicycler, Tricycler, SPADES en fonction de la longueur de la séquence, du nombre de contigs, de la profondeur, du pourcentage GC, du nombre de ARN_t et ARN_r, du nombre de CDS et CDS hypothétique, du N50 et du L50

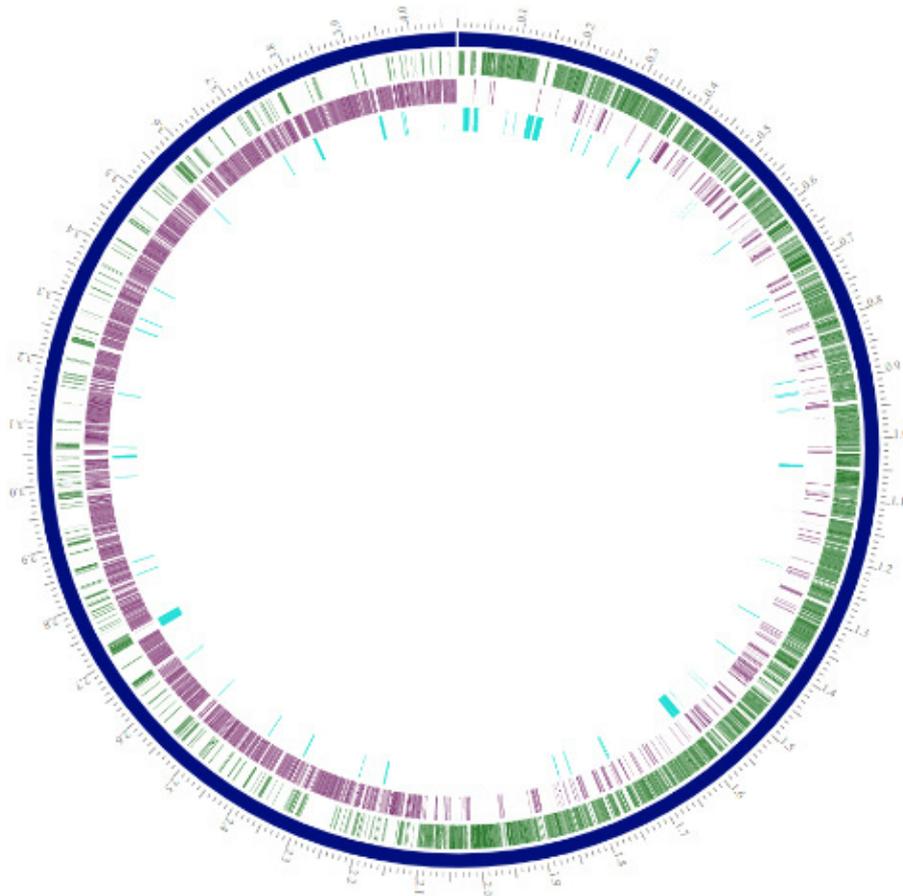


Figure 1 Image obtenue sur BV-BRC du génome complet et circulaire de *Clostridioides difficile* s0756_078 assemblé par Tricycler. En bleu, la séquence du chromosome complet et les coordonnées des gènes. Les CDS forward sont représentés en vert et les CDS reverse sont représentés en mauve. En bleu turquoise, sont annotés les « non-CDS features »

séquençage par nanopore diminuent en qualité au cours du séquençage. Cela signifie que lorsque les « reads » sont classés par ordre temporel, les premiers « reads » seront de meilleure qualité que les derniers. Cette étape de « subsampling » permet ensuite d'utiliser trois assembleurs différents : Flye, Miniasm+minipolish et Raven. Trycycler va ensuite clustériser ensemble les « subsamples » des différents assembleurs. Les contigs sont groupés en per-réplicon dans un cluster à partir de la distance MASH. Il s'agit d'une distance permettant de réduire les grandes séquences en graphes. Elles seront ainsi plus faciles à analyser afin de déterminer la distance de mutation. Cette distance est synonyme du nombre de mutations séparant deux séquences. Le but est d'éliminer les mauvais contigs, c'est-à-dire les contigs incomplets ou mal-assemblés.

Dans chaque cluster, les séquences venant des trois assembleurs sont assemblées en contigs. Cette étape permet à Trycycler d'accomplir plusieurs tâches. Tout d'abord, une vérification des contigs est faite. Les contigs doivent être \pm de la même taille avec une distance MASH réduite. Ensuite, Trycycler va s'assurer que les contigs se trouvent dans la même orientation. Cette étape se fait d'abord par conversion des séquences en la séquence reverse complémentaire. Par la suite, les problèmes de circularisation sont réparés pour l'obtention d'un génome circulaire. A ce stade, Trycycler détermine le point de départ du génome circulaire de deux façons : soit, Trycycler trouve une séquence « starters » dans sa base de données de gènes connus pour être des « starting positions », soit, en cherchant une séquence unique aléatoire si aucune séquence homologue n'est trouvée dans la base de données.

Lors de l'étape suivante, les contigs des différents assembleurs vont être assemblés en fonction d'un alignement multiple réalisé par MUSCLE. Les contigs sont assemblés en fonction de l'alignement multiple pour former un consensus par cluster. S'il y a plusieurs clusters consensus, ils doivent être assemblés ensemble. Cependant, ce n'est pas le cas ici et cette étape a été passée puisqu'un seul cluster a pu être généré.

Pour terminer, la séquence consensus est polie par Polypolish à partir des « short-reads » Illumina permettant l'élimination des erreurs de séquençage propres à MinION. En effet, ce type de séquençage donne de plus « long reads » mais avec un nombre d'erreurs plus important qu'avec le séquençage Illumina.

4. Annotation :

a) *NCBI* :

Le génome de référence a été annoté sur le **National Center for Biotechnology Information**, une « database » de génome (« Clostridioides difficile (ID 535) - Genome - NCBI » s. d.) C'est par cette annotation que nous allons pouvoir identifier les gènes. En effet, NCBI détermine les CDS, leur nom de gène ainsi que celui de la protéine correspondante et leur « enzyme commission number » (E.C

number) s'ils en ont un. L'E.C number permet une classification des réactions enzymatiques, ce numéro associe donc une réaction enzymatique à une/plusieurs enzymes capables de la réaliser.

b) KEGG :

Kyoto Encyclopedia of Genes and Genomes est une encyclopédie de gènes et de génomes qui permet de définir un gène par sa fonction métabolique dans un « pathway ». Pour cela, KEGG utilise un « KEGG Ortholog number » (KO), un numéro qui rassemble des gènes ayant une fonction métabolique/chimique semblable. A partir de ce « KO number », il existe plusieurs points d'entrée dans la « database » de KEGG, notamment le KEGG module. Un module est défini comme un ensemble de « KO number » formant un même processus biochimique.

A partir de GhostKOALA, un système d'annotation interne à la database, KEGG attribue des « KO number » à partir des séquences protéiques en fasta (Kanehisa, Sato, et Morishima 2016). Ces « KO number » sont eux-mêmes associés à un ou plusieurs modules.

B. Analyse transcriptomique :

1. Extraction ARN total :

A partir des échantillons liquides des plaques « transwells », l'ARN total a été extrait à l'aide du kit RNeasy PowerMicrobiome en temps 12 h de chaque « transwell » et ont subi un traitement à la DNase.

L'absence de contamination en ADN bactérien a été vérifié avec une PCR ciblant le gène 16S sur chaque échantillon.

2. Séquençage RNAseq par Illumina :

Les ARN extraits ont été transférés au GIGA. Ils ont été soumis à une analyse de qualité QC (Agilent 2100 bioanalyzer) et les ARN de qualités suffisantes ont été conservé afin de réaliser un séquençage RNAseq par Illumina. Le séquenceur Illumina avec le système NovaSeq S4 V1 306 et le kit RiboZero ont été utilisés.

Les ADNc de chaque condition à temps 12 h ont été séquencés au laboratoire GIGA. Une librairie correspond à un réplica d'une condition expérimentale. 18 librairies différentes ont été séquencées pour nos deux analyses avec trois conditions répliquées trois fois chacune : deux librairies sans matière fécale WOF réplica 1-2-3, deux librairies matière fécale 3 réplica 1-2-3 et deux librairies matière fécale 4 réplica 1-2-3.

3. Premières étapes de nettoyage de librairies :

a) *FastP* :

Avec le logiciel FastP v0.23.4 (« Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp - Chen - 2023 - iMeta - Wiley Online Library » s. d.), les « reads » ont été évalués par rapport à leur contenu en GC et le taux de séquences répétées avant d'être filtrés.

b) *SortMeRNA* :

SortMeRNA v2.1b, 03/03/2016, est un outil bio-informatique en Base qui permet la séparation des ARNr des ARNm par mapping sur une collection de données d'ARNr bactérien (ARNr 16S et les ARNr 23S des bactéries, les ARNr 16S et les ARNr 23S des Archées, les ARNr 18S et les ARNr 28S des Eucaryotes ainsi que les ARNr 5S). Les séquences reconnues seront flaggées avant d'être écartées (Kopylova, Noé, et Touzet 2012). Seuls les « reads » correspondant à des ARNm seront analysés.

4. Analyse transcriptomique :

a) *Mapping sur génome avec subread et subjunc* :

La première étape consiste en la création d'un index à partir de notre séquence de référence. En effet, dans le but de faire du « mapping », la fonction subread-buildindex v2.0.1 crée un index facilitant le mapping en divisant notre séquence de référence en « reads » plus courts.

C'est à partir de cet index que nous allons pouvoir identifier les gènes exprimés à l'aide de la fonction subjunc v2.0.1. Les reads séquencés des ADNc vont alors être mappés sur notre séquence de référence pour attribuer les coordonnées de chaque gène.

b) *Table de comptage avec featureCounts* :

Afin de déterminer le taux d'expression des ARNm transformés en ADNc, nous utilisons le logiciel FeatureCounts v2.0.1. Ce logiciel va convertir le nombre de « reads » séquencés de nos différentes librairies en comptage de gène.

La fonction featureCounts va créer des tables de comptage pour chaque pool de « reads » des différentes conditions de « transwells ». Les tables de comptage reprennent les comptes des gènes mappés de chaque librairie en les associant avec leurs coordonnées par un gène ID.

c) *Analyse DESEQ* :

A partir des tables de comptage obtenues par FeatureCounts, une analyse différentielle de l'abondance a été réalisée avec R sur Rstudio 2022.12.0+353 avec deux packages : DESeq2 1.36.0 (Love, Huber, et Anders 2014).

DESeq2 est un package Rstudio permettant la comparaison pairée dont le résultat est exprimé en « \log_2 fold change » accompagné de la p -valeur et p -valeur ajustée. Il s'agit d'un ratio d'expression pour chaque gène entre la condition contrôle et la condition testée.

Pour ce faire, DESeq2 normalise les tables de comptage afin d'éliminer le biais lié aux différentes tailles de bibliothèques en fonction des conditions. Ensuite, il évalue la dispersion. Celle-ci permet d'évaluer à combien l'échantillon fluctue autour de la moyenne. La mesure d'analyse différentielle de l'abondance se fait ensuite avec un General Linear Model (Negative Binomial GLM) qui suit la régression de Poisson et un test de Wald dans le but de démontrer la signifiante corrigée par la suite à l'aide du test Benjamini-Hochberg (« false discovery rate ») (Love, Huber, et Anders 2014).

5. Analyse par enrichissement :

a) *Annotation des gene counts avec les données KEGG, et E.C.:*

A partir de notre génome de référence et de nos deux dictionnaires obtenus à partir de KoalaGhost et NCBI, nous avons pu établir une correspondance entre les tables de comptage et les informations descriptives contenues dans les dictionnaires.

b) *FeedMaurice :*

FeedMaurice est un script perl qui permet l'enrichissement des données transcriptomiques obtenues. Un enrichissement en transcriptomique est appelé un « gene set enrichment ». Il s'agit d'une analyse bio-informatique visant à ressortir une des listes de gènes afin de les mettre en évidence avec leur taux d'expression en fonction d'une voie métabolique, d'une fonction enzymatique (Subramanian et al. 2005). Pour cela, il est nécessaire d'avoir un logiciel d'enrichissement possédant le génome de référence. Cependant, aucun outil bio-informatique actuel ne possède un génome de référence pour *C. difficile*. FeedMaurice est un script d'enrichissement dans lequel n'importe quel génome de référence peut être introduit. Ce script va mettre en évidence le module métabolique et la fonction enzymatique des gènes avec leur « \log_2 fold change ». Pour cela, il utilise la référence annotée sur NCBI et un dictionnaire de KO-modules venant de l'encyclopédie KEGG pour produire trois outputs majeurs :

Le premier reprend l'ensemble des gènes, leur KO, leur « E.C number », leur « \log_2 fold change », la p -valeur et p -valeur ajustée. Le deuxième contient la liste des KO exprimés et la médiane des « \log_2 fold change » des gènes de même KO. Le troisième « output » présente la liste des modules exprimés, la médiane des « \log_2 fold change » des gènes faisant partie d'un même module.

C. Validation par qPCR :

1. Choix des cibles et Design PCR :

Pour valider les résultats transcriptomiques, certains gènes sélectionnés vont être amplifiés par PCR et qPCR.

A partir des résultats transcriptomiques, 6 gènes ont été sélectionnés en fonction de leur « baseMean » et leur « log₂ fold change » pour ensuite concevoir des amorces (Table 1).

Grâce à featureCounts, nous obtenons le « baseMean » de chaque gène. Celui-ci correspond au nombre moyen de « reads » pour un gène dans les différentes conditions en fonction de la taille de la librairie. Un « baseMean » élevé correspond à une expression moyenne élevée. Cela permettra de mettre en évidence la séquence amplifiée plus facilement en qPCR. Pour continuer, à l'aide du « log₂ fold change » correspondant au ratio d'expression entre deux conditions, nous pouvons également choisir un gène avec un ratio éloigné de 1 afin d'être plus assuré d'observer une différence en qPCR.

Les amorces ont été conçues à partir de leur séquence génomique sur BV-BRC 3.28.5. (Olson et al. 2023). Les amorces choisies sont constituées de 20 nucléotides avec un produit de PCR entre 150 et 200 nucléotides. Afin d'assurer leur spécificité, tous les couples d'amorces ont été blastés sur NCBI Blast. De plus, l'outil Multiple Primer Analyzer (ThermoFisher) a permis d'identifier de potentiels hétérodimères.

2. Extraction ADNg :

Dans l'objectif de valider nos amorces et d'avoir des standards pour les qPCR, nous devons extraire l'ADN génomique de *Clostridioides difficile* s0756-078.

L'extraction d'ADN se fait à l'aide du kit Blood and Tissue à partir d'une culture solide. Une extraction typique d'ADN commence par une lyse cellulaire chimique et mécanique. Il faut ensuite cliver les protéines et inhiber les DNases. L'ADN sera par la suite déposé sur une colonne pour s'y fixer dans le but de le purifier avant l'élution.

La 1^{ère} étape du protocole du Kit Blood and Tissue correspond à la lyse chimique qui consiste au dépôt de 180 µl de tampon de lyse contenant la bonne quantité de lysozymes (20 mg par ml de tampon) dans un Eppendorf. À l'aide d'une oese, nous ajoutons une colonie en l'écrasant sur les parois de l'Eppendorf avant une incubation d'une heure à 37°C sous agitation dans le Thermomix.

Après incubation, il faut ajouter 5 petites billes et 2 grosses billes à la solution dans le but d'augmenter les surfaces de tension durant la lyse mécanique. On ajoute également de la protéinase

K permettant le clivage des protéines et 200 µl de buffer « AI », un inhibiteur de DNases. Il faut ensuite laisser incuber pendant une 1 h à 56°C avant de récupérer le surnageant.

Dans un nouvel Eppendorf avec 200 µl d'éthanol 96%, nous déposons 350 µl de surnageant. Par la suite, la solution doit être homogénéisée avant d'être transférée sur colonne pour être centrifugée 1 min à 12 000 g. Le filtrat est jeté et la colonne est placée sur un nouveau tube collecteur. La colonne est ensuite lavée une 1^{ère} fois avec 500 µl de solution de lavage « AW1 » et centrifugée 1 min à 12 000 g. Le filtrat est à nouveau jeté et la colonne est placée sur un nouveau tube collecteur. La colonne est lavée une 2^{ème} fois avec 500 µl de solution de lavage « AW2 » et centrifugée 3 min à 12 000 g. Après avoir changé de tube collecteur, la colonne est centrifugée à vide afin d'éliminer les restes d'éthanol.

Pour finir, l'ADN est élué avec 50 µl de buffer AE et une centrifugation d'1 min à 6000 g. Afin d'améliorer le rendement, cette dernière étape est répétée une seconde fois. L'ADN récolté est stocké dans le congélateur à -20°C dans le local de stockage ADN.

a) *Quantification de l'ADN par nanodrop :*

Après extraction, il est intéressant d'avoir une idée de la pureté et de la quantité d'ADN extrait. Le Nanodrop 2000 permet la quantification par mesure d'absorbance en 260 et 280 avec 1 µl d'échantillon. La pureté est évaluée avec le rapport de l'absorbance en 260/280 qui doit se trouver entre 1,8 et 2. Cela permet d'évaluer la contamination par les protéines. En effet, l'ADN absorbe à 260 et 280, tandis que les protéines n'absorbent qu'en 280.

3. Validation design :

a) *PCR gradient :*

Pour vérifier la température de fusion, nous avons réalisé des PCR gradients sur une PCR Eppendorf MasterCycler gradient en utilisant l'ADN génomique extrait d'une souche de *Clostridioides difficile* 0756. La PCR gradient a été réalisée dans un volume de 20 µl contenant 10,2 µl d'eau moléculaire, 1,6 µl de MgCl₂ 25 mM, 2 µl de Buffer 10x, 5 µl BSA 20 ng/ml, 0,2 µl Taq 5 U/µl, 2 µl dNTP 2 mM, 1 µl Primer Forward, 1 µl Primer Reverse et 2 µl d'ADN génomique. Le programme PCR se déroule en 35 cycles après la dénaturation initiale à 95°C pendant 5min. Un cycle commence par une dénaturation à 95°C pendant 30 sec, suivie par une hybridation de 30 sec avec un gradient de 52°C à 59°C et d'une élongation à 72°C pendant 30 sec. Le programme se termine ensuite par une élongation finale à 72°C pendant 5 min.

Couple	Région	Séquence	Longueur du produit	Tm	%GC
1 <i>eutA</i>	Forward primer	AATATCAGGAAAAGGGGCTG	154	57	45
	Reverse Primer	AAGCCTCCACCTATATCCA		57	45
2 <i>eut B</i>	Forward primer	CCAGTTGATGACTCTGTTGA	140	57	45
	Reverse Primer	GGAGCACCATTTCTTATTGC		57	45
3 <i>hadA</i>	Forward primer	CATTATGCAGGTCTAGCACT	194	54,93	45
	Reverse Primer	GTCATTAATGGGCTGTTTGG		54,91	45
4 <i>rnfG</i>	Forward primer	AGACCTGGAGATGGTCTTA	150	54,84	45
	Reverse Primer	TGTTACCAATGTCAACACCT		54,79	40
5 <i>flaA</i>	Forward primer	TGTTGATGCTGGATGGATAG	192	54,90	45
	Reverse Primer	ACTCCATAGTCAGCGATTTC		54,97	45
6 <i>polyDEA</i>	Forward primer	TATTGCATTTCCCTCGTCAA	197	55,01	40
	Reverse Primer	ACGACTATTCAAACAACGGA		55,01	40

Table 1 Conception des amorces sélectionnées

b) PCR :

La validation des amorces a été réalisée par PCR classique (Techno Flexigene) en utilisant l'ADN génomique extrait d'une souche de *Clostridioides difficile* 0756. La PCR a été réalisée dans un volume de 30 µl contenant 15,3 µl d'eau moléculaire, 2,4 µl de MgCl₂ 25 mM, 3 µl de Buffer 10x, 3,75 µl BSA 20 ng/ml, 0,3 µl Taq 5U/ µl, 3 µl dNTP 2mM, 1,5 µl Primer Forward, 1,5 µl Primer Reverse et 3 µl d'ADN génomique. Le programme PCR se déroule en 35 cycles après la dénaturation initiale à 95°C pendant 5 min. Un cycle commence par une dénaturation à 95°C pendant 30 sec, suivie par une hybridation de 30 sec à la température de fusion recommandé (voir tableau) et d'une élongation à 72°C pendant 30 sec. Le programme se termine ensuite par une élongation finale à 72°C pendant 5 min.

c) Gel électrophorèse :

Pour vérifier l'amplification par PCR, une migration par électrophorèse sur un gel d'agarose à 2% a été effectuée dans un bain de TBS à 100 volts pendant 30 min. La visualisation des brins amplifiés a été faite sur Image Lab 5.0 avec BioRad Gel Doc EZ imager.

d) Purification :

La purification de nos produits PCR a été effectuée à l'aide du kit Promega « Wizard® SV gel and PCR clean-Up system ». Pour ce faire, les 25 µl de produit PCR ont été déposés sur une colonne avec 50 µl de membrane binding solution pour ensuite être lavés 2x avec la membrane wash solution (70% d'éthanol). L'ADN a été élué dans 50 µl d'eau « nuclease free » et déposé au congélateur à -20°C.

e) Quantification de l'ADN amplifié au PicoGreen :

Afin de connaître la quantité exacte d'ADN amplifié contenu dans nos échantillons, une quantification au PicoGreen a été réalisée avec une courbe étalon à partir d'une dilution d'ADN lambda (100µg/ml) (invitrogen Ref p7589C, life technologies corporation) dans du tampon TE 1x composé de Tris-HCL et EDTA. Sur une plaque 96puits, 1 µl des échantillons et de leur dilution 10x sont déposés avec 99 µl de tampon TE 1x ainsi que 100 µl des dilutions de l'ADN standard. Dans chaque puits, nous ajoutons 100 µl PicoGreen 1x. Les résultats sont ensuite visualisés sur le spectrophotomètre Fluostar optima BMG labtech. C'est suite à la quantification au picogreen que nos standards de qPCR ont été réalisés.

f) Séquençage des produits PCR :

Puisque la PCR ne permet pas de vérifier spécifiquement quelle séquence est amplifiée, il faut passer par une étape de séquençage afin de valider l'amplification spécifique d'une séquence à partir des amorces choisies. Pour ce faire, des Eppendorfs d'une concentration de 0,2 à 0,6µg/µl pour 150pb

de produit PCR dans un volume final de 10 μ l, 5 μ l d'amorce « forward » ont été préparé. Cette concentration est dépendent de la taille du fragment amplifié.

4. qPCR :

Sur la qPCR BioRad CFX96 RealTime System, nous avons réalisé des qPCRs sur nos échantillons de ADNc venant des ARNm extraits à 12h. Les qPCR ont été réalisées dans un volume final de 12,5 μ l contenant 1,25 μ l d'eau moléculaire, 6,25 μ l de MasterMix Takyon™ TOX SYBR et 2,5 μ l BSA 20 ng/ml, 0,25 μ l « Primer Forward », 0,25 μ l « Primer Reverse » et 2 μ l d'ADN standard ou d'ADNc. Le programme PCR se déroule en 39 cycles après l'activation de la polymérase Takyon (Eurogentec) à 50°C pendant 2 min. Un cycle commence par une dénaturation à 95°C pendant 20 min, suivie par une hybridation de 20 sec à 95°C et une étape d'élongation pendant 1 min à température de fusion.

librairie	"reads" brutes	"reads" nettoyés	taux de duplication
WOF_1_1	11 952 966	11 844115	24.64%
WOF_1_2	11 197 127	11 140 587	26.36%
WOF_1_3	10 853 960	10 795 986	26.75%
F04_1_1	10 640 093	10 498 085	34.72%
F04_1_2	9 959 621	9 738 546	40.16%
F04_1_3	10 967 073	10 832 246	40.36%
F03_1_1	10 535 106	10 429 019	35.09%
F03_1_2	11 828 067	11 747 133	22.09%
F03_1_3	11 744 240	11 669 049	33.00%
WOF_2_1	13 256 089	13 167 936	50.91%
WOF_2_2	12 342 942	12 288 261	43.25%
WOF_2_3	13 148 423	13 087 385	58.65%
F04_2_1	13 614 015	13 422 740	82.56%
F04_2_2	10 512 821	10 208 000	90.13%
F04_2_3	10 722 551	10 505 005	82.42%
F03_2_1	12 040 676	11 293 169	90.30%
F03_2_2	11 071 515	10 494 076	89.86%
F03_2_3	13 176 565	12 834 145	85.89%

Table 3 Taux de traitement des "reads" pour l'analyse 1 et 2 des différents réplicas pour les conditions WOF, F03 et F04 collecté à 12h

librairie	nombre de "reads" total	ARNr	ARN	%ARNr	%ARN
WOF_1_1	2228174	5011289	1867641	21.16%	78.84%
WOF_1_2	2281174	836589	13920585	37.52%	62.48%
WOF_1_3	21591972	8998316	12593656	41.67%	58.33%
F04_1_1	20996170	7435483	13560687	35.41%	64.59%
F04_1_2	19477092	7162076	12315016	36.77%	63.23%
F04_1_3	21664492	7681911	13982581	35.46%	64.54%
F03_1_1	20858038	18915412	1942626	90.60%	9.31%
F03_1_2	23494266	23203756	290507	98.76%	1.24%
F03_1_3	23338098	22007550	1330548	94.30%	5.70%
WOF_2_1	26 335 872	21 449 228	4 886 044	81.45%	18.44%
WOF_2_2	24 576 522	18 125 435	6 451 087	73.75%	26.25%
WOF_2_3	26 174 770	12 166 712	14 008 058	46.48%	53.52%
F04_2_1	26 845 480	24 703 921	2 141 559	92.02%	7.98%
F04_2_2	20 416 000	18 872 721	1 543 279	92.44%	7.56%
F04_2_3	21 010 010	20 157 316	852 694	95.94%	4.06%
F03_2_1	22 586 338	21 068 107	1 518 231	93.28%	6.72%
F03_2_2	20 988 152	19 417 108	1 571 044	92.51%	7.49%
F03_2_3	25 668 290	23 780 375	1 887 915	92.64%	7.36%

Table 4 Pourcentage de ARNr et ARN pour l'analyse 1 et 2 des différents réplicas pour les conditions WOF, F03 et F04 collecté à 12h

IV. Résultats :

A. Assemblage du génome :

Préalablement à ce travail, un premier assemblage d'une librairie de séquençage Illumina de la souche S756-078 de *C. difficile* a été obtenu en utilisant SPAdes. SPAdes est un logiciel d'assemblage utilisant les séquences « short reads » provenant d'Illumina. Il utilise des « reads » de haute qualité mais de courtes séquences. Cependant, le génome assemblé obtenu comportait encore plusieurs contigs non joints et n'avait pas pu être circularisé. Afin d'obtenir un génome circularisé et complet, une seconde librairie de séquençage a été obtenue avec de long fragments d'ADN (Séquençage MinION).

Le point de départ de notre travail a été de réaliser des assemblages hybrides à l'aide de Tricycler et Unicycler. L'avantage des assemblages hybrides est de pouvoir combiner les qualités des deux techniques de séquençage. En effet, la technologie Illumina permet d'obtenir de courtes séquences avec peu d'erreurs contrairement aux longues séquences « error-prone » de la technologie MinION. La combinaison de ces deux techniques permet la circularisation du génome avec un taux d'erreur moins élevé.

Unicycler est un assembleur utilisant les séquences « short reads » pour l'assemblage hybride par polishing avec les « long read MinION ». C'est-à-dire que les séquences Illumina sont assemblées par des ponts formés de séquences long « reads » MinION. L'utilisation des « short reads » Illumina a permis d'obtenir un assemblage de haute qualité circularisé grâce aux séquences MinION. Tricycler est une version plus récente d'Unicycler, reposant sur un assemblage avec des longues séquences et un polishing utilisant une librairie de reads courts. Nous avons voulu comparer les deux outils car Tricycler est plus exigeant en termes de couverture, du fait qu'il travaille au départ sur un échantillon de la librairie d'origine.

Grâce à Tricycler, nous obtenons un génome circularisé formé d'un seul chromosome d'une longueur de 4071782 bp avec un pourcentage G-C de 29,01%. La souche S756-078 ne possède pas de plasmide (table 2).

Une fois l'assemblage terminé, nous avons utilisé dans un premier temps le service d'annotation automatique du site BV-BRC pour les assemblages Unicycler et Tricycler. Le but était de comparer ces nouvelles annotations avec celle disponible pour l'assemblage SPAdes (résultats antérieurs). Les résultats de l'annotation sont illustrés dans la table 2.

Lorsque nous comparons ces trois assemblages, nous remarquons l'apparition de 30 kb (Table 2) par rapport à SPADES. En moyenne, un gène bactérien équivaut à 1kb, Tricycler a permis

librairie	nombre de "reads" total	"read" mappés	"read" non mappés	% de mapping
WOF_1_1	9 324 016	9 120 908	203 108	97.80%
WOF_1_2	6 947 503	6 813 308	134 195	98.10%
WOF_1_3	6 277 514	6 146 555	130 959	97.90%
F04_1_1	6 763 482	6 486 261	277 221	95.90%
F04_1_2	6 142 318	5 899 530	242 788	96.00%
F04_1_3	6 976 688	6 743 036	233 652	96.70%
F03_1_1	10 429 019	3 148 413	7 280 606	30.20%
F03_1_2	136 408	116 681	19 727	85.50%
F03_1_3	655 493	625 476	30 017	95.40%
WOF_2_1	2 400 506	2 339 726	60 780	97.50%
WOF_2_2	3 156 932	3 106 078	50 854	98.40%
WOF_2_3	6 938 404	6 853 020	85 384	98.80%
F04_2_1	1 055 453	855 453	200 000	81.10%
F04_2_2	759 010	328 578	430 432	43.30%
F04_2_3	416 769	261 993	154 776	62.30%
F03_2_1	743 614	263 802	479 812	35.50%
F03_2_2	768 419	267 650	500 769	34.80%
F03_2_3	928 091	358 314	569 777	38.60%

Table 5 Pourcentage de reads mappés sur notre génome de référence pour l'analyse 1 et 2 des différents réplicas pour les conditions WOF, F03 et F04 collecté à 12h.

librairie	Alignement total	Alignements assignés	% de couverture
WOF_1_1	18 648 032	16 795 700	90.10%
WOF_1_2	907 966	814 400	89.70%
WOF_1_3	12 555 028	11 488 567	91.50%
F04_1_1	13 526 964	12 040 216	89.00%
F04_1_2	12 284 636	10 840 800	88.20%
F04_1_3	13 953 376	12 459 707	89.30%
F03_1_1	20 858 038	6 128 053	29.40%
F03_1_2	272 816	204 460	74.90%
F03_1_3	1 310 986	1158 281	88.40%
WOF_2_1	4801 012	4 447 544	92.0%
WOF_2_2	6 313 864	5 896 672	93.4%
WOF_2_3	13 876 808	12 959 424	93.4%
F04_2_1	2 110 906	1 527 170	72.3%
F04_2_2	1518020	593 695	39.1%
F04_2_3	833 538	450 401	54.0%
F03_2_1	1 487 228	503 894	34.0%
F03_2_2	1 536 838	508 823	33.10%
F03_2_3	1 856 182	655 166	35.30%

Table 6 Pourcentage de "reads" identifié par assignation des alignements

d'améliorer la séquence avec l'apparition de plus ou moins 30 gènes. Pour continuer, l'avantage d'utiliser les séquences « long read » obtenues par MinION par rapport à l'assemblage Unicycler, est l'apparition de séquences difficiles à séquencer comme des transposons (séquence d'ADN mobile composé d'une répétition de nucléotides).

Le nombre de séquences codantes (CDS) détectées est assez semblable entre les assemblages Unicycler et Trycycler. Par contre, ce nombre est significativement plus faible que pour l'assemblage SPAdes. Après analyse de ces différences, il s'avère que dans la majorité des cas, ce nombre plus faible de CDS provienne de la plus grande fiabilité de l'assemblage : diminution de CDS dédoublées dans des contigs non joints et diminution d'erreur de cadre de lecture aboutissant à la création de CDS supplémentaires.

Sur base de la taille totale la plus grande et du nombre de CDS le plus cohérent, nous avons choisi l'assemblage Trycycler pour la publication sur Genbank. Dès lors, nous avons utilisé le service d'annotation automatique Prokaryotic Genome Annotation Pipeline du NCBI (Li et al. 2021) et publié le génome avec l'accession suivante : NZ_CP107552.1. La figure 4 ci-jointe illustre le génome séquencé.

B. Analyse transcriptomique :

1. Préambule :

Préalablement, nous avons réalisé des expériences de « transwells » dans le but de récupérer les ARNm exprimés par *C. difficile* en absence d'un voisinage bactérien (WOF) et en présence de la matière fécale 3 ou 4 (F03 ou F04). Plusieurs expériences ont été réalisées : 1 analyse au temps 6 h, 2 analyses au temps 12 h et 1 analyse au temps 24 h. Dans ce manuscrit, nous nous sommes concentrés sur les ARNm extraits en 12 h nommées analyse 1 et analyse 2. Après conversion en ADNc, ils ont été séquencés par Illumina seq.

2. Traitements des librairies :

Les librairies brutes obtenues par Illumina ont été traitées par fastp et sortmeRNA (Table 3&4).

3. Obtention des tables de comptages avec featureCounts :

Grâce au mapping sur notre génome de référence, la majorité des gènes séquencés ont pu être identifiés. La majorité de nos librairies ont un pourcentage de mapping satisfaisant excepté F03 dont le mapping reste faible sur plusieurs répliques (Table 5)

A partir de cette identification, featureCounts a déterminé le taux d'expression de chaque gène par comptage. Cela nous a permis d'obtenir des tables de comptage reprenant le nombre de

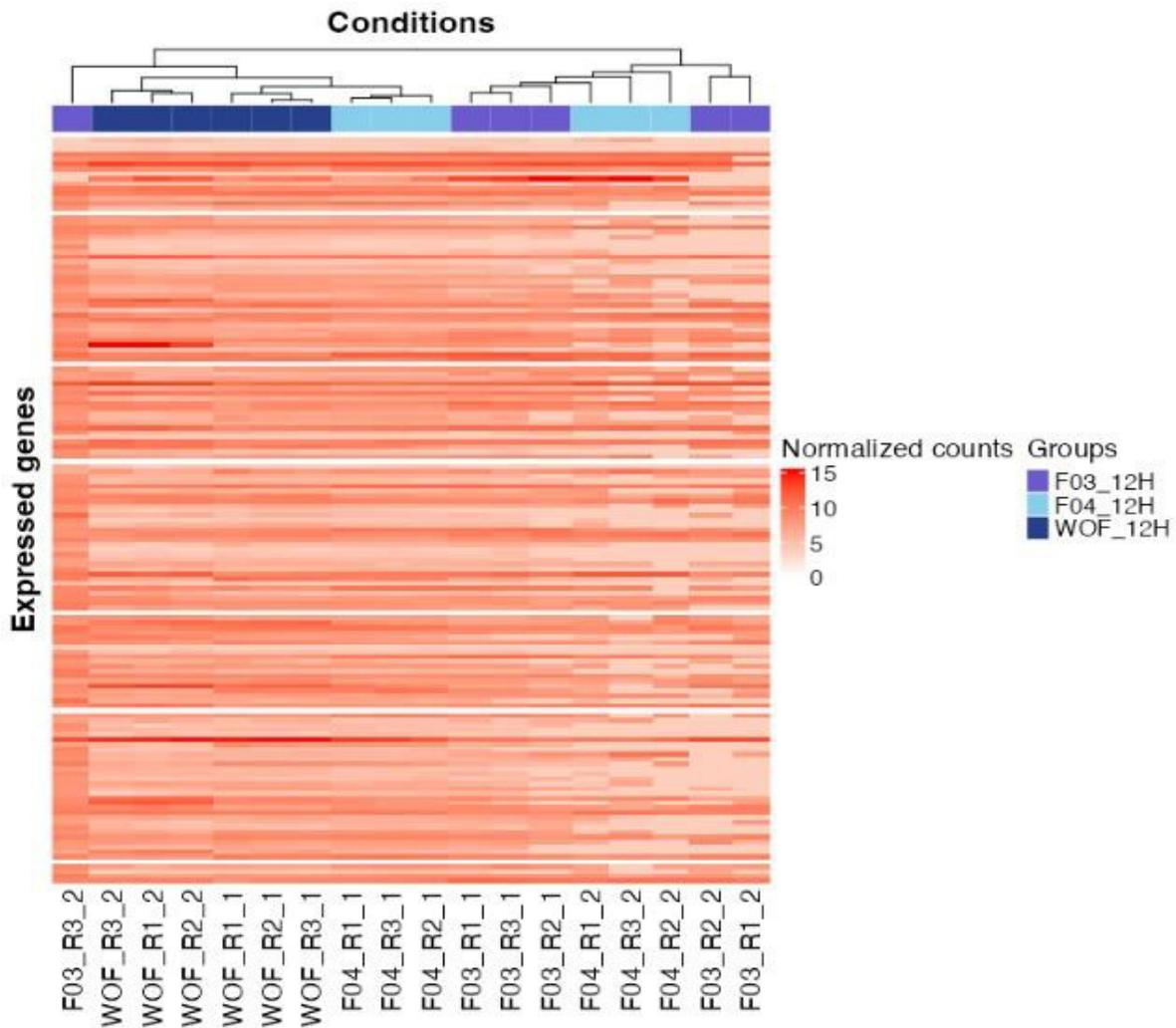


Figure 2 : Heatmap de l'expression globale des gènes de *Clostridioides difficile*. Chaque colonne représente une librairie de cDNA séquencés et l'axe y correspond au génome de *Clostridioides difficile* dont chaque rangée représente un gène/un CDS. Les couleurs du graphique indiquent le taux d'expression normalisé des gènes : plus la couleur est rouge plus l'expression normalisé par Deseq2 est élevée. Le bandeau de couleur associe les librairies à une condition environnemental.

reads identifiés dans chaque condition des deux analyses en 12 h. Pour F03, nous observons également un pourcentage de couverture faible (Table 6)

4. Analyse DESEQ2 :

Deseq2 est un logiciel conçu pour réaliser des analyses différentielles dans le but de mettre en évidence des différences significatives à partir des tables de comptages de deux conditions. Dans notre cas, Deseq2 a généré des « log2 fold change » équivalant au ratio d'expression du gène entre deux conditions. Trois comparaisons ont été générées à partir des tables de comptage : F03 vs WOF, F04 vs WOF et F04 vs F03.

5. Comparaison du taux d'expression :

a) Expression globale des gènes :

Par le « heatmaps » (figure 4), nous avons pu mettre en évidence l'expression globale des gènes de *Clostridioides difficile* dans différentes conditions environnementales au temps 12 h. Nos trois groupes sont la condition sans matière fécale (WOF-Without Feces), avec matière fécale 3 (F03-Feces 03) et avec matière fécale 4 (F04-Feces 04).

Le heatmap permet la visualisation des taux d'expression normalisés par Deseq2 pour les CDS du génome de *Clostridioides difficile*. Sur l'axe des X se trouvent nos trois groupes et leurs réplicas pour les échantillons collectés à 12h. Les rangées correspondent à un gène/CDS et sont ordonnées en fonction du chromosome de *Clostridioides difficile*. Le taux d'expression des gènes est représenté par un gradient de couleur, plus le rouge est foncé plus l'expression est élevée. Par un bandeau de couleur, les différentes librairies de séquençages sont associées à leur condition environnementale. Un regroupement des colonnes, représenté par un dendrogramme, a été obtenu par la méthode euclidienne ou complète (complete linkage). Grâce à ce regroupement, permettant de connecter les librairies dont l'expression globale des gènes est similaire, nous pouvons conclure que l'absence de bactéries compétitrices affecte l'expression des gènes de *Clostridioides difficile* de manière identique puisque les réplicas WOF sont regroupés ensemble. En revanche, les réplicas des échantillons avec matières fécales sont clustérisés entre réplicas d'une même expérience mais les expériences sont peu répétables entre elles.

Nous pouvons également conclure que l'introduction de bactéries compétitrices affectent de manière globale l'expression des gènes. En effet, des changements de taux d'expression sont observés sur toute la longueur du chromosome.

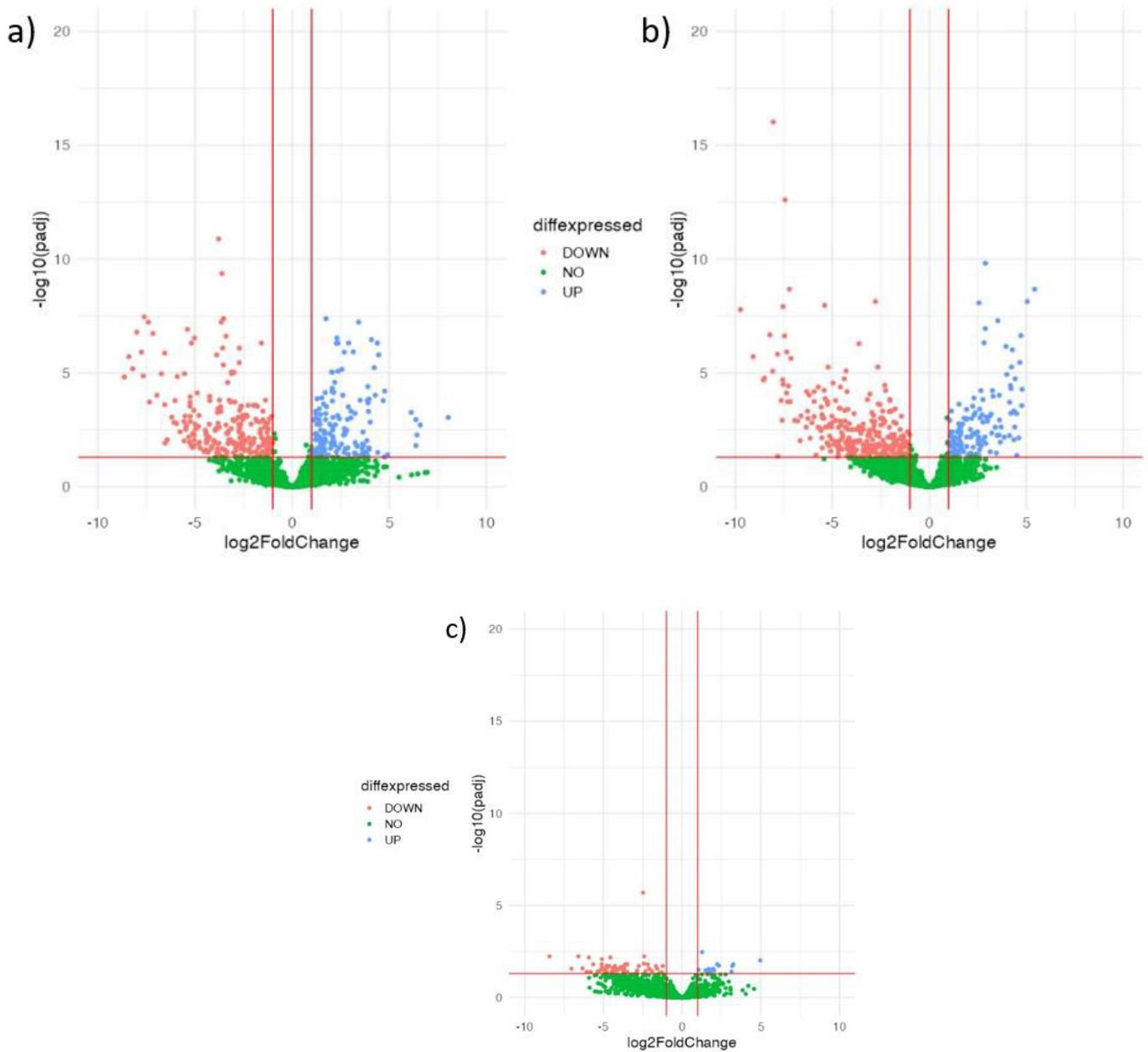


Figure 6 Volcano plot de l'analyse différentielle d'abondance de l'expression des gènes à 12h. L'axe des X montre le logarithme du delta d'expression entre la condition contrôle et la condition testée. L'axe Y représente la valeur négative du logarithme en base 10 de la p-valeur. Les gènes sur-exprimés (par rapport à la condition) à droite ou sous-exprimés (par rapport à la condition) à gauche et qui passent également le seuil de la p-valeur de 0,05 sont respectivement coloré en bleu si le $\log_2\text{FoldChange}$ est plus grand que 1 ou en rouge si le $\log_2\text{FoldChange}$ est plus petit que -1. Les gènes dont le $\log_2\text{foldchange}$ ne dépasse pas le seuil sont représentés en vert. Cela vaut également pour les gènes dont la p-valeur ne dépasse pas le seuil de signifiante. Les seuils sont représentés par une ligne rouge. a) Volcano plot de l'analyse différentielle d'abondance de l'expression des gènes entre la condition sans matière fécale (WOF) et avec matière fécale 3 (F03) à 12h, b) Volcano plot de l'analyse différentielle d'abondance de l'expression des gènes entre la condition sans matière fécale (WOF) et avec matière fécale 4 (F04) à 12h, c) Volcano plot de l'analyse différentielle d'abondance de l'expression des gènes entre la condition matière fécale 3 (F03) et avec matière fécale 4

b) *Changement d'expression significatif entre deux conditions :*

L'entièreté des ADNc séquencés de chaque condition ont subi une analyse différentielle de l'abondance de l'expression des gènes à 12 h avec Deseq2. La moyenne des expressions des différents gènes de chaque réplica est utilisée pour réaliser l'analyse statistique. Le résultat donne le « \log_2 fold change » de l'expression des gènes en fonction de la condition contrôle et de la condition testée. Par un volcano plot (figure 5), la signifiante est représentée en Y par le \log_{-10} de la p -valeur. Cela signifie que plus la p -valeur est faible, plus le point sera haut sur l'axe. L'axe des X reprend le « \log_2 fold change » entre la condition contrôle et la condition testée. Les gènes dont le \log_2 du « fold change » dépasse le seuil de 1 et dont la p -valeur est plus grande que 0,05 sont représentés en bleu. Les gènes dont le « \log_2 fold change » dépasse le seuil de -1 et dont la p -valeur est plus grande que 0,05 sont représentés en rouge. Les gènes dont le « \log_2 fold change » est non-significatif ou qui ne dépassent pas le seuil de changement d'expression de 1 et -1 sont représentés en vert. En présence de F03 par rapport à WOF, 12,45 % des gènes de *C. difficile* ont un changement d'expression significatif. En présence de F04 par rapport à WOF, 13,63 % des gènes de *C. difficile* ont un changement d'expression significatif. En présence de F04 par rapport à F03, 1,98 % des gènes de *C. difficile* ont un changement d'expression significatif.

Nous pouvons noter que l'ajout de matières fécales provoque un changement d'expression de même tendance pour les deux matières fécales par rapport à l'environnement seul. Par ailleurs, grâce à la comparaison entre les deux conditions testées, on observe un faible pourcentage de gènes dont l'expression change de manière significative.

En somme, l'introduction de bactéries compétitrices dans l'environnement de *C. difficile* affecte de manière globale l'expression de ces gènes. Cependant, le comportement de *C. difficile* ne semble pas changer de manière fondamentale entre les deux matières fécales.

C. *Enrichissement par FeedMaurice :*

L'enrichissement par le script FeedMaurice a permis de sortir différentes informations. Pour commencer, nos analyses ont mis en évidence 14 modules significatifs (Table 7). Il s'agit de modules composés de gènes dont le « \log_2 fold changes » est significatif et supérieur à 1 ou inférieur à -1. La majorité de ces modules jouent un rôle dans la biosynthèse, notamment des acides-aminés. Cette observation est également valable en présence de matières fécales 3 et/ou 4.

Pour continuer, quand nous nous intéressons aux modules significatifs dont la médiane des « \log_2 fold change » est supérieure à 1 ou inférieure à -1 obtenus lors de la comparaison entre les deux matières fécales, 29 modules ont un changement d'expression significatif (Table 8). Cependant, il faut mettre en évidence que la plupart de ces modules ne sont pas complets. C'est-à-dire que *C.*

KEGG Module	Log2 fold change F03vsWOF	Log2 fold change F04vsWOF	Fonction	Nombre d'éléments du module présent
M00019	-1,15	-1,62	Biosynthèse valine/isoleucine	2/4
M00026	-1,36	-1,95	Biosynthèse histidienne	5/6
M00061	1,09	-1,15	Dégradation du D-glucuronate	2/5
M00088	-1,16	-2	Biosynthèse du corps cétonique	2/5
M00095	-2,08	-1,23	Biosynthèse isoprénoïde en C5	1/7
M00140	-1,3	-1,01	Interconversion unité C1	3/3
M00373	-2,08	-1,23	Voie éthylmalonylique	1/12
M00432	-2,73	-2,17	Biosynthèse leucique	3/3
M00535	-2,71	-2,42	Biosynthèse de l'isoleucine	2/3
M00554	-1,12	-1,47	Biosynthèse du sucre nucléotidique	1/2
M00565	-1,41	-1,88	Biosynthèse du tréhalose	3/6
M00849	-2,08	-1,23	Biosynthèse isoprénoïde en C5	1/6
M00883	-2,29	-1,3	Biosynthèse de l'acide lipoïque	1/3
M00896	1,45	-1,12	Biosynthèse de thiamine	2/4

Table 7 Tableau reprenant les Modules significatif dont la médiane des foldchange des gènes ID qui font partis de ce module est plus élevé que 1 ou moins élevé que -1 entre la condition matière fécale F03 avec la condition matière fécale F04 par rapport à la condition contrôle WOF.

KEGG Module	F04vsF03 Log2 fold change	Fonction	Complétude
M00008	-2,53	Voie Entner-Doudoroff	1/4
M00020	1,26	Biosynthèse de la sérine	1/3
M00021	-1,79	Biosynthèse de la cystéine	2/2
M00027	-1,29	Dérivation GABA (gamma-aminobutyrate)	1/3
M00028	-1,74	Biosynthèse de l'ornithine	4/4
M00035	-1	Dégradation de la méthionine	1/4
M00046	-2,6	Dégradation de la pyrimidine	1/3
M00061	-2,23	Dégradation du D-glucuronate	2/5
M00114	-1,16	Biosynthèse de l'ascorbate	1/8
M00119	-2,38	Biosynthèse du pantothénate	3/5
M00125	-2,13	Biosynthèse de la riboflavine	6/7
M00127	-2,21	Biosynthèse de thiamine	6/7
M00133	-1,59	Biosynthèse des polyamines	4/4
M00140	1,33	Interconversion unité C1	3/3
M00141	-2,43	Interconversion unité C1	1/3
M00368	-1	Biosynthèse de l'éthylène	1/3
M00377	1,25	Voie acétylCoA réductrice (voie Wood-Ljungdahl)	7/7
M00550	-1,4	Dégradation des ascorbates	1/5
M00609	-1	Biosynthèse de la cystéine	3/6
M00631	-1,49	Dégradation du D-galacturonate	3/5
M00876	-3,46	Staphyloferrine Une biosynthèse	1/3
M00878	1,02	Dégradation du phénylacétate	6/9
M00895	-2,21	Biosynthèse de la thiamine	1/8
M00896	-2,57	Biosynthèse de la thiamine	2/4
M00897	-1,46	Biosynthèse de la thiamine	2/5
M00899	-2,58	Voie de récupération de la thiamine	5/5
M00913	-1,82	Biosynthèse du pantothénate	4/5
M00914	-1,25	Coenzyme A biosynthèse	3/7
M00953	-1	Biosynthèse de l'acide muginéique	1/5

Table 8 : Tableau reprenant les Modules significatif dont la médiane des foldchange des gènes ID de ce module est plus élevé que 1 ou moins élevé que -1 entre la condition matière fécale F03 (contrôle) et la condition matière fécale F04.

KEGG Orthologue (KO)	Log2 fold change F03vsWOF	Log2 fold change F04vsWOF	Fonction
K04028	4,76	4,79	Utilisation de l'éthanolamine protéine EutN
K04019	4,69	4,37	Utilisation de l'éthanolamine protéine EutA
K04023	4,39	4,28	Ethanolamine transporteur EutH
K04027	4,22	4,72	Utilisation de l'éthanolamine protéine EutM
K04030	4,08	3,96	Utilisation de l'éthanolamine protéine EutQ
K03735	4,01	4,36	Éthanolamine ammoniac-lyase, grande sous-unité B
K04032	3,96	4,67	Éthanolamine utilisation cobalamine adénosyltransférase EutT
K03736	3,91	4,43	Éthanolamine ammoniac-lyase, petite sous-unité C
K00132	3,89	4,24	acétaldéhyde déshydrogénase (acétylation) ->EutE
K04026	3,86	4,19	Utilisation de l'éthanolamine
K00248	3,02	3,3	Butyryl-CoA déshydrogénase
K20265	2,91	2,3	Glutamate:GABA antiporteur
K04031	2,85	2,77	Utilisation de l'éthanolamine protéine EutS
K20026	2,81	3,19	(R) -2-hydroxyisocaproyl-CoA déshydratase bêta-unité hadC
K23876	2,69	2,94	(R) -2-hydroxyacyl-CoA déshydraté activant ATPase
K15024	2,47	4,41	Phosphotransacetylase putative
K03739	2,43	2,6	Protéine membranaire impliquée dans l'exportation de D-alanine
K18887	2,36	2,89	Cassette de liaison ATP, sous-famille B, pompe à efflux
K11070	2,3	1,02	Spermidine/putrescine transport system permease protein
K18888	2,28	2,56	Cassette de liaison ATP, sous-famille B, pompe à efflux
K03617	2,22	2,13	Ferredoxine translocante H ⁺ /Na ⁺ :NAD ⁺ sous-unité A
K02346	2,19	2,48	ADN polymérase IV
K03614	2,14	1,72	Ferredoxine translocante H ⁺ /Na ⁺ :NAD ⁺ sous-unité D
K03612	2,12	1,8	Ferredoxine translocante H ⁺ /Na ⁺ :NAD ⁺ sous-unité G
K01005	2,11	1,79	Acide polyisoprényl-teichoïque--acide peptidoglycane teichoïque transférase
K16923	2,11	-1,61	Facteur de couplage énergétique
K03430	2,1	2,24	2-aminoéthylphosphonate-pyruvate transaminase
K03613	2,07	1,61	Ferredoxine translocante H ⁺ /Na ⁺ :NAD ⁺ sous-unité E
K03616	2,06	1,58	Ferredoxine translocante H ⁺ /Na ⁺ :NAD ⁺ sous-unité B
K06901	2,04	2,23	Adénine/guanine/hypoxanthine perméase
K03305	2,03	1,33	Transporteur d'oligopeptides dépendant des protons POT
K07335	1,75	1,14	Protéine de base de la membrane A et protéines connexes
K02011	1,72	1,24	Protéine perméase du système de transport fer(III)
K03615	1,7	1,22	Ferredoxine translocante H ⁺ /Na ⁺ :NAD ⁺ sous-unité C
K01271	1,68	1,93	Xaa-Pro dipeptidase
K02050	1,58	-5,94	Protéine de perméase du système de transport familial NitT/TauT

K02110	1,52	1,63	F-type H ⁺ -transport ATPase sous-unité c
K02027	1,41	1,14	Protéine de liaison du substrat à plusieurs systèmes de transport du sucre
K03498	1,36	1,35	Protéine d'absorption de potassium du système trk/ktr
K12942	1,34	1,6	Protéine de transport aminobenzoyl-glutamate
K03216	1,27	1,28	ARNt (cytidine/uridine-2'-O-)-méthyltransférase
K03499	1,15	1,26	Protéine d'absorption de potassium du système trk/ktr
K00615	1,11	1,02	Transcétone
K09762	-1	-1,11	Protéine de division cellulaire WhiA
K01883	-1,21	-1,44	Cystéine-ARNt synthétase
K03086	-1,26	-1,63	Facteur sigma primaire de polymérase de l'ARN
K05592	-1,28	-1,15	Hélicase ARN DeaD dépendant de l'ATP
K03168	-1,29	-1,44	ADN topoisomérase I
K00876	-1,38	-1,37	Uridine kinase
K00928	-1,43	-1,19	Aspartate kinase
K03177	-1,47	-1,36	ARNt pseudouridine55 synthase
K02343	-1,65	-1,42	ADN polymérase III sous-unité gamma/tau
K03797	-1,72	-1,96	Protéase de traitement terminal carboxylique
K22405	-1,79	-1,5	NADH oxydase (H ₂ O-formation)
K22278	-1,8	-4,27	Peptidoglycane-N-acétylglucosamine deacétylase
K06888	-1,85	-1,75	Protéine non caractéristique
K06436	-1,99	-2,37	Spore tégument
K07029	-1,99	-2,03	Diacylglycérol kinase (ATP)
K00975	-2,16	-1,81	Glucose-1-phosphate adényltransférase
K01693	-2,22	-1,94	Imidazoleglycérol-phosphate déshydratase
K00198	-2,24	-1,87	Sous-unité catalytique de déshydrogénase de CO anaérobie
K01439	-2,26	-3,26	Succinyl-diaminopimelate desuccinylase
K02437	-2,32	-1,57	Système de clivage glycine protéine H
K01448	-2,43	-3,67	N-acétylmuramoyl-L-alanine amidase
K00688	-2,45	-2,24	Glycogène phosphorylase
K00929	-2,46	-2,64	Butyrate kinase
K02117	-2,53	-2,28	V/A-type H ⁺ /Na ⁺ -transporter ATPase sous-unité A
K02123	-2,6	-2,27	V/A-type H ⁺ /Na ⁺ -transporter ATPase sous-unité I
K02051	-2,64	-7,52	Protéine de liaison de substrat du système de transport familial NitT/TauT
K02118	-2,64	-2,61	V/A-type H ⁺ /Na ⁺ -transporter ATPase sous-unité B
K07171	-2,64	-2,42	ARNm interférase MazF
K07658	-2,71	-1,83	Système à deux composants, famille OmpR, régulateur de réponse de synthèse de phosphatase alcaline PhoP
K01938	-2,72	-1,56	Formate--tétrahydrofolate ligase
K06217	-2,74	-2,34	Phosphate, protéine inductible de famine PhoH et protéines connexes
K00832	-2,75	-2,81	Acide aminé aromatique transaminase
K11202	-2,89	-2,6	Système PTS de type fructose composant EIIB

K04758	-4,72	-4,68	Protéine de transport de fer ferreux A
K01118	-4,8	-3,7	NADH-azoréductase dépendant de la FMN
K06397	-4,82	-4,07	Protéine de sporulation de stade III AH
K04564	-5,03	-6,62	Superoxyde dismutase, famille Fe-Mn
K05337	-5,03	-5,4	Ferredoxine
K07217	-5,11	-5,35	Catalase contenant du mn
K04769	-5,25	-5,57	Régulateur transcriptionnel de la famille AbrB, protéine de sporulation de stade V T
K01449	-5,26	-4,33	N-acétylmuramoyl-L-alanine amidase
K01989	-5,38	-6,4	Tryptophane putatif/tyrosine, protéine de liaison du substrat
K06410	-5,56	-7,64	Sous-unité de synthase dipicolinate A
K06398	-5,95	-4,75	Sporulation de la protéine A de stade IV
K06411	-6,03	-6,19	Sous-unité de synthase dipicolinate A
K00433	-6,56	-5,36	Chloroperoxydase non hémique
K03091	-6,96	-7,13	Facteur sigma spécifique à la sporulation de l'ARN polymérase
K02768	-7,16	-5,22	Composant EIIA du système fructose PTS
K06333	-7,34	-9,08	Protéine spore coat JB
K06419	-7,61	-7,54	Petite protéine B soluble dans l'acide
K06334	-7,75	-8,21	Protéine spore coat JB

Table 9 Tableau reprenant les numéro KO significatif dont la médiane des "Log2 fold change" des gènes ID de ce numéro KO est plus élevé que 1 ou moins élevé que -1 lors de la comparaison entre la condition sans matière fécale (contrôle) et la condition F03 et F04

K07723	-2,99	-2,54	Famille CopG régulateur transcriptionnel / antitoxine EndoAI
K02119	-3	-2,01	V/A-type H ⁺ /Na ⁺ -transport ATPase sous-unité C
K02120	-3,03	-2,72	V/A-type H ⁺ /Na ⁺ -transport ATPase sous-unité D
K04517	-3,05	-2,96	Déshydrogénase préphénate
K03671	-3,11	-1,9	Thioredoxine 1
K03294	-3,14	-2,57	Acide aminé de base/polyamine antiporteur, famille APA
K07636	-3,16	-2,79	Système à deux composants, famille OmpR, capteur de régularisation de phosphate histidine kinase PhoR
K00179	-3,17	-2,96	Indolepyruvate ferredoxine oxydaoreductase, sous-unité alpha
K02238	-3,26	-3,01	Protéine de compétence ComEC
K19504	-3,3	-7,21	Glucoselysine-6-phosphate deglycase
K10670	-3,32	-1,79	Composant complexe glycine/sarcosine/bêtaïne réductase A
K01736	-3,35	-3,41	Chorismate synthase
K00800	-3,4	-3,08	3-phosphoshikimate 1-carboxyvinyltransferase
K10671	-3,4	-2,41	Composant complexe de glycine réductase B sous-unité alpha et bêta
K01738	-3,46	-1,55	Cystéine synthase
K21577	-3,53	-2,06	Glycine/sarcosine/bêtaïne réductase composant C sous-unité bêta
K14170	-3,54	-3,07	Chorismate mutase / préphénate déshydratase
K01258	-3,58	-3,62	Tripeptide aminopeptidase
K11063	-3,59	-3,35	Toxine A/B
K07258	-3,6	-4,9	Type sérine D-Ala-D-Ala carboxypeptidase (protéine 5/6 liant la pénicilline)
K10672	-3,65	-2,28	Tomposant complexe glycine réductase B sous-unité gamma
K03856	-3,74	-4,29	3-désoxy-7-phosphoheptulonate synthase
K01735	-3,75	-3,53	3-dehydroquinone synthase
K03790	-3,79	-5,52	[protéine ribosomale S5]-alanine N-acétyltransférase
K05810	-3,82	-2,92	Purine-nucleoside/S-méthyl-5'-thioadenosine phosphorylase / adénosine deaminase
K05832	-3,84	-5,43	Tryptophane putatif/tyrosine, protéine perméase du système de transport
K01775	-3,89	-3,19	Alanine racémase
K04759	-3,96	-5,42	Protéine de transport du fer ferreux B
K00891	-4,01	-3,2	Shikimate kinase
K03594	-4,23	-3,39	Bactérioferritine
K06283	-4,26	-5,94	Régulateur transcriptionnel de la famille DeoR putatif, protéine de sporulation de stade III D
K03798	-4,31	-1,03	Division cellulaire protéase FtsH
K05833	-4,4	-6,18	Système de transport putatif tryptophane/tyrosine protéine de liaison ATP
K06374	-4,43	-4,84	Protéine de maturation des spores B

KEGG Orthologue (KO)	Log2 fold change F04vsF03	Fonction
K00864	-6,33	Glycérol kinase
K21672	-6,06	2,4-diaminopentanoate déshydrogénase
K00145	-5,82	N-acétyl-gamma-glutamyl-phosphate réductase
K07010	-5,36	Glutamine amidotransférase putative
K02036	-5,21	Système de transport de phosphate protéine de liaison ATP
K03077	-4,99	L-ribulose-5-phosphate-4-épimérase
K19294	-4,9	O-acétyltransférase complexe alginate AlgI
K07260	-4,79	Zinc D-Ala-D-Ala carboxypeptidase
K19509	-4,79	Fructoselysine/glucoselysine PTS system EIID component
K05814	-4,77	Sn-glycérol 3-phosphate système de transport perméase protéine
K01783	-4,63	Phosphate ribulose 3-épimérase
K20859	-4,27	Phosphoribosyl 1,2-phosphate cyclique 1,2-diphosphodiesterase
K21071	-4,22	Phosphofructokinase dépendante de l'ATP/phosphofructokinase dépendante du diphosphate
K19504	-3,91	Glucoselysine-6-phosphate deglycase
K05832	-3,6	Tryptophane putatif/tyrosine, protéine perméase du système de transport
K01989	-3,48	Tryptophane putatif/tyrosine, protéine de liaison du substrat
K19505	-2,88	Régulateur transcriptionnel dépendant de sigma-54, activateur transcriptionnel d'opéron
K06400	-2,84	ADN recombinase
K16923	-2,43	Facteur de couplage énergétique composant spécifique du substrat du système de transport
K02777	-2,34	Sucre PTS système EIIA composant
K00600	-2,33	Glycine hydroxyméthyltransférase
K01940	-2,09	Argininosuccinate synthase
K11071	-1,91	Spermidine/putrescine transport system permease protein
K11069	-1,66	Spermidine/putrescine transport system substrate-binding protein
K05808	-1,58	Protéine de modulation sigma-54 putative
K11070	-1,28	Spermidine/putrescine transport system permease protein
K00198	1,04	Sous-unité catalytique de déshydrogénase de monoxyde de carbone anaérobie
K00196	1,27	Sous-unité anaérobie de soufre de fer et de monoxyde de carbone déshydrogénase
K00384	1,72	Thioredoxine réductase (NADPH)
K03687	1,77	Chaperon moléculaire GrpE
K03705	1,97	Répresseur transcriptionnel thermo-inductif
K03696	2,03	ATP-dependent Clp protease ATP-binding subunit ClpC
K01777	2,22	Racémase proline
K03307	3,19	Soluté:Na ⁺ symporter, famille SSS
K03695	3,25	ATP-dependent Clp protease ATP-binding subunit ClpB

Table 10 Tableau reprenant les numéro KO significatif dont la médiane des "Log2 fold change" des gènes ID de ce numéro KO est plus élevé que 1 ou moins élevé que -1 lors de la comparaison entre la condition F03 et la condition F04.

CDS	Log2 fold change F03vsWOF	Log2 fold change F04vsWOF	Fonction putative par similarité
J5O10_02860	3,50	5,05	Amidohydrolase
J5O10_09910	3,78	4,37	TIGR02536 famille éthanolamine
J5O10_09920	3,59	4,13	Protéine contenant du domaine BMC
J5O10_02240	3,24	3,61	Flavoprotéine de transfert d'électron, sous-unité bêta
J5O10_02225	2,66	3,12	R)-2-hydroxyisocaproyl-CoA déshydratase sous-unité HadB
J5O10_02245	2,71	3,01	Sous-unité alpha/FixB de flavoprotéine de transfert d'électrons
J5O10_04655	2,61	2,67	Protéine contenant le domaine du dicluster 4Fe-4S
J5O10_02215	2,70	2,52	Isocaprényl-CoA:2-hydroxyisocaproate CoA-transferase HadA
J5O10_11565	3,14	2,31	Protéine de la famille FUSC
J5O10_17110	2,06	2,15	Protéine de la famille de la nitroréductase
J5O10_16440	2,35	1,93	MATE famille efflux transporteur
J5O10_06255	2,09	1,65	Protéine de la famille des protéases du site 2
J5O10_03255	2,05	1,60	Transporteur d'échange CIC famille H(+)/Cl(-)
J5O10_07445	1,49	1,59	Polysaccharide désacétylase
J5O10_14470	1,37	1,54	Protéine de biosynthèse polysaccharidique
J5O10_12345	1,65	1,53	Protéine peptidoglycane
J5O10_14465	2,06	1,53	Peptidase de la famille C40
J5O10_14535	1,63	1,50	Protéine de fixation des parois cellulaires Cwp7
J5O10_06720	1,63	1,45	Transporteur de cations métalliques divalent
J5O10_11185	1,77	1,29	Protéine de la famille VanW
J5O10_15145	1,53	1,19	L,D-transpeptidase/protéine de liaison peptidoglycane
J5O10_08860	-1,51	-1,13	2-hydroxyacyl-CoA déshydratase
J5O10_06400	-1,83	-1,27	Sous-unité protéolytique de la protéase CLP dépendante d'ATP
J5O10_09000	-1,21	-1,34	FAD-dépendante oxydaoreductase
J5O10_16575	-2,35	-1,58	Protéine contenant le domaine de la méthyltransférase
J5O10_04135	-2,33	-1,83	NUDIX hydrolase
J5O10_06810	-2,59	-2,52	Protéine contenant le domaine de l'adénylation des AA
J5O10_16645	-3,68	-2,63	Protéine de la famille des polysaccharides désacétylase
J5O10_12170	-3,61	-2,65	Protéine de la famille GrdX
J5O10_04130	-1,87	-2,78	Protéine HEAT
J5O10_11655	-5,24	-3,57	Acide biliaire récepteur germinant pseudoprotease CspC
J5O10_09195	-3,24	-3,64	Shikimate déshydrogénase
J5O10_16695	-4,33	-3,92	Hydrogénase de fer
J5O10_14620	-4,04	-3,95	TIGR04086 famille de protéines membranaires
J5O10_12450	-3,84	-4,19	Protéine contenant un ruban de zinc
J5O10_03040	-2,85	-4,24	AAA family ATPase
J5O10_14850	-3,94	-4,60	Protéine de la famille de la rubrerythrine
J5O10_18330	-4,41	-4,68	Protéine de la famille Staygreen

J5O10_11360	-3,96	-5,02	N-acétylmuramoyl-L-alanine amidase
J5O10_16890	-5,52	-5,03	Protéine de liaison ATP
J5O10_16660	-4,37	-5,05	Protéine de la famille YncE
J5O10_08465	-5,12	-5,17	Protéine de la famille YfcC
J5O10_10990	-5,57	-5,29	Protéine de la famille des flavins réductase
J5O10_01205	-3,88	-5,43	Protease YyaC des spores
J5O10_09230	-6,56	-5,94	Protéine contenant du domaine cupin
J5O10_04700	-4,74	-6,05	Facilitateur de diffusion cation transporteur familial
J5O10_01620	-4,41	-6,32	Protéine de sporulation
J5O10_17085	-6,55	-6,66	Protéine de la famille 2 de la glycosyltransférase
J5O10_06965	-5,76	-7,16	Peroxydoxyxine
J5O10_16650	-6,73	-7,34	Protéine de spore soluble dans l'acide de type alpha/bêta
J5O10_17080	-6,08	-7,35	Exosporium glycoprotéine BclA3 de type collagène
J5O10_02975	-6,43	-7,58	« spore cortex-lytique germination » protéine SleC
J5O10_16785	-3,59	-7,825	Sous-unité IIC du PTS transporteur du sucre
J5O10_12400	-8,64	-8,492	Protéine CotJA associée au pelage des spores
J5O10_05205	-8,41	-9,74	Protéine morphogénétique de l'exosporium CdeC

Table 11 Tableau reprenant les gènes dont le foldchange significatif est plus élevé que 1 ou moins élevé que -1 lors de la comparaison entre la condition matière fécale F03 avec la condition matière fécale F04 par rapport à la condition contrôle WOF

difficile ne possède pas toutes les étapes composant le module. Il se peut donc que ces modules ne soient pas réalisés par *C. difficile*.

Lorsque nous analysons les KO significatifs dont la médiane des « \log_2 fold changes » des gènes correspondant au KO est supérieure à 1 ou inférieure à -1, nous pouvons ressortir plusieurs KO faisant référence à une même fonction lors de la comparaison entre la condition sans matière fécale et la condition en présence de matière fécale 3 ou 4 (Table 9). Lorsque l'on compare les conditions avec matières fécales, nous pouvons voir 32 numéros KO dont la médiane des « \log_2 fold change » est au-dessus du seuil (+1 et -1) de manière significative (Table 10).

L'enrichissement avec FeedMaurice a permis de mettre en évidence les gènes qui possédaient un numéro KO. Cependant, il faut prendre en compte les gènes sans numéro KO. A partir d'une liste extraite des fichiers, nous avons pu analyser manuellement ces gènes. 55 Gènes sans numéro KO ont un « \log_2 fold change » significatif au-dessus du seuil de 1 et en-dessous du seuil de -1 (Table 11). Lorsque l'on compare les deux conditions en présence de matières fécales, on peut voir 27 gènes sans numéro KO dont le \log_2 foldchange est au-dessus du seuil (+1 et -1) de manière significative (Table 12).

Grâce au heatmap, nous avons mis en évidence que l'introduction de bactéries compétitrices affecte de façon globale le comportement de *C. difficile*. En effet, différents mécanismes et métabolismes sont affectés par la présence d'un voisinage bactérien

1. Métabolisme :

Concernant le métabolisme, nous avons mis en évidence un déséquilibre d'expression des gènes liés au métabolisme de Stickland. En effet, nous retrouvons notamment une diminution significative du taux d'expression de certains gènes qui constituent le chemin oxydatif comme *iorA* (J5O10_12285). D'autres gènes comme *iorB* (J5O10_12280), *buk* (J5O10_12540) et *vorB* (J5O10_12555) sont également des gènes qui participent à l'oxydation montrent une diminution d'expression en présence de matière fécale de façon non significative. En revanche, plusieurs gènes du chemin réducteur montrent une augmentation d'expression en présence de matière fécale. En effet, nous retrouvons le gène *hadA* (J5O10_02215), une coenzyme A transférase qui permet de réapprovisionner le cycle réducteur en Coenzyme A (Hofmann et al. 2018; Neumann-Schaal, Jahn, et Schmidt-Hohagen 2019).

Pour continuer, Stickland est capable d'utiliser plusieurs acides aminés pour produire de l'énergie. C'est en fonction de leur potentiel rédox qu'ils seront oxydés ou réduits. Dans le branchement oxydatif, nous retrouvons la valine, l'isoleucine, le tryptophane et l'histidine et les acides aminés tel que la proline, la glycine et la méthionine sont réduits. Le changement métabolique est

également mis en évidence avec l'expression de certaines oxydoréductases. De ce fait, l'expression de certaines enzymes spécifiques aux acides aminés du chemin réducteur comme la proline-réductase et son gène *prdA* (J5O10_16620) qui augmente de façon significative en présence de F04 et de façon non-significative en présence de F03 contre une diminution du gène *grdB* codant pour la glycine réductase.

Non seulement, les gènes liés à la dégradation des acides aminés, augmente mais nous avons également mis en évidence une diminution des modules de biosynthèses des acides aminés notamment celui de la leucine (M00432) et isoleucine (M00535), des acides aminés utilisés par la partie oxydative de Stickland.

Le métabolisme de Stickland est couplé avec le complexe RNF dans le but de produire de l'énergie en passant par le chemin réducteur. En présence de matière fécale, les gènes codant pour le complexe RNF montrent une augmentation significative par rapport à leur expression en absence de bactéries. Par exemple, nous pouvons citer *rnfG* (J5O10_05455, K03612), *rnfD* (J5O10_05450, K03614), *rnfB* (J5O10_05470, K03616) dont l'expression augmente presque de 2x par rapport à la condition contrôle.

Pour continuer, *C. difficile* possède un autre chemin métabolique extrêmement conservé. Il s'agit du chemin métabolique de Wood-Ljungdahl (WLP). Lorsque nous cherchons l'expression des gènes composant le chemin métabolique Wood-Ljungdahl, il y a une diminution de l'expression en fructose (K02768) est significativement moins exprimé que lorsque *C. difficile* est seul. En outre, lorsque l'on compare l'expression des gènes codant pour des PTS, nous avons mis en évidence deux gènes dont l'expression est significativement différente en fonction de la matière fécale. Effectivement, le gène du système PTS de la fructoselysine/glucoselysine (K19509) et le gène du PTS du sucre (K02777) sont significativement moins exprimés dans la condition F04 par rapport à F03. Une fois les carbohydrates à l'intérieur de la bactérie, ils sont métabolisés en pyruvate et acétyl-CoA dans le métabolisme central. En présence de F03, *C. difficile* exprime significativement plus le GAP, le gène d'une enzyme de la glycolyse, que dans un environnement seul (Neumann-Schaal, Jahn, et Schmidt-Hohagen 2019).

C. difficile est capable d'utiliser plusieurs types de carbohydrates, notamment les tréhaloses (Neumann-Schaal, Jahn, et Schmidt-Hohagen 2019). Dans nos expériences, le module de la biosynthèse des tréhaloses est significativement moins exprimé en présence de matières fécales par rapport à la condition WOF.

Le fer est un élément essentiel à la croissance bactérienne, c'est pourquoi les bactéries possèdent de nombreuses enzymes et des transporteurs de hautes affinités pour le fer. *C. difficile* n'est pas une exception, nous retrouvons plusieurs de ces enzymes dans son métabolisme.

Dans le métabolisme central, l'expression des gènes de certaines ferrédoxines (K03612, K03613, K03614, K03615, K03616, K03617) est diminuée lorsque *C. difficile* se trouve dans un voisinage bactérien.

2. Virulence :

L'introduction de bactéries dans l'environnement de *C. difficile* modifie également l'expression des gènes liés à la virulence. Pour commencer, les gènes codant pour les toxines TcdA et TcdB sont significativement moins exprimés dans les deux conditions F0 par rapport à WOF.

En outre, la sporulation de *C. difficile* est également diminuée dans les conditions F03 et F04 par rapport à la condition contrôle. Par contre, concernant les gènes régulateurs, nos résultats ne montrent qu'une diminution significative de *spo0A* en présence de F04 tandis que la différence d'expression de *codY* est non-significative en présence de matière fécale.

C. difficile est une bactérie capable d'utiliser l'éthanolamine trouvé dans les membranes sous forme de phosphatidyléthanolamine. C'est grâce à l'opéron « *eut* » que l'éthanolamine peut être transformé en nitrogène et en acétyl-CoA. A partir de l'enrichissement par feedmaurice, des gènes de l'opéron éthanolamine ont pu être mis en évidence à partir du dictionnaire KO. Cet opéron est composé de 19 gènes régulés positivement en présence d'éthanolamine grâce aux gènes *eutV* et *eutW* mais aussi par inhibition de *sigF*, un facteur de sporulation. Le « log₂ fold change » de ces gènes ne changent pas de façon significative en présence de matière fécale. En revanche, nous pouvons observer une tendance non significative concordant avec l'activation de cet opéron puisque nous avons une augmentation positive de l'expression de *eutV* et *eutW* et une diminution de l'expression de *sigF* (J5O10_04045) en présence de matière fécale.

En outre, la métabolisation de l'éthanolamine débute par une transformation en ammoniac ou en acétaldéhyde grâce à l'éthanolamine ammonia lyase. Cette enzyme est codée par deux gènes *eutB* et *eutC* dont l'expression est augmentée de manière significative en présence de matière fécales (Roof et Roth 1989). Pour continuer, en plus de produire des composants volatiles, l'éthanolamine est métabolisée en produit toxique pour la cellule, l'acétaldéhyde. Pour éviter la dispersion de ses métabolites et protéger la cellule, *C. difficile* va exprimer des microcompartiments bactériens (BMC) (Pitts et al. 2012). Il s'agit d'une capsule servant à la fois de barrière semi-perméable mais aussi d'environnement chaperon contenant les enzymes nécessaires à la métabolisation de l'éthanolamine

Nom	Gène_ID	Basemean	F03-WOF	F04-WOF
<i>tcdB</i>	J5O10_03470	2331,12767	-3,36	-3,49
<i>tcdA</i>	J5O10_03485	4019,99587	-3,81	-3,27
<i>eutA</i>	J5O10_09865	1456,79455	4,69	4,37
<i>eutB</i>	J5O10_09870	2780,24709	4,01	4,36

Table 13 Choix des gènes sélectionnés pour la validation des résultats transcriptomiques

(Penrod et Roth 2006). Chez *C. difficile*, il a été montré que le gène *eutS* est notamment capable de former des micro-compartiments lorsqu'il est surexprimé (Pitts et al. 2012). En présence de bactéries, nos résultats montrent une augmentation significative de l'expression de *eutS* mais également de gènes codant pour des protéines composant la BMC (J5O10_09920).

C. difficile est capable de produire un biofilm (Meza-Torres et al. 2021). En présence de matières fécales, nos résultats montrent une augmentation de certains gènes liés au biofilm comme des polysaccharide désacétylases et des polysaccharides biosynthèses en présence de matières fécales comme J5O10_07445, J5O10_14470 liés aux polysaccharides mais aussi les gènes de transport de polyamines.

Pour continuer, les polyamines comme la putrescine et la spermidine sont reconnus pour offrir plusieurs bénéfices aux bactéries. En présence de matières fécales, nous pouvons observer une augmentation de certains gènes liés aux transports des spermidines et putriscines (K11070). D'une part, en présence de F03, il y a également une augmentation significative du gène codant pour un transporteur ABC des spermidines. D'autres part, trois numéros KO (K11071, K11069, K11070) correspondant au transport des spermidines sont exprimés significativement moins en présence de F04 par rapport à F03.

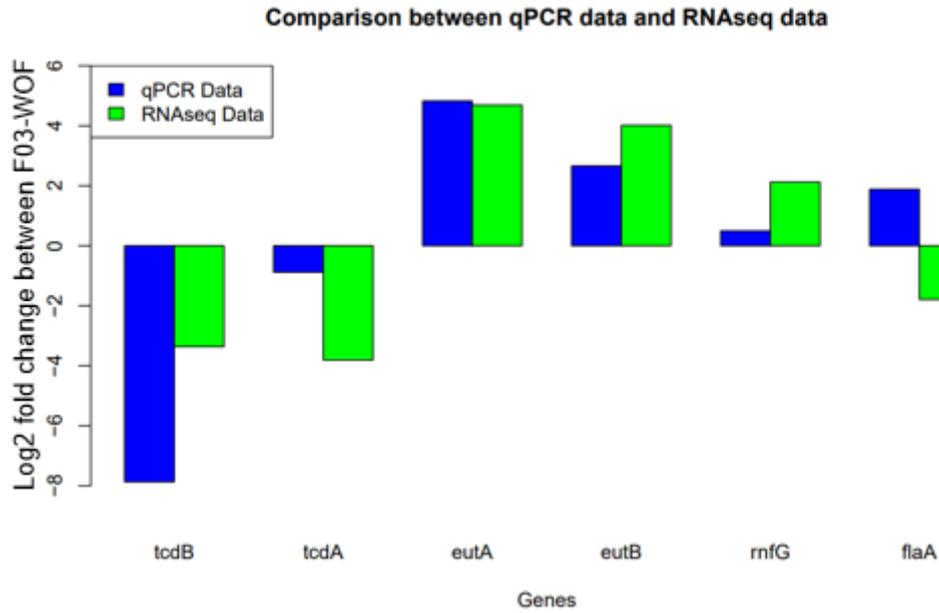
D. Validation par qPCR :

Afin de valider nos résultats transcriptomiques, nous avons réalisé des qPCR sur le cDNA des ARNm récoltés à 12 h. La qPCR nous permet de comparer la tendance des « \log_2 fold change » sur quelques gènes sélectionnés.

Dans cet objectif, nous avons sélectionné 6 gènes avec un ratio de changement d'expression significatif entre la condition testée et la condition contrôle. Ces gènes ont également été sélectionnés car leur nombre de reads séquencés moyens (« basemean ») est élevé afin de garantir une détection de leur ADNc en qPCR (Table 13).

Lorsque nous comparons les résultats de « \log_2 fold change » des delta-delta CT et du RNA-seq entre F03 et WOF, la tendance du « \log_2 fold change » est comparable entre les résultats obtenus en qPCR et les résultats obtenus en transcriptomique pour *tcdA*, *tcdB*, *eutA*, *eutB*, *rnfG* mais pas pour *flaA* (figure 6). Pour continuer, lorsque nous comparons le « \log_2 fold change » des delta-delta CT et du RNA-seq entre F04 et WOF obtenu par nos deux méthodes, la tendance est comparable pour les gènes *tcdB*, *eutA*, *eutB*, *rnfG* mais pas pour *tcdA* et *flaA* (figure 7).

a)



b)

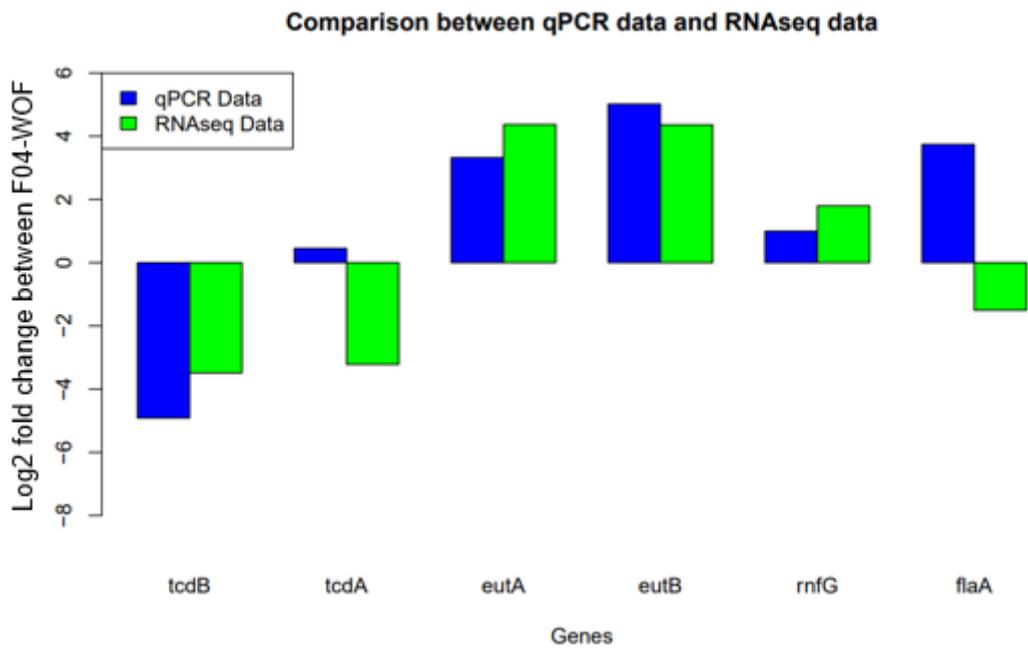


Figure 7 Comparaison des log2 fold change obtenue en qPCR et en RNAseq a) entre F03 et WOF b) entre F04 et WOF

V. Discussion :

La présence de *Clostridioides difficile* dans les selles n'est pas synonyme d'infection puisqu'elle est définie par la présence de la bactérie mais également par la présence des toxines et de diarrhées (Martinez, Taminiou, et al. 2022). Les principales responsables du développement de colites pseudomembraneuses sont les toxines ainsi que l'activation du système immunitaire. L'étape déterminante de la production de toxines est une colonisation aboutie. La colonisation de *C. difficile* semble être favorisée en cas d'une dysbiose hypothétisant que le comportement de *C. difficile* change en fonction de son environnement. Le propos de ce travail est d'étudier l'expression des gènes de *C. difficile* en fonction de condition de co-culture avec ou sans microbiote intestinal à travers une analyse transcriptomique.

Lors de la préparation des bibliothèques par Fastp et SortmeRNA, la condition F03 n'a pas montré un pourcentage de mapping satisfaisant. Cela se reporte également sur le pourcentage d'alignement obtenu par subread. En sachant que le filtre des « transwells » n'est pas complètement imperméable, des bactéries contaminantes pourraient se retrouver dans notre « pool » d'ARN séquencés. Il faut également noter que F03 provient d'un individu plus jeune que F04. Notre hypothèse est basée sur la différence des bactéries entre deux microbiotes. Effectivement, les bactéries contenues dans le microbiote d'un individu jeune sont potentiellement plus nombreuses et plus mobiles, ce qui pourrait expliquer une plus grande contamination du compartiment contenant *C. difficile*.

Les résultats obtenus montrent que l'introduction de bactéries compétitrices dans l'environnement de *C. difficile* affecte de manière globale l'expression de ces gènes. En effet, l'analyse DESeq2 nous révèle qu'entre la condition sans microbiote et avec microbiote (F03 ou F04) à 12 d'incubation, l'expression de plus de 10% des gènes de la bactérie est significativement altérée. Cependant, le comportement de *C. difficile* ne semble pas changer de manière fondamentale entre les deux matières fécales car le nombre de gènes significatifs est de l'ordre du 1%.

Un des premiers facteurs permettant la survie de *C. difficile* dans des environnements hostiles est sa capacité à sporuler. La sporulation chez *C. difficile* débute après 16 h à 24 h de colonisation. A 12 h, nous pourrions commencer à observer une augmentation du taux d'expression des gènes liés à la sporulation. Cependant, la comparaison entre la condition WOF et F03-F04 montre une diminution de l'expression des gènes liés à la sporulation. Une première hypothèse est qu'en présence de bactéries compétitrices, *C. difficile* pourrait avoir un retard de croissance par rapport à la population de la condition WOF. Or, la croissance de *C. difficile* dans les différentes conditions avait préalablement été mesurée et les résultats montrent qu'en 12h, *C. difficile* en présence de bactérie et en absence de

bactéries se trouve au même stade de croissance. Par conséquent, la diminution d'expression des gènes de sporulation n'est probablement pas liée à un retard de la croissance.

En outre, la sporulation de *C. difficile* est activée en cas d'un manque de nutriment. Notre deuxième hypothèse pourrait être que *C. difficile* trouve les nutriments dont il a besoin en présence de nos deux matières fécales testées. La sporulation est un processus régulé par de nombreux gènes, principalement *spoOA* (J5O10_05835) et *codY* (J5O10_06160). Ils sont à l'origine d'une cascade de réaction entraînant l'initialisation de la sporulation. Ces deux facteurs sont sensibles à la disponibilité de nutriment. En effet, *spoOA* est activé à la suite de sa phosphorylation médiée par des signaux. Le gène *codY* est quant à lui médié par la concentration en GTP et en acides aminés branchés. Ceux-ci diminuent lors d'une limitation en nutriments. Nos résultats n'ont pas mis en évidence une différence d'expression significative. Cependant, d'autres gènes en aval dans la cascade de régulation de la sporulation montrent une diminution d'expression.

Pour continuer, les toxines produites par *C. difficile* sont également régulées par la présence de nutriment. De plus, l'expression des toxines est étroitement liée au phénomène de sporulation. En effet, ces deux mécanismes sont régulés par les mêmes gènes : *spoOA* et *codY*. Cette diminution d'expression concorde avec celle observée pour les gènes de sporulation. En somme, nous avons mis en évidence une diminution de la sporulation et des toxines. Cette diminution pourrait notamment faire suite à l'approvisionnement de nutriments par les bactéries environnantes.

Le lien entre la sporulation et les besoins nutritifs de notre bactérie peuvent être abordés sous l'angle de la métabolisation de l'éthanolamine (Pitts et al. 2012; Nawrocki et al. 2018). *C. difficile*, à travers l'opéron *eut* est capable d'utiliser l'éthanolamine comme source d'énergie. Par ailleurs, ces résultats coïncident avec l'hypothèse que la présence d'éthanolamine enclenche l'expression de l'opéron « *eut* » (Nawrocki et al. 2018). En effet, l'éthanolamine étant un composant de la membrane cellulaire, *C. difficile* ne retrouve pas ce composant dans un environnement dépourvu de bactéries, ce qui explique une expression plus faible de l'opéron *eut* dans la condition contrôle. De plus, l'expression du gène *eutS* induite par la capacité de métaboliser l'éthanolamine pourrait également avoir un effet sur la virulence. Chez plusieurs autres espèces bactériennes, l'utilisation d'éthanolamine permet la régulation de facteurs de virulence. Chez *C. difficile*, l'effet de l'éthanolamine sur la virulence est moins évident puisque *in vitro*, le changement d'expression des toxines n'est pas significatif. Cependant, *in vivo* sur le modèle hamster, la capacité d'utiliser l'éthanolamine par *C. difficile* a montré un délai dans le développement de la maladie (Nawrocki et al. 2018). Cela pourrait être dû à un retard de l'expression des toxines par une régulation indirecte. Cette hypothèse pourrait expliquer pourquoi

l'expression des gènes des toxines TcdA et TcdB est diminuée en présence de matière fécale dans nos expériences de "transwell".

Un deuxième processus biologique permettant la survie de *C. difficile* dans le tube digestif est sa capacité à former ou à participer à un biofilm. Dans l'intestin, de nombreuses bactéries sont capables de former un biofilm, une matrice constituée de polysaccharides (Tremblay et Dupuy 2022). Cette matrice offre plusieurs avantages de survie notamment en formant un réservoir nutritionnel et en offrant une résilience aux agents anti-bactériens. En présence de matières fécales, nos résultats montrent une augmentation de certains gènes liés au biofilm et à la synthèse de polysaccharides comme on peut le retrouver dans nos résultats malgré une absence des grands gènes régulateurs.

Les polyamines jouent entre autres un rôle dans la formation du biofilm chez d'autres bactéries *Bacillus*. Effectivement, l'absence de spermidine engendre une diminution de la formation du biofilm chez *Bacillus subtilis*. En outre cet effet sur le biofilm, les spermidines et les putrescines jouent également un rôle dans la protection contre le stress oxydatif (Shah et Swiatlo 2008). La plupart des bactéries trouvent leur spermidine dans leur environnement, cela pourrait expliquer pourquoi l'expression de leurs gènes augmente en présence de matières fécales.

Clostridioides difficile est une bactérie anaérobie dont le métabolisme n'utilise pas d'oxygène. Au lieu de cela, elle utilise des processus tels que la respiration anaérobie et la fermentation. Son métabolisme s'adapte en fonction de l'environnement qui l'entoure, notamment en fonction de la présence de nutriments.

Par exemple, Stickland est un chemin métabolique qui utilise la dégradation des acides aminés par une réaction d'oxydo-réduction pour produire de l'énergie et un gradient de proton (Stickland 1934). En présence de matière fécale, nous avons pu mettre en évidence un découplage entre le chemin oxydatif et réducteur du métabolisme de Stickland. De plus, *Clostridioides difficile* utilise certains acides aminés de manière préférentielle.

Effectivement, en présence de proline, l'opéron *prd* va inhiber l'expression de l'opéron *grd* codant pour la glycine réductase (Bouillaut, Self, et Sonenshein 2013). Nos résultats reflètent cette inhibition puisque nous observons une diminution significative d'expression du gène *grdB* en comparaison à l'augmentation de *prdA* en présence de matière fécale.

Nous pouvons conclure qu'en présence d'un voisinage bactérien, *C. difficile* privilégie le chemin réducteur de Stickland en utilisant notamment la proline. Lorsque le couplage rédox est à l'équilibre, ce chemin réducteur joue un rôle de balance pour les réactions d'oxydation en formant des NAD⁺ (Neumann-Schaal, Jahn, et Schmidt-Hohagen 2019). Cependant, d'autres espèces comme

Clostridium sporogenes montrent que le chemin réducteur peut participer à la production d'énergie, notamment en utilisant la proline. La consommation de proline par réduction donne lieu à la formation d'une force motrice de proton permettant la génération d'ATP par ATP synthase (Liu et al. 2022; Lovitt, Kell, et Morris 1986). Chez *C. difficile*, il a également été démontré que le chemin réducteur participe à la formation d'un gradient de proton par couplage au complexe RNF.

En parallèle, *C. difficile* utilise un second chemin de fermentation qui est la voie Wood-Ljungdahl. Il s'agit d'un chemin métabolique qui se sert de 2 CO₂ comme accepteur d'électron final et pour former du NADH. Wood-Ljungdahl est utile lorsqu'il n'y a plus d'acide aminé dans l'environnement, il permet à *C. difficile* de se multiplier de manière autotrophe avec du CO₂ et des H₂ (Köpke, Straub, et Dürre 2013; Neumann-Schaal, Jahn, et Schmidt-Hohagen 2019). Les gènes qui régulent la voie Wood-Ljungdahl sont extrêmement conservés chez *Clostridioides difficile*. Ils sont eux-mêmes régulés en fonction des carbohydrates présents dans l'intestin. Pour continuer, la production de toxine également régulée par l'environnement nutritif serait responsable de l'expression des gènes du WLP. Dans nos expériences de "transwell", nous avons montré une diminution du taux de toxines en présence de bactéries ainsi que l'utilisation du métabolisme réducteur de Stickland pour produire de l'énergie. Nous pouvons conclure que dans ces conditions de croissances, *C. difficile* parvient à trouver les nutriments dont il a besoin. Lorsque nous cherchons l'expression des gènes composant le chemin métabolique Wood-Ljungdahl, il y a une diminution de l'expression en présence de matières fécales correspondant avec notre hypothèse.

En complément, *Clostridioides difficile* utilise les carbohydrates à partir de la glycolyse et glycogénèse. Le métabolisme central est régulé en fonction de la présence d'acides aminés dans l'environnement et de l'apport en carbohydrates (Neumann-Schaal, Jahn, et Schmidt-Hohagen 2019; Marshall et al. 2023). Les carbohydrates se trouvent en grande quantité dans les intestins et sont consommés par de nombreuses bactéries. C'est pourquoi *C. difficile* compte sur la mise en place d'un environnement dysbiotique où il pourra accéder à une quantité plus importante de carbohydrates

Premièrement, les transporteurs PTS du fructose sont significativement moins exprimés lorsque *C. difficile* se trouve dans un voisinage bactérien. En outre, les carbohydrates sont consommés par les bactéries compétitrices. Cependant, lors d'une dysbiose, puisque les bactéries consommatrices de carbohydrates diminuent, *C. difficile* peut utiliser cette source de nutriments en augmentant l'expression des gènes liés au métabolisme central et de transports comme les PTS. Dans nos expériences, les différences entre les deux matières fécales pourraient s'expliquer par le fait que F03 sont les selles d'un individu plus âgé. Or, avec l'âge le microbiote change et les bactéries

consommatrices de carbohydrates de F03 pourraient être moins présentes que dans les selles F04 provenant d'individus plus jeunes.

Pour poursuivre, *C. difficile* est capable d'utiliser le tréhalose comme source de carbohydrate. Ce carbohydrate pourrait affecter la virulence de *C. difficile*. En effet, des études ont montré qu'en présence de tréhalose, les toxines TcdA et TcdB seraient augmentées par rapport à la condition sans tréhalose. Dans nos expériences, la diminution des modules liés aux tréhaloses est diminuée corrélant avec la diminution des toxines (Marshall et al. 2023).

Enfin, lors d'une limitation en fer, l'expression de certains gènes change comme les ferrédoxines du métabolisme central. Pour compenser leur diminution, *C. difficile* va augmenter l'expression des gènes RNF couplés aux enzymes liés à l'utilisation de la proline comme la proline racemase. Ces changements d'expression sont observés de façon significative entre la condition WOF et les conditions avec matières fécales. Pour continuer, les transporteurs de fer sont également diminués en cas d'absence de leur substrat. Dans un voisinage bactérien, *C. difficile* n'est plus le seul à utiliser le fer. Nos résultats mettent en évidence cette diminution de l'accès au fer (Berges et al. 2018).

Suite à l'analyse transcriptomique, nous avons mesuré l'expression de certains gènes cibles en utilisant une méthodologie différente mais également basée sur l'ADNc, la PCR quantitative (qPCR). Nous avons sélectionné une série de gènes et créé un design PCR spécifique à chacun. Pour terminer, nous avons comparé les tendances de « \log_2 fold change » entre deux conditions et ceux obtenus en qPCR. Pour les 6 gènes sélectionnés, 5 d'entre eux montrent une tendance similaire pour la comparaison entre F03 et WOF et 4 d'entre eux montrent une tendance similaire pour la comparaison entre F04 et WOF. L'article « Benchmarking of RNA-sequencing analysis workflows using wholetranscriptome RT-qPCR expression data » de Celine Everaert a mis en évidence que 15 à 19% des gènes ne peuvent pas être comparés en RNAseq et qPCR. Les gènes sélectionnés font peut-être partie de cette catégorie (Everaert et al. 2017).

Il convient de noter que notre étude présente certaines limitations pour une meilleure compréhension des changements d'expression des gènes de *C. difficile*. Pour commencer, les analyses transcriptomiques à 12 h n'offrent qu'une image partielle des changements d'expression au cours du temps. Actuellement, des analyses complémentaires à 24h vont permettre de mettre en évidence les différences d'expression des gènes de *C. difficile* au cours du temps et de l'environnement. Pour continuer, nos expériences se sont basées sur seulement deux types de matière fécale. Il serait nécessaire d'inclure plusieurs microbiotes de composition plus diverse pour émettre des hypothèses plus solides sur les changements de comportement de *C. difficile* en fonction du voisinage bactérien.

à l'aide du sihumix, un modèle de microbiote intestinal humain. Ce modèle permettra l'analyse transcriptomique du microbiote en présence de *C. difficile* mais également de moduler sa composition afin de mettre en évidence son impact sur l'expression génique.

VI. Conclusion

En conclusion, nos expériences de « transwells » ont révélé des changements d'expressions significatifs des gènes de *C. difficile* en fonction de la présence ou en absence d'un environnement bactérien. L'insertion de microbiote provoque des changements d'expression des gènes sur l'ensemble du génome de *C. difficile*. Nous pouvons également noter qu'en fonction du microbiote ajouté, certains gènes sont également exprimés de façon différente. Ces changements d'expression se retrouvent sur l'ensemble du génome.

En somme, nos découvertes soulignent l'importance de l'environnement bactérien dans la régulation de l'expression des gènes de *C. difficile*. Les interactions entre *C. difficile* et d'autres bactéries semblent influencer son comportement. Il est probable que ces changements d'expression jouent un rôle dans la compétition bactérienne et l'adaptation de *C. difficile* à son environnement de par la modulation de certaines voies métaboliques et de virulence.

VII. Bibliographie :

- Abukhalil, Abdallah Damin, Liana AbuKhdeir, Malak Hamed, Ni'meh Al Shami, Hani A. Naseef, Banan M. Aiesh, et Ali Sabateen. 2021. « Characteristics, Risk Factors, and Prevalence of *Clostridioides Difficile* Among Hospitalized Patients in a Tertiary Care Hospital in Palestine ». *Infection and Drug Resistance* 14 (novembre): 4681-88. <https://doi.org/10.2147/IDR.S333985>.
- Awad, Milena M, Priscilla A Johanesen, Glen P Carter, Edward Rose, et Dena Lyras. 2014. « Clostridium difficile virulence factors: Insights into an anaerobic spore-forming pathogen ». *Gut Microbes* 5 (5): 579-93. <https://doi.org/10.4161/19490976.2014.969632>.
- Baloh, Marko, et Joseph A. Sorg. 2022. « Clostridioides Difficile Spore Germination: Initiation to DPA Release ». *Current Opinion in Microbiology* 65 (février): 101-7. <https://doi.org/10.1016/j.mib.2021.11.001>.
- Berges, Mareike, Annika-Marisa Michel, Christian Lassek, Aaron M. Nuss, Michael Beckstette, Petra Dersch, Katharina Riedel, et al. 2018. « Iron Regulation in Clostridioides difficile ». *Frontiers in Microbiology* 9 (décembre): 3183. <https://doi.org/10.3389/fmicb.2018.03183>.
- Borriello, S P. 1998. « Pathogenesis of Clostridium difficile infection. » *Journal of Antimicrobial Chemotherapy* 41 (suppl_3): 13-19. https://doi.org/10.1093/jac/41.suppl_3.13.
- Bouillaut, Laurent, William T. Self, et Abraham L. Sonenshein. 2013. « Proline-Dependent Regulation of Clostridium Difficile Stickland Metabolism ». *Journal of Bacteriology* 195 (4): 844-54. <https://doi.org/10.1128/JB.01492-12>.
- Brown, Kevin A., Nagham Khanafer, Nick Daneman, et David N. Fisman. 2013. « Meta-Analysis of Antibiotics and the Risk of Community-Associated Clostridium Difficile Infection ». *Antimicrobial Agents and Chemotherapy* 57 (5): 2326-32. <https://doi.org/10.1128/AAC.02176-12>.
- Buddle, Jessica E., et Robert P. Fagan. s. d. « Pathogenicity and virulence of Clostridioides difficile ». *Virulence* 14 (1): 2150452. <https://doi.org/10.1080/21505594.2022.2150452>.
- Calabi, Emanuela, Franco Calabi, Alan D. Phillips, et Neil F. Fairweather. 2002. « Binding of Clostridium difficile Surface Layer Proteins to Gastrointestinal Tissues ». *Infection and Immunity* 70 (10): 5770-78. <https://doi.org/10.1128/IAI.70.10.5770-5778.2002>.
- Chan, Hung, Shan Zhao, Lin Zhang, Jeffery Ho, Czarina C. H. Leung, Wai T. Wong, Yuanyuan Tian, et al. 2018. « Clostridium Difficile Toxin B Induces Autophagic Cell Death in Colonocytes ». *Journal of Cellular and Molecular Medicine* 22 (4): 2469-77. <https://doi.org/10.1111/jcmm.13555>.
- Chitnis, Amit S., Stacy M. Holzbauer, Ruth M. Belflower, Lisa G. Winston, Wendy M. Bamberg, Carol Lyons, Monica M. Farley, et al. 2013. « Epidemiology of Community-Associated Clostridium

- Difficile Infection, 2009 through 2011 ». *JAMA Internal Medicine* 173 (14): 1359-67. <https://doi.org/10.1001/jamainternmed.2013.7056>.
- « Clostridioides (Clostridium) Difficile Infections - Annual Epidemiological Report for 2016–2017 ». 2022. 16 novembre 2022. <https://www.ecdc.europa.eu/en/publications-data/clostridioides-difficile-infections-annual-epidemiological-report-2016-2017>.
- « Clostridioides difficile (ID 535) - Genome - NCBI ». s. d. Consulté le 23 juin 2023. https://www.ncbi.nlm.nih.gov/genome/535?genome_assembly_id=2054681.
- Cowardin, Carrie A., et William A Petri. 2014. « Host Recognition of Clostridium difficile and the Innate Immune Response ». *Anaerobe* 30 (décembre): 205-9. <https://doi.org/10.1016/j.anaerobe.2014.08.014>.
- Czepiel, Jacek, Mirosław Drózdź, Hanna Pituch, Ed J. Kuijper, William Perucki, Aleksandra Mielimonka, Sarah Goldman, Dorota Wultańska, Aleksander Garlicki, et Grażyna Biesiada. 2019. « Clostridium Difficile Infection: Review ». *European Journal of Clinical Microbiology & Infectious Diseases* 38 (7): 1211-21. <https://doi.org/10.1007/s10096-019-03539-6>.
- Đapa, Tanja, Rosanna Leuzzi, Yen K. Ng, Soza T. Baban, Roberto Adamo, Sarah A. Kuehne, Maria Scarselli, Nigel P. Minton, Davide Serruto, et Meera Unnikrishnan. 2013. « Multiple Factors Modulate Biofilm Formation by the Anaerobic Pathogen Clostridium Difficile ». *Journal of Bacteriology* 195 (3): 545-55. <https://doi.org/10.1128/JB.01980-12>.
- Dawson, Lisa F., Esmeralda Valiente, Alexandra Faulds-Pain, Elizabeth H. Donahue, et Brendan W. Wren. 2012. « Characterisation of Clostridium Difficile Biofilm Formation, a Role for Spo0A ». *PLOS ONE* 7 (12): e50527. <https://doi.org/10.1371/journal.pone.0050527>.
- Denève, Cécile, Claudine Deloménie, Marie-Claude Barc, Anne Collignon, et Claire Janoir. 2008. « Antibiotics involved in Clostridium difficile-associated disease increase colonization factor gene expression ». *Journal of Medical Microbiology* 57 (6): 732-38. <https://doi.org/10.1099/jmm.0.47676-0>.
- Dial, Sandra, J. A. C. Delaney, Alan N. Barkun, et Samy Suissa. 2005. « Use of Gastric Acid–Suppressive Agents and the Risk of Community-Acquired Clostridium difficile–Associated Disease ». *JAMA* 294 (23): 2989-95. <https://doi.org/10.1001/jama.294.23.2989>.
- DiCandia, Michael A., Adrienne N. Edwards, Joshua B. Jones, Grace L. Swaim, Brooke D. Mills, et Shonna M. McBride. 2022. « Identification of Functional Spo0A Residues Critical for Sporulation in Clostridioides Difficile ». *Journal of Molecular Biology* 434 (13): 167641. <https://doi.org/10.1016/j.jmb.2022.167641>.

- Donlan, Rodney M. 2002. « Biofilms: Microbial Life on Surfaces ». *Emerging Infectious Diseases* 8 (9): 881-90. <https://doi.org/10.3201/eid0809.020063>.
- Donnelly, M. Lauren, Kelly A. Fimlaid, et Aimee Shen. 2016. « Characterization of Clostridium Difficile Spores Lacking Either SpoVAC or Dipicolinic Acid Synthetase ». *Journal of Bacteriology* 198 (11): 1694-1707. <https://doi.org/10.1128/JB.00986-15>.
- D'Silva, Kristin M., Raaj Mehta, Michael Mitchell, Todd C. Lee, Vibha Singhal, Marnie Goodwin Wilson, et Emily G. McDonald. 2021. « Proton Pump Inhibitor Use and Risk for Recurrent Clostridioides Difficile Infection: A Systematic Review and Meta-Analysis ». *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, janvier, S1198-743X(21)00035-5. <https://doi.org/10.1016/j.cmi.2021.01.008>.
- Engevik, Melinda A., Mary Beth Yacyshyn, Kristen A. Engevik, Jiang Wang, Benjamin Darien, Daniel J. Hassett, Bruce R. Yacyshyn, et Roger T. Worrell. 2015. « Human Clostridium difficile infection: altered mucus production and composition ». *American Journal of Physiology - Gastrointestinal and Liver Physiology* 308 (6): G510-24. <https://doi.org/10.1152/ajpgi.00091.2014>.
- Farooq, Priya D., Nathalie H. Urrunaga, Derek M. Tang, et Erik C. von Rosenvinge. 2015. « Pseudomembranous Colitis ». *Disease-a-Month: DM* 61 (5): 181-206. <https://doi.org/10.1016/j.disamonth.2015.01.006>.
- Gencic, Simonida, et David A. Grahame. 2020. « Diverse Energy-Conserving Pathways in Clostridium difficile: Growth in the Absence of Amino Acid Stickland Acceptors and the Role of the Wood-Ljungdahl Pathway ». *Journal of Bacteriology* 202 (20): e00233-20. <https://doi.org/10.1128/JB.00233-20>.
- Gupta, Akshita, et Ashwin N. Ananthakrishnan. 2021. « Economic Burden and Cost-Effectiveness of Therapies for Clostridioides Difficile Infection: A Narrative Review ». *Therapeutic Advances in Gastroenterology* 14: 17562848211018654. <https://doi.org/10.1177/17562848211018654>.
- HALL, IVAN C., et ELIZABETH O'TOOLE. 1935. « INTESTINAL FLORA IN NEW-BORN INFANTS: WITH A DESCRIPTION OF A NEW PATHOGENIC ANAEROBE, BACILLUS DIFFICILIS ». *American Journal of Diseases of Children* 49 (2): 390-402. <https://doi.org/10.1001/archpedi.1935.01970020105010>.
- Heather, James M., et Benjamin Chain. 2016. « The sequence of sequencers: The history of sequencing DNA ». *Genomics* 107 (1): 1-8. <https://doi.org/10.1016/j.ygeno.2015.11.003>.
- Hecht, G., C. Pothoulakis, J. T. LaMont, et J. L. Madara. 1988. « Clostridium Difficile Toxin A Perturbs Cytoskeletal Structure and Tight Junction Permeability of Cultured Human Intestinal Epithelial

- Monolayers ». *The Journal of Clinical Investigation* 82 (5): 1516-24. <https://doi.org/10.1172/JCI113760>.
- Hennequin, Claire, Claire Janoir, Marie-Claude Barc, Anne Collignon, et Tuomo Karjalainen. 2003. « Identification and characterization of a fibronectin-binding protein from *Clostridium difficile* ». *Microbiology* 149 (10): 2779-87. <https://doi.org/10.1099/mic.0.26145-0>.
- Hofmann, Julia D., Andreas Otto, Mareike Berges, Rebekka Biedendieck, Annika-Marisa Michel, Dörte Becher, Dieter Jahn, et Meina Neumann-Schaal. 2018. « Metabolic Reprogramming of *Clostridioides difficile* During the Stationary Phase With the Induction of Toxin Production ». *Frontiers in Microbiology* 9 (août): 1970. <https://doi.org/10.3389/fmicb.2018.01970>.
- « Iron Regulation in *Clostridioides difficile* - PMC ». s. d. Consulté le 20 juin 2023. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6311696/>.
- Kanehisa, Minoru, Yoko Sato, et Kanae Morishima. 2016. « BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences ». *Journal of Molecular Biology* 428 (4): 726-31. <https://doi.org/10.1016/j.jmb.2015.11.006>.
- Kirby, Jonathan M., Helen Ahern, April K. Roberts, Vivek Kumar, Zoe Freeman, K. Ravi Acharya, et Clifford C. Shone. 2009. « Cwp84, a Surface-associated Cysteine Protease, Plays a Role in the Maturation of the Surface Layer of *Clostridium difficile* ». *The Journal of Biological Chemistry* 284 (50): 34666-73. <https://doi.org/10.1074/jbc.M109.051177>.
- Kirk, Joseph A., Dana Gebhart, Anthony M. Buckley, Stephen Lok, Dean Scholl, Gillian R. Douce, Gregory R. Govoni, et Robert P. Fagan. 2017. « New Class of Precision Antimicrobials Redefines Role of *Clostridium Difficile* S-Layer in Virulence and Viability ». *Science Translational Medicine* 9 (406): eaah6813. <https://doi.org/10.1126/scitranslmed.aah6813>.
- Kochan, Travis J., Matthew H. Foley, Michelle S. Shoshiev, Madeline J. Somers, Paul E. Carlson, et Philip C. Hanna. 2018. « Updates to *Clostridium Difficile* Spore Germination ». *Journal of Bacteriology* 200 (16): e00218-18. <https://doi.org/10.1128/JB.00218-18>.
- Köpke, Michael, Melanie Straub, et Peter Dürre. 2013. « *Clostridium Difficile* Is an Autotrophic Bacterial Pathogen ». *PloS One* 8 (4): e62157. <https://doi.org/10.1371/journal.pone.0062157>.
- Kopylova, Evguenia, Laurent Noé, et Hélène Touzet. 2012. « SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data ». *Bioinformatics* 28 (24): 3211-17. <https://doi.org/10.1093/bioinformatics/bts611>.
- Kordus, Shannon L., Audrey K. Thomas, et D. Borden Lacy. 2022. « *Clostridioides Difficile* Toxins: Mechanisms of Action and Antitoxin Therapeutics ». *Nature Reviews. Microbiology* 20 (5): 285-98. <https://doi.org/10.1038/s41579-021-00660-2>.

- Lawson, Paul A., Diane M. Citron, Kerin L. Tyrrell, et Sydney M. Finegold. 2016. « Reclassification of *Clostridium Difficile* as *Clostridioides Difficile* (Hall and O'Toole 1935) Prévot 1938 ». *Anaerobe* 40 (août): 95-99. <https://doi.org/10.1016/j.anaerobe.2016.06.008>.
- Lewis, Brittany B., Charlie G. Buffie, Rebecca A. Carter, Ingrid Leiner, Nora C. Toussaint, Liza C. Miller, Asia Gobourne, Lilan Ling, et Eric G. Pamer. 2015. « Loss of Microbiota-Mediated Colonization Resistance to *Clostridium Difficile* Infection With Oral Vancomycin Compared With Metronidazole ». *The Journal of Infectious Diseases* 212 (10): 1656-65. <https://doi.org/10.1093/infdis/jiv256>.
- Li, Wenjun, Kathleen R. O'Neill, Daniel H. Haft, Michael DiCuccio, Vyacheslav Chetvernin, Azat Badretdin, George Coulouris, et al. 2021. « RefSeq: Expanding the Prokaryotic Genome Annotation Pipeline Reach with Protein Family Model Curation ». *Nucleic Acids Research* 49 (D1): D1020-28. <https://doi.org/10.1093/nar/gkaa1105>.
- Linevsky, Joanne K., Charalabos Pothoulakis, Sarah Keates, Michel Warny, Andrew C. Keates, J. Thomas Lamont, et Ciarán P. Kelly. 1997. « IL-8 release and neutrophil activation by *Clostridium difficile* toxin-exposed human monocytes ». *American Journal of Physiology-Gastrointestinal and Liver Physiology* 273 (6): G1333-40. <https://doi.org/10.1152/ajpgi.1997.273.6.G1333>.
- Liu, Yuanyuan, Haoqing Chen, William Van Treuren, Bi-Huei Hou, Steven K. Higginbottom, et Dylan Dodd. 2022. « *Clostridium Sporogenes* Uses Reductive Stickland Metabolism in the Gut to Generate ATP and Produce Circulating Metabolites ». *Nature Microbiology* 7 (5): 695-706. <https://doi.org/10.1038/s41564-022-01109-9>.
- Love, Michael I., Wolfgang Huber, et Simon Anders. 2014. « Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2 ». bioRxiv. <https://doi.org/10.1101/002832>.
- Lovitt, Robert W., Douglas B. Kell, et J.Gareth Morris. 1986. « Proline reduction by *Clostridium sporogenes* is coupled to vectorial proton ejection ». *FEMS Microbiology Letters* 36 (2-3): 269-73. <https://doi.org/10.1111/j.1574-6968.1986.tb01708.x>.
- Marshall, Andrew, John W. McGrath, Robert Graham, et Geoff McMullan. 2023. « Food for Thought-The Link between *Clostridioides Difficile* Metabolism and Pathogenesis ». *PLoS Pathogens* 19 (1): e1011034. <https://doi.org/10.1371/journal.ppat.1011034>.
- Martinez, Elisa, Cristina Rodriguez, Sébastien Crèvecoeur, Sarah Lebrun, Véronique Delcenserie, Bernard Taminiau, et Georges Daube. 2022. « Impact of environmental conditions and gut microbiota on the in vitro germination and growth of *Clostridioides difficile* ». *FEMS Microbiology Letters* 369 (1): fnac087. <https://doi.org/10.1093/femsle/fnac087>.

- Martinez, Elisa, Bernard Taminiau, Cristina Rodriguez, et Georges Daube. 2022. « Gut Microbiota Composition Associated with Clostridioides Difficile Colonization and Infection ». *Pathogens (Basel, Switzerland)* 11 (7): 781. <https://doi.org/10.3390/pathogens11070781>.
- Matamouros, Susana, Patrick England, et Bruno Dupuy. 2007. « Clostridium Difficile Toxin Expression Is Inhibited by the Novel Regulator TcdC ». *Molecular Microbiology* 64 (5): 1274-88. <https://doi.org/10.1111/j.1365-2958.2007.05739.x>.
- Meador-Parton, Jennifer, et David L. Popham. 2000. « Structural Analysis of Bacillus subtilis Spore Peptidoglycan during Sporulation ». *Journal of Bacteriology* 182 (16): 4491-99. <https://doi.org/10.1128/jb.182.16.4491-4499.2000>.
- Meza-Torres, Jazmin, Emile Auria, Bruno Dupuy, et Yannick D. N. Tremblay. 2021. « Wolf in Sheep's Clothing: Clostridioides difficile Biofilm as a Reservoir for Recurrent Infections ». *Microorganisms* 9 (9): 1922. <https://doi.org/10.3390/microorganisms9091922>.
- Nawrocki, Kathryn L., Daniela Wetzel, Joshua B. Jones, Emily C. Woods, et Shonna M. McBride. 2018. « Ethanolamine Is a Valuable Nutrient Source That Impacts Clostridium Difficile Pathogenesis ». *Environmental Microbiology* 20 (4): 1419-35. <https://doi.org/10.1111/1462-2920.14048>.
- Nerber, Hailee N., et Joseph A. Sorg. 2021. « The Small Acid-Soluble Proteins of Clostridioides Difficile Are Important for UV Resistance and Serve as a Check Point for Sporulation ». *PLoS Pathogens* 17 (9): e1009516. <https://doi.org/10.1371/journal.ppat.1009516>.
- Neumann-Schaal, Meina, Dieter Jahn, et Kerstin Schmidt-Hohagen. 2019. « Metabolism the Difficile Way: The Key to the Success of the Pathogen Clostridioides difficile ». *Frontiers in Microbiology* 10. <https://www.frontiersin.org/articles/10.3389/fmicb.2019.00219>.
- Nibbering, Britt, Dale N. Gerding, Ed J. Kuijper, Romy D. Zwiittink, et Wiep Klaas Smits. 2021. « Host Immune Responses to Clostridioides difficile: Toxins and Beyond ». *Frontiers in Microbiology* 12 (décembre): 804949. <https://doi.org/10.3389/fmicb.2021.804949>.
- Olson, Robert D., Rida Assaf, Thomas Brettin, Neal Conrad, Clark Cucinell, James J. Davis, Donald M. Dempsey, et al. 2023. « Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): A Resource Combining PATRIC, IRD and ViPR ». *Nucleic Acids Research* 51 (D1): D678-89. <https://doi.org/10.1093/nar/gkac1003>.
- Penrod, Joseph T., et John R. Roth. 2006. « Conserving a Volatile Metabolite: a Role for Carboxysome-Like Organelles in Salmonella enterica ». *Journal of Bacteriology* 188 (8): 2865-74. <https://doi.org/10.1128/JB.188.8.2865-2874.2006>.
- Pitts, Alison C., Laura R. Tuck, Alexandra Faulds-Pain, Richard J. Lewis, et Jon Marles-Wright. 2012. « Structural Insight into the Clostridium Difficile Ethanolamine Utilisation

- Microcompartment ». *PLOS ONE* 7 (10): e48360. <https://doi.org/10.1371/journal.pone.0048360>.
- Quick, Joshua, Aaron R. Quinlan, et Nicholas J. Loman. 2014. « A Reference Bacterial Genome Dataset Generated on the MinION™ Portable Single-Molecule Nanopore Sequencer ». *GigaScience* 3: 22. <https://doi.org/10.1186/2047-217X-3-22>.
- Ransom, Eric M., Gabriela M. Kaus, Phuong M. Tran, Craig D. Ellermeier, et David S. Weiss. 2018. « Multiple factors contribute to bimodal toxin gene expression in *Clostridioides (Clostridium) difficile* ». *Molecular microbiology* 110 (4): 533-49. <https://doi.org/10.1111/mmi.14107>.
- Rinninella, Emanuele, Pauline Raoul, Marco Cintoni, Francesco Franceschi, Giacinto Abele Donato Miggiano, Antonio Gasbarrini, et Maria Cristina Mele. 2019. « What is the Healthy Gut Microbiota Composition? A Changing Ecosystem across Age, Environment, Diet, and Diseases ». *Microorganisms* 7 (1): 14. <https://doi.org/10.3390/microorganisms7010014>.
- Rodriguez, Cristina, Lamia Bouchafa, Kate Soumillion, Eleonore Ngyuvula, Bernard Taminiau, Johan Van Broeck, Michel Delmée, et Georges Daube. 2019. « Seasonality of *Clostridium Difficile* in the Natural Environment ». *Transboundary and Emerging Diseases* 66 (6): 2440-49. <https://doi.org/10.1111/tbed.13301>.
- Roof, D M, et J R Roth. 1989. « Functions required for vitamin B12-dependent ethanolamine utilization in *Salmonella typhimurium* ». *Journal of Bacteriology* 171 (6): 3316-23. <https://doi.org/10.1128/jb.171.6.3316-3323.1989>.
- Rutherford, Steven T., et Bonnie L. Bassler. 2012. « Bacterial Quorum Sensing: Its Role in Virulence and Possibilities for Its Control ». *Cold Spring Harbor Perspectives in Medicine* 2 (11): a012427. <https://doi.org/10.1101/cshperspect.a012427>.
- Shah, Pratik, et Edwin Swiatlo. 2008. « A Multifaceted Role for Polyamines in Bacterial Pathogens ». *Molecular Microbiology* 68 (1): 4-16. <https://doi.org/10.1111/j.1365-2958.2008.06126.x>.
- Shen, Aimee, Adrienne N. Edwards, Mahfuzur R. Sarker, et Daniel Paredes-Sabja. 2019. « Sporulation and Germination in Clostridial Pathogens ». *Microbiology Spectrum* 7 (6). <https://doi.org/10.1128/microbiolspec.GPP3-0017-2018>.
- Simpson, Morgan, Alyse Frisbee, Pankaj Kumar, Carsten Schwan, Klaus Aktories, et William A. Petri. 2022. « *Clostridioides Difficile* Binary Toxin Is Recognized by the Toll-Like Receptor 2/6 Heterodimer to Induce a Nuclear Factor- κ B Response ». *The Journal of Infectious Diseases* 225 (7): 1296-1300. <https://doi.org/10.1093/infdis/jiaa620>.
- Sorg, Joseph A., et Abraham L. Sonenshein. 2008. « Bile Salts and Glycine as Cogerminants for *Clostridium difficile* Spores ». *Journal of Bacteriology* 190 (7): 2505-12. <https://doi.org/10.1128/JB.01765-07>.

- Stickland, Leonard Hubert. 1934. « Studies in the metabolism of the strict anaerobes (genus *Clostridium*) ». *Biochemical Journal* 28 (5): 1746-59.
- Subramanian, Aravind, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, Amanda Paulovich, et al. 2005. « Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles ». *Proceedings of the National Academy of Sciences of the United States of America* 102 (43): 15545-50. <https://doi.org/10.1073/pnas.0506580102>.
- Taggart, Megan G., William J. Snelling, Patrick J. Naughton, Roberto M. La Ragione, James S. G. Dooley, et Nigel G. Ternan. 2021. « Biofilm Regulation in *Clostridioides Difficile*: Novel Systems Linked to Hypervirulence ». *PLOS Pathogens* 17 (9): e1009817. <https://doi.org/10.1371/journal.ppat.1009817>.
- Thursby, Elizabeth, et Nathalie Juge. 2017. « Introduction to the human gut microbiota ». *Biochemical Journal* 474 (11): 1823-36. <https://doi.org/10.1042/BCJ20160510>.
- Tremblay, Yannick Dn, et Bruno Dupuy. 2022. « The Blueprint for Building a Biofilm the *Clostridioides Difficile* Way ». *Current Opinion in Microbiology* 66 (avril): 39-45. <https://doi.org/10.1016/j.mib.2021.12.002>.
- « Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp - Chen - 2023 - iMeta - Wiley Online Library ». s. d. Consulté le 19 juin 2023. <https://onlinelibrary.wiley.com/doi/10.1002/imt2.107>.
- Waligora, Anne-Judith, Claire Hennequin, Peter Mullany, Pierre Bourlioux, Anne Collignon, et Tuomo Karjalainen. 2001. « Characterization of a Cell Surface Protein of *Clostridium difficile* with Adhesive Properties ». *Infection and Immunity* 69 (4): 2144-53. <https://doi.org/10.1128/IAI.69.4.2144-2153.2001>.
- Webb, Brandon J., Aruna Subramanian, Bert Lopansri, Bruce Goodman, Peter Bjorn Jones, Jeffrey Ferraro, Edward Stenehjem, et Samuel M. Brown. 2020. « Antibiotic Exposure and Risk for Hospital-Associated *Clostridioides Difficile* Infection ». *Antimicrobial Agents and Chemotherapy* 64 (4): e02169-19. <https://doi.org/10.1128/AAC.02169-19>.
- Weber, David J., Deverick J. Anderson, Daniel J. Sexton, et William A. Rutala. 2013. « Role of the Environment in the Transmission of *Clostridium Difficile* in Health Care Facilities ». *American Journal of Infection Control* 41 (5 Suppl): S105-110. <https://doi.org/10.1016/j.ajic.2012.12.009>.
- Wick, Ryan R., Louise M. Judd, Louise T. Cerdeira, Jane Hawkey, Guillaume Méric, Ben Vezina, Kelly L. Wyres, et Kathryn E. Holt. 2021. « Trycycler: consensus long-read assemblies for bacterial genomes ». *Genome Biology* 22 (1): 266. <https://doi.org/10.1186/s13059-021-02483-z>.

- Wick, Ryan R., Louise M. Judd, Claire L. Gorrie, et Kathryn E. Holt. 2017. « Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads ». *PLoS Computational Biology* 13 (6): e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>.
- Wright, Anne, Denise Drudy, Lorraine Kyne, Katherine Brown, et Neil F. Fairweather. 2008. « Immunoreactive cell wall proteins of *Clostridium difficile* identified by human sera ». *Journal of Medical Microbiology* 57 (6): 750-56. <https://doi.org/10.1099/jmm.0.47532-0>.
- Zhu, Duolong, Joseph A. Sorg, et Xingmin Sun. 2018. « *Clostridioides Difficile* Biology: Sporulation, Germination, and Corresponding Therapies for *C. Difficile* Infection ». *Frontiers in Cellular and Infection Microbiology* 8: 29. <https://doi.org/10.3389/fcimb.2018.00029>.