



HAL
open science

Impact of antibiotics on the human gut phageome: from method to application

Camille d'Humières

► **To cite this version:**

Camille d'Humières. Impact of antibiotics on the human gut phageome: from method to application. Life Sciences [q-bio]. Université Paris Cité, 2020. English. NNT: 2020UNIP7014 . tel-04076281

HAL Id: tel-04076281

<https://pasteur.hal.science/tel-04076281>

Submitted on 20 Apr 2023

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

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



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Thèse de doctorat de

l'Université de Paris

Ecole doctorale FIRE (ED 474)

UMR1137, INSERM, Universités de Paris et Paris Nord

Génomique évolutive des microbes, Institut Pasteur

Impact des antibiotiques sur le phageome intestinal humain: de la méthode à l'application.

Par Camille d'Humières

Thèse de doctorat de microbiologie

Directeur : Erick Denamur

Co-Directeur : Eduardo Rocha

Présentée et soutenue publiquement à Paris le 24 janvier 2020

Président du jury : Pr Stéphane Bonarcorsi, IAME, Inserm, Université de Paris

Rapporteuses : Pr Marie-Cécile Ploy, Inserm, Université de Limoges

Dr Mireille Ansaldi, CNRS, Aix-Marseille Université

Examineurs : Dr David Vallenet, CEA, Université Paris-Saclay

Directeur de thèse : Pr Erick Denamur, IAME, Inserm, Université de Paris

Co-directeurs de thèse : Dr Eduardo Rocha, CNRS, Institut Pasteur

Résumé

Titre

Impact des antibiotiques sur le phageome intestinal humain: de la méthode à l'application.

Résumé

L'antibiorésistance est un problème majeur de santé publique. Mieux comprendre l'impact des antibiotiques sur le microbiote intestinal et l'acquisition de résistances est essentiel. Dans le cadre d'une étude clinique randomisée contrôlée (CEREMI), deux antibiotiques (ceftriaxone et céfotaxime) aux spectres d'activité antimicrobienne similaires, mais aux modes d'élimination différents, ont été administrés à 22 volontaires sains. Les selles de ces derniers ont été collectées à différents temps sur une période de 6 mois afin d'évaluer l'impact des antibiotiques sur la composition et l'évolution des communautés bactériennes et phagiques (virus bactériens) du microbiote intestinal ainsi que la présence de gènes de résistance aux antibiotiques.

L'étude du phageome intestinal a nécessité le développement de méthodologies adaptées tant au niveau expérimental que bioinformatique. Suite à la comparaison de différentes méthodes d'isolement de phages issus de fèces, la technique de concentration au polyéthylène glycol s'est révélée être la plus adaptée à notre étude au regard du grand nombre d'échantillons (>100) à analyser par la suite, de par sa simplicité, sa reproductibilité, son faible coût et de la grande diversité des phages obtenus. Un pipeline bioinformatique d'identification et d'annotation des phages, combinant des approches de métagénomique et de génomique comparative a été spécifiquement développé. Il permet, à partir des reads issus du séquençage métagénomique des phages, d'assembler, d'annoter syntaxiquement et fonctionnellement les différents contigs et enfin de prédire automatiquement s'ils sont des phages. Ces méthodologies ont ainsi été appliquées à l'ensemble des échantillons de selles issues de l'étude CEREMI.

Les analyses de la communauté phagique ont révélé que l'usage de ces deux antibiotiques entraîne une forte perturbation du phageome intestinal avec un retour progressif à l'équilibre au 30ème jour. L'usage de ceftriaxone semble cependant avoir un effet plus important sur la baisse de diversité phagique. Malgré cette perturbation, le phageome intestinal reste hautement individu-spécifique et ne semble pas impliqué dans le transfert de gènes de résistance aux antibiotiques.

En parallèle, l'étude de la communauté bactérienne, par des comptages bactériens et l'analyse des 16S, a montré que les deux antibiotiques entraînent une diminution rapide et drastique de la diversité bactérienne, sans pour autant sélectionner de bacilles à Gram négatifs résistants aux céphalosporines de troisième génération. Contrairement au phageome, les deux traitements semblent avoir un impact similaire sur le bactériome.

Mots clefs

Microbiote intestinal, antibiotiques, phages, phageome intestinal, gènes de résistance aux antibiotiques.

Abstract

Title

Impact of antibiotics on the human gut phageome: from method to application.

Abstract

Antimicrobial resistance is a major public health problem. A better understanding of the impact of antibiotics on the intestinal microbiota and the development of resistance is essential. In a randomized controlled clinical trial (CEREMI), two antibiotics (ceftriaxone and cefotaxime) with similar antimicrobial activity spectrum, but different modes of elimination, were administered to 22 healthy volunteers. The stools of the latter were collected at different times over a period of 6 months, in order to evaluate the impact of antibiotics on the composition and evolution of bacterial and phage (bacterial viruses) communities of the intestinal microbiota as well as on the presence of resistance genes.

The study of the intestinal phageome required the development of methodologies adapted to both experimental and bioinformatics levels. Following the comparison of different methods of isolating phages from feces, the polyethylene glycol concentration technique proved to be the most suitable for our study given the large number of samples (>100) to be analysed later, its simplicity, reproducibility, low cost and the great diversity of the phages obtained. A bioinformatics pipeline for phage identification and annotation, combining metagenomic and comparative genomic approaches, has been specifically developed. It allows, from the reads resulting from the metagenomic sequencing of phages, to assemble, annotate syntactically and functionally the different contigs, and finally to automatically predict if they are phages. These methodologies were thus applied to all stool samples from the CEREMI study.

The analyses of the phage community revealed that the use of these two antibiotics causes a strong disturbance of the intestinal phageome with a gradual return to equilibrium on the 30th day. However, the use of ceftriaxone seems to have a greater effect on the decrease in phage diversity. Despite this disruption, the intestinal phageome remains highly individual-specific and does not appear to be involved in the transfer of antibiotic resistance genes.

In parallel, the study of the bacterial community, through bacterial counts and 16S analysis, showed that both antibiotics lead to a rapid and drastic decrease in bacterial diversity, without selecting Gram-negative bacilli resistant to third generation cephalosporins. Unlike the phageome, both treatments seem to have a similar impact on the bacteriome

Keywords

Intestinal microbiota, antibiotics, phages, intestinal phageome, antibiotic resistance genes.

Remerciements

Merci aux **membres de mon jury** de me faire l'honneur de juger mon travail. Mireille Ansaldi et Marie Cécile Ploy qui ont accepté d'être les rapporteuses de cette thèse, David Vallenet qui a accepté d'être examinateur et Stéphane Bonacorsi qui me fait l'honneur de présider !

Merci à mes deux directeurs de thèse Erick et Eduardo. Merci sincèrement pour votre humanité, votre gentillesse et votre encadrement. Deux grossesses pendant une thèse ce n'est pas rien! Je me suis toujours sentie accompagnée et épaulée de votre part. J'ai pu mener à bien cette thèse et ma vie de famille en parallèle, je vous remercie sincèrement de m'avoir soutenue.

Erick, merci de votre humanité, bienveillance et disponibilité. Vous m'accompagnez depuis 5 ans maintenant avec toujours beaucoup d'enthousiasme. Merci pour votre confiance et vos bons conseils. Merci pour le modèle que vous représentez à mes yeux dans la gestion de votre (grande) équipe.

Eduardo, merci d'avoir accepté de m'accueillir dans ton équipe et dans ce bel établissement qu'est Pasteur. Je pense que lorsque ma première question a été « comment on allume l'ordinateur » tu as pris conscience de mon ignorance totale! Merci pour ton enseignement, ta disponibilité, ta rigueur et ton humanité. Merci de m'avoir fait découvrir l'univers de la bioinformatique, j'ai trouvé ça assez hostile au départ et j'étais complètement perdue mais je dois avouer que j'y ai pris gout !!

Merci à **Antoine Andremont** de m'avoir accueilli dans son service en 2013, de m'avoir toujours soutenue et de m'avoir présenté à Erick pour la Thèse de Sciences. Une pensée particulière pour **Mr Bingen**.

Merci à l'équipe « **gem** » pour son accueil et la précieuse aide apportée par beaucoup d'entre vous : Jorges, Amandine (X2), Olaya, Marc, Matthieu, fanny, Eugen, Antoine. Un spécial merci à Rémi qui m'a consacré de son temps avec toujours le sourire !

Merci à **Marie** de m'avoir prise sous son aile au départ ! Ouverture du terminal, le béaba de « bash » et les premières lignes de codes c'est grâce à toi.

Merci aux **stylés de 7 E** ... sans qui rien n'aurait été pareil, j'ai nommé Jean, Aude et Camille. Que d'excellent souvenir dans notre « jungle ». Les 2 premières années de thèse n'ont pas été que faciles et sincèrement grâce à vous c'était un bonheur de venir au labo !! Longue vie à notre amitié! Un spécial merci à Jean sans qui python ne serait encore qu'un serpent pour moi. La pédagogie selon Jean, un poème.

Merci à **Maël**, tout a commencé entre un sbatch et un srunch et a continué pendant nos nombreux déjeuners, qui disons-le, nous ont permis d'un peu changer le monde...

Merci à la team « thésard » de **IAME**, je n'ai pas été beaucoup là mais vous voir donnait toujours le sourire : Méril, Anne-Claire, Marine, Marion, André et momo.

Merci à tous ceux qui m'ont apporté leur précieuse aide à un moment de l'aventure : Laurent Debarbieux, Nicolas Dufour, Mélanie Magnan, MÉRIL (re 😊), Djamel, Amine, Stevonn.

Merci à l'équipe de **Bichat** : Laurence, Etienne, Nathalie, Naouale, Emilie, Catherine, Emmanuelle pour les moments intenses vécus à l'hôpital. Merci de m'avoir permis de rendre ma thèse dans des bonnes conditions en me libérant du temps ! Un merci particulier à Laurence et Etienne pour avoir pris le temps de discuter avec moi.

A mes relecteurs(trices) : Antoine, Gaële, Aude, papa et Fred (un spécial merci à Gaële pour les passages en anglais) ! Disponibilité, réactivité, rapidité → impressionnant !

Merci aux **copains** qui n'ont rien à voir avec la choucroute mais avec qui ça fait toujours autant plaisir de ricaner !

Merci à [A compléter au cas ou j'ai oublié quelqu'un😊]

A mes **parents et ma famille** pour leur soutien inconditionnel !

Et évidemment le meilleur pour la fin à **Fred** et à nos deux petits coquins chats : **Augustin** et **Eléonore**. La vita è bella.

Table des matières

1	INTRODUCTION GENERALE	9
1.1	GENES DE RESISTANCE AUX ANTIBIOTIQUES	9
1.1.1	PROBLEMATIQUE ACTUELLE	9
1.1.2	APPROCHE ECOLOGIQUE	12
1.1.3	ELEMENTS GENETIQUES MOBILES	13
1.2	LE MICROBIOTE INTESTINAL	15
1.2.1	PRESENTATION ET COMPOSITION	15
1.2.2	METHODE D'ANALYSE DU MICROBIOTE INTESTINAL	18
1.2.3	EFFET DES ANTIBIOTIQUES SUR LE MICROBIOTE INTESTINAL	21
1.3	LES PHAGES	24
1.3.1	GENERALITES SUR LES PHAGES : DEFINITION, COMPOSITION, CLASSIFICATION	24
1.3.2	ETAPE D'ADSORPTION : SPECIFICITE DE L'HOTE DES PHAGES	29
1.3.3	CYCLES DE REPRODUCTION PHAGIQUE	29
1.3.4	TRANSDUCTION	32
1.3.5	TRANSDUCTION ET TRANSFERT GENETIQUE	35
1.3.6	LES PHAGES DU MICROBIOTE INTESTINAL	36
1.4	PHAGES ET GENES DE RESISTANCE AUX ANTIBIOTIQUES	40
1.4.1	PHAGES ET GENES DE RESISTANCE AUX ANTIBIOTIQUES : OBSERVATIONS	40
1.4.2	TRANSFERT DE GENES DE RESISTANCE AUX ANTIBIOTIQUES ENTRE BACTERIES ET PHAGES	41
1.4.3	PHAGES : RESERVOIR DE GENES DE RESISTANCE AUX ANTIBIOTIQUES AU SEIN DU MICROBIOTE INTESTINAL	43
1.4.4	ANALYSE BIO-INFORMATIQUE DES GENES DE RESISTANCE : BASES DE DONNEES DISPONIBLES ET METHODOLOGIES UTILISEES DANS LA LITTERATURE	46
1.4.5	CONTROVERSE SUR LE TRANSFERT DE GENES DE RESISTANCE PAR LES PHAGES	49
1.5	PHAGE ET MICROBIOTE INTESTINAL : PROCEDURE D'EXTRACTION ET D'ANALYSE	51
1.5.1	ISOLEMENT	52
1.5.2	AMPLIFICATION DE L'ADN	56
1.5.3	ANALYSE BIO-INFORMATIQUE	58
1.5.4	MESURE DE DIVERSITE ET APPLICATION EN METAGENOMIQUE	64
2	OBJECTIFS ET QUESTIONS POSEES	66
3	TRAVAUX ORIGINAUX	68
3.1	UNE METHODE SIMPLE, REPRODUCTIBLE ET A MOINDRE COUT POUR ANALYSER LE PHAGEOME INTESTINAL : DE L'ISOLEMENT DES PHAGES A L'APPROCHE BIO-INFORMATIQUE	68
3.1.1	PRESENTATION DU CONTEXTE DE L'ETUDE	68
3.1.2	ARTICLE PUBLIE	69
3.1.3	PRINCIPAUX RESULTATS DE L'ARTICLE ET PERSPECTIVES	101
3.2	ETUDE CLINIQUE D'IMPACT DE LA CEFOTAXIME ET DU CEFTRIAXONE SUR LE MICROBIOTE INTESTINAL CHEZ DES VOLONTAIRES SAINS : RESULTATS BACTERIOLOGIQUES	103
3.2.1	PRESENTATION DU CONTEXTE DE L'ETUDE CLINIQUE	103
3.2.2	ARTICLE PUBLIE	105
3.2.3	PRINCIPAUX RESULTATS DE L'ARTICLE ET PERSPECTIVES	131
3.3	ETUDE CLINIQUE D'IMPACT DE LA CEFOTAXIME ET DU CEFTRIAXONE SUR LE MICROBIOTE INTESTINAL CHEZ DES VOLONTAIRES SAINS : ANALYSE DU PHAGEOME	133
3.3.1	INTRODUCTION	133
3.3.2	ARTICLE EN PREPARATION	135
3.3.3	PRINCIPAUX RESULTATS DE L'ARTICLE ET PERSPECTIVES	169

4. CONCLUSIONS ET PERSPECTIVES	171
5. BIBLIOGRAPHIE	181

Table des illustrations

Figures

FIGURE 1: EVOLUTION DU TAUX D' <i>ESCHERICHIA COLI</i> RESISTANT AUX CEPHALOSPORINES DE TROISIEME GENERATION ENTRE 2003 ET 2017. SOURCE: ECDC, EARS-NET.....	10
FIGURE 2: CARTE DE L'EUROPE DES <i>ESCHERICHIA COLI</i> RESISTANTS AUX CEPHALOSPORINES DE TROISIEME GENERATION, SOURCE: ECDC, EARS-NET, 2017.....	10
FIGURE 3 : ASSOCIATION ENTRE LE NOMBRE DE CAS D'INFECTIONS CAUSEES PAR DES BACTERIES RESISTANTES AUX ANTIBIOTIQUES ET LE NOMBRE DE MORTS ATTRIBUABLE A LA RESISTANCE AUX ANTIBIOTIQUES DANS L'UE ET L'ESPACE ECONOMIQUE EUROPEEN EN 2015, D'APRES ³	12
FIGURE 4: ROLE DES DIFFERENTS FACTEURS DANS LEUR IMPLICATION DANS LA RESISTANCE AUX ANTIBIOTIQUES, D'APRES ⁷	13
FIGURE 5: SCHEMA RECAPITULATIF DES NOTIONS D'EMERGENCE ET DE TRANSMISSION DE GENES DE RESISTANCE AUX ANTIBIOTIQUES.....	14
FIGURE 6: SCHEMA EXPLICATIF DES DIFFERENTS MECANISMES DE TRANSFERTS HORIZONTAUX DE MATERIELS GENETIQUES ENTRE DEUX BACTERIES D'APRES ⁷	15
FIGURE 7: PRINCIPAUX GROUPES TAXONOMIQUES PRESENTS DANS LE MICROBIOME INTESTINAL HUMAIN D'APRES ²⁴	17
FIGURE 8: ILLUSTRATION DES DIFFERENTES TECHNIQUES D'ANALYSE DU MICROBIOTE INTESTINAL, DE LEUR SENSIBILITE ET DU TYPE DE BACTERIES DETECTABLES D'APRES ²⁶	18
FIGURE 9: STRUCTURE SECONDAIRE DE L'ARNr 16S D' <i>ESCHERICHIA COLI</i> D'APRES ³² . IL EST CONSTITUE DE 9 REGIONS HYPERVARIABLES (V1-V9) PROPRE AUX ESPECES BACTERIENNES SEPARÉES PAR DES REGIONS CONSERVÉES PERMETTANT LA FIXATION D'AMORCES UNIVERSELLES.....	20
FIGURE 10: PROPOSITION POUR L'EVALUATION DE L'IMPACT DES ANTIBIOTIQUES SUR LE MICROBIOTE INTESTINAL D'APRES ⁴⁶ . VRE : VANCOMYCIN-RESISTANT ENTEROCOCCI, ESBL-PE : EXTENDED-SPECTRUM BETA-LACTAMASE-PRODUCING <i>ENTEROBACTERIACEAE</i> , CPE : CARBAPENEMASE-PRODUCING <i>ENTEROBACTERIACEAE</i> , NDRB : MULTIDRUG-RESISTANT BACTERIA.....	24
FIGURE 11: MORPHOTYPES DES DIFFERENTES FAMILLES DE BACTERIOPHAGES, D'APRES ⁵⁵	26
FIGURE 12: REPRESENTATION SCHEMATIQUE TYPIQUE D'UN BACTERIOPHAGE DE L'ORDRE DES <i>CAUDOVIRALES</i>	27
FIGURE 13: PHOTOS PRISE AU MICROSCOPE ELECTRONIQUE. CREDITS: C. D'HUMIERES, L. DEBARBIEUX.....	28
FIGURE 14: SCHEMA EXPLICATIF DES DIFFERENTS TYPES DE PHAGES: VIRULENTS ET LYTIQUES. M. TOUCHON, NON PUBLIE.....	32
FIGURE 15: PHOTOGRAPHIE D'UNE LYSE BACTERIENNE PAR UN PHAGE, D'APRES ⁶⁶	32
FIGURE 16: ILLUSTRATION DE L'EXPERIENCE DE ZINDER ET LEDERBERG.....	33
FIGURE 17: SCHEMA EXPLICATIF DES DEUX TYPES DE TRANSDUCTION: GENERALISEE ET SPECIALISEE. D'APRES ⁶⁸	34
FIGURE 18: PRESENTATION DU MODELE D'AUTOTRANSDUCTION D'APRES ⁶⁹	35
FIGURE 19: ANALYSE DU PROTEOME DE PHAGES CONNUS POUR LA RECHERCHE DE GENES DE RESISTANCE AUX ANTIBIOTIQUES AVEC DIFFERENTES BASES DE DONNEES ET DIFFERENTS SEUILS BIO-INFORMATIQUES. LA LIGNE EN POINTILLEE INDIQUE LE NOMBRE DE GENES DE RESISTANCES ATTENDUS (n=2). D'APRES ¹³³	51
FIGURE 20: ILLUSTRATION DE LA NOTION DE "VIRAL DARK MATTER PROBLEM" REPRESENTANT LE POURCENTAGE DE READS OU CONTIGS NON MAPPE SUR LES BASES DE DONNEES POUR DIFFERENTS DATASETS DE PHAGEOME, D'APRES AGGARWALA ¹⁴⁶	59

Table des illustrations

Tableaux

TABLEAU 1 : PRESENTATION DES DIFFERENTES BASES DE DONNEES DE GENES DE RESISTANCE	46
TABLEAU 2: REVUE DE LA LITTERATURE DES DIFFERENTES APPROCHES BIO-INFORMATIQUES POUR RECHERCHER DES GENES DE RESISTANCE AUX ANTIBIOTIQUES DANS DES DONNEES METAGENOMIQUES DE PHAGEOME INTESTINAL.	48
TABLEAU 3: VUE D'ENSEMBLE DES DIFFERENTES METHODOLOGIES D'ISOLEMENT DES PHAGES DANS LES SELLES	55
TABLEAU 4: TABLEAU RECAPITULATIF DES ETUDES PUBLIEES DANS LA LITTERATURE UTILISANT, OU NON, UNE ETAPE D'AMPLIFICATION AVANT LE SEQUENCAGE D'ADN DE PHAGE ISSUS DU MICROBIOTE INTESTINAL.	58
TABLEAU 5: PRESENTATION DES DIFFERENTS LOGICIELS DE PREDICTION DE SEQUENCES DE PHAGES	62

1 Introduction générale

1.1 Gènes de résistance aux antibiotiques

1.1.1 Problématique actuelle

Les antibiotiques ont contribué de manière considérable aux progrès réalisés en médecine humaine durant les 60 dernières années en luttant efficacement contre les maladies bactériennes. Durant « l'âge d'or » (1960-1980), les nouveaux antibiotiques mis sur le marché ont permis aux médecins de toujours disposer de molécules efficaces. Aujourd'hui les mises sur le marché de nouveaux antibiotiques sont rares, spécialement pour les bacilles à Gram négatifs.

Toutefois, on assiste depuis une vingtaine d'années à une émergence de souches bactériennes résistantes aux antibiotiques du fait de l'apparition de nouveaux mécanismes de résistance associés à une diminution de mise sur le marché de nouveaux antibiotiques.

Le réseau européen de surveillance de la résistance aux antimicrobiens (EARS-Net), basée sur l'étude de 8 espèces bactériennes d'importance en pathologie humaine, montre que, depuis 1998, la résistance bactérienne est croissante dans tous les pays d'Europe (Figure 1). L'augmentation de la prévalence de la résistance aux céphalosporines de troisième génération (CIIG) des souches d'*Escherichia coli*, premier organisme responsable de bactériémie, est le phénomène le plus inquiétant. Elle est ainsi passée, en France, de 0,1% en 1999 à 10,2 % en 2017 (Figure 2). Cette situation est très préoccupante et pourrait conduire à des cas d'impasse thérapeutique notamment chez les sujets fragilisés en réanimation. L'organisation mondiale de la santé (OMS) a identifié l'antibiorésistance comme un problème majeur de santé publique. ¹

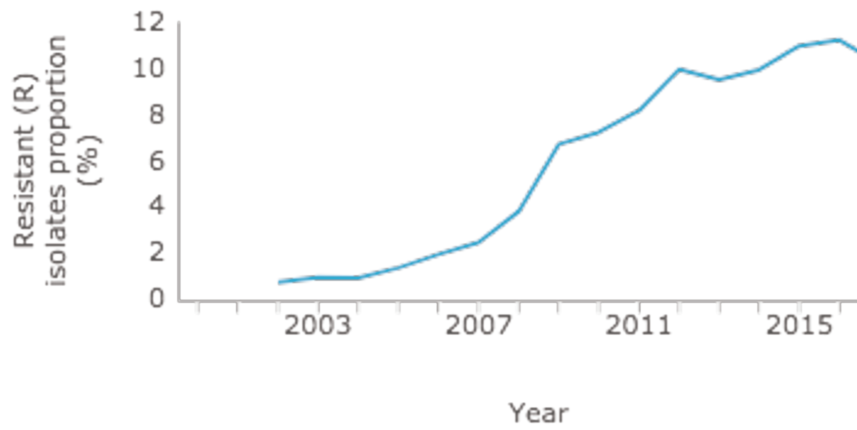


Figure 1: Evolution du taux d'*Escherichia coli* résistant aux céphalosporines de troisième génération entre 2003 et 2017. Source: ECDC, EARS-Net.

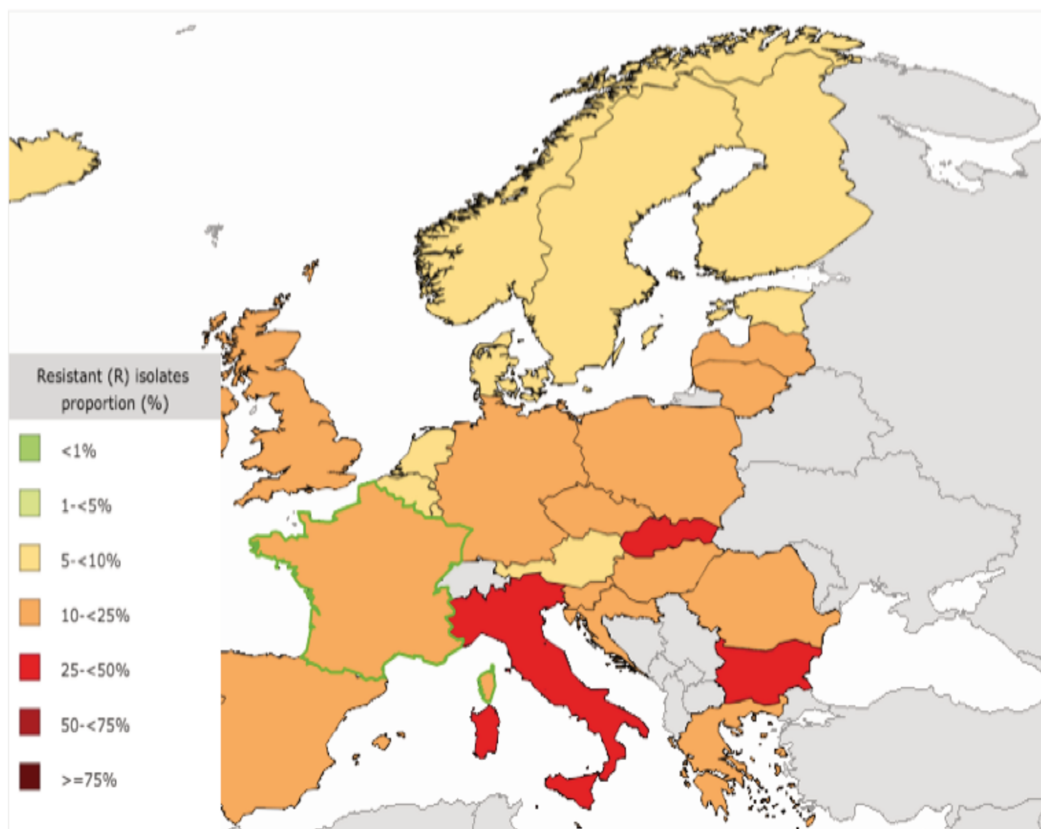


Figure 2: Carte de l'Europe des *Escherichia coli* résistants aux céphalosporines de troisième génération, source: ECDC, EARS-Net, 2017.

En 2015, l'étude Burden BMR a étudié le poids des infections à Bactéries Multi-Résistantes (BMR)² et a estimé pour la première fois la morbidité et la mortalité des infections à bactéries multirésistantes en France en 2012. Trois chiffres clés sont présentés : il y a 158 000 cas d'infection à BMR en France par an dont 16 000 infections invasives (méningites, bactériémie) et 12 500 décès par an associés à ces infections. Plus récemment, une étude³ a analysé pour la première fois le nombre de cas, de décès et d'espérance de vie corrigées de l'incapacité (DALYs pour disability-adjusted life years) imputable aux infections causées par des bactéries résistantes aux antibiotiques dans l'Union Européenne et l'Espace économique européen en 2015. La Figure 3, issue de cette étude, nous montre l'association entre le nombre de cas, le nombre de morts et l'espérance de vie corrigées de l'incapacité liée aux infections causées par des BMR en Europe. L'étude estime qu'environ 33 000 personnes décèdent chaque année des suites directes d'une infection due à une bactérie résistante aux antibiotiques et que le fardeau de ces infections est comparable à celui combiné de la grippe, de la tuberculose et du VIH. En France, il est estimé que 24 806 patients par an développent une infection liée à une bactérie résistante et que 5543 d'entre eux décèdent des suites de ces infections. Ces études confirment le problème actuel lié à l'antibiorésistance et le besoin urgent de renforcer la mobilisation contre cette problématique inquiétante. Actuellement une approche dite « One health », une seule santé, consiste en la collaboration de différentes disciplines (santé humaine, santé animale, environnement) pour la gestion et la prévention des crises sanitaires et notamment pour la lutte contre l'antibiorésistance.

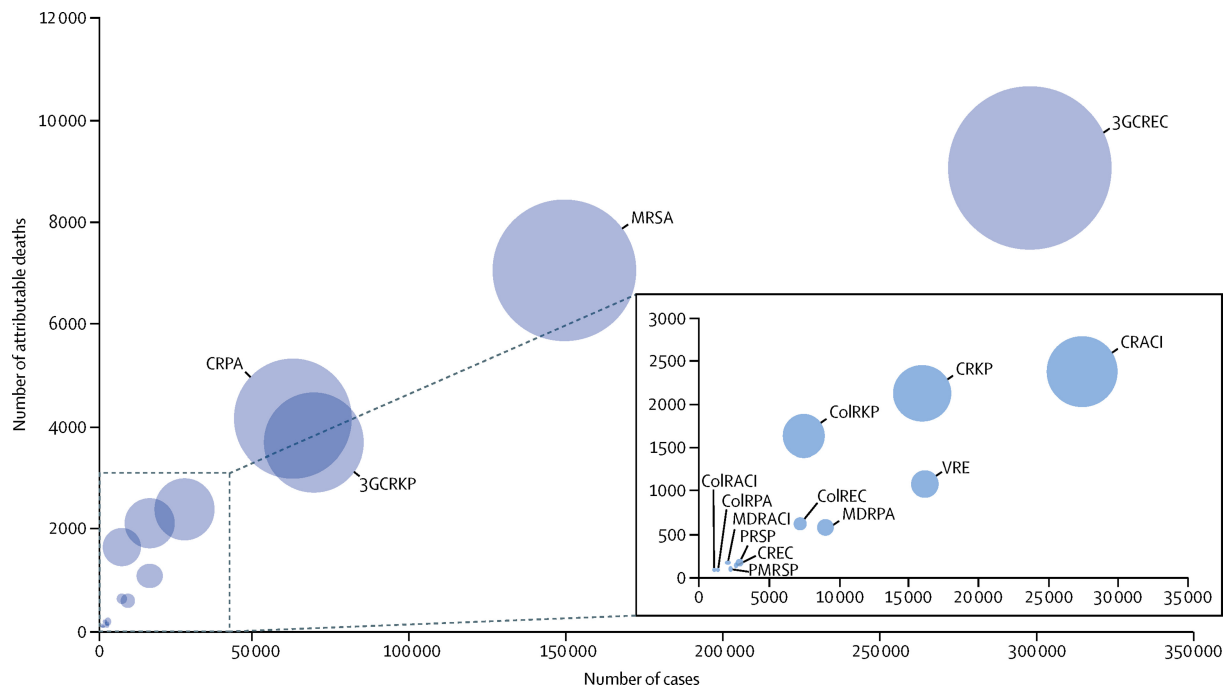


Figure 3 : Association entre le nombre de cas d'infections causées par des bactéries résistantes aux antibiotiques et le nombre de morts attribuable à la résistance aux antibiotiques dans l'UE et l'Espace économique européen en 2015, d'après ³.

ColRACI=colistin-résistant *Acinetobacter* spp. CRACI=carbapénem-résistant *Acinetobacter* spp. MDRACI=multidrog-résistant *Acinetobacter* spp. VRE=vancomycine-résistant *Enterococcus faecalis* and *Enterococcus faecium*. ColREC=colistin-résistant *Escherichia coli*. CREC=carbapénem-résistant *E. coli*. 3GCREC=third-generation cephalosporin-résistant *E. coli*. ColRKP=colistin-résistant *Klebsiella pneumoniae*. RKP=carbapénem-résistant *K. pneumoniae*. 3GCRKP=third-generation cephalosporin-résistant *K. pneumoniae*. ColRPA=colistin-résistant *Pseudomonas aeruginosa*. CRPA=carbapénem-résistant *P. aeruginosa*. MDRPA=multidrog-résistant *P. aeruginosa*. MRSA=meticilline-résistant *Staphylococcus aureus*. PRSP=penicilline-résistant, *Streptococcus pneumoniae*. PMRSP=penicilline-résistant and macrolide-résistant *S. pneumoniae*

1.1.2 Approche écologique

La question des mécanismes d'émergence de la résistance bactérienne est complexe. La résistance bactérienne peut être due : (i) à des mutations dans le génome de la bactérie (ii) à l'acquisition de gènes de résistance par transfert génétique.

Chez l'humain, la sélection de souches résistantes aux antibiotiques se produit vraisemblablement peu dans les foyers infectieux (bien que possible pour des infections chroniques, e.g: mucoviscidose, tuberculose) mais surtout au niveau des flores commensales, particulièrement du microbiote intestinal⁴. En effet, le microbiote intestinal est un milieu complexe hébergeant 10^{13} à 10^{14} bactéries appartenant à plus de 500 espèces ⁵. Les grandes tailles de populations résidant dans un milieu physiologique, la promiscuité bactérienne, ainsi que l'afflux constant de bactéries du milieu extérieur provenant de l'environnement ou d'autres hôtes y favorisent l'émergence de l'antibiorésistance. La pression de sélection exercée par les antibiotiques va ensuite sélectionner les souches antibiorésistantes. Cette pression de sélection s'exerce chez l'humain dans le cadre de l'usage (ou du mésusage) des antibiotiques mais également

chez les animaux d'élevage où la prescription est particulièrement importante ^{6,7} (Figure 4). De nombreuses études épidémiologiques chez des patients ⁸⁻¹⁰, des sujets sains ¹¹ ou chez des volontaires sains recevant des antibiotiques¹², ont documenté ce schéma. Par exemple, chez des volontaires sains recevant de la ciprofloxacine (fluoroquinolone) il a été observé une émergence de la résistance de 25% des *E. coli* dans la flore fécale, majoritairement lorsque les concentrations de ciprofloxacine dans les selles étaient inférieures à la concentration minimale inhibitrice¹².

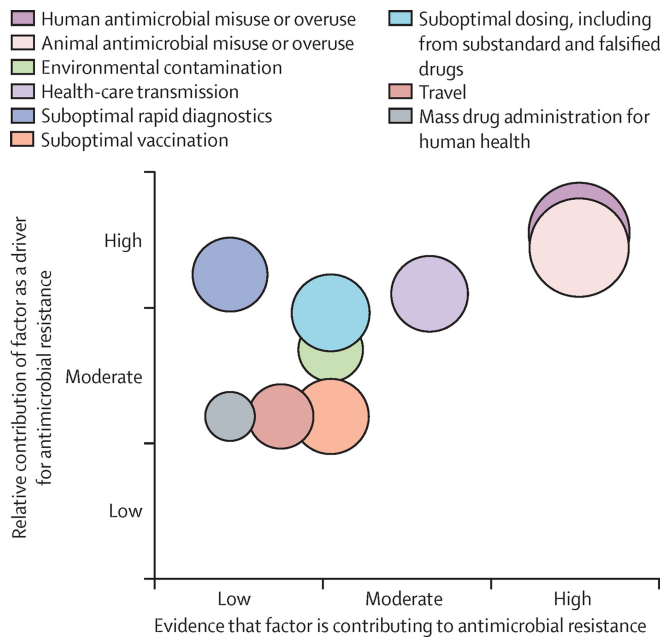


Figure 4: Rôle des différents facteurs dans leur implication dans la résistance aux antibiotiques, d'après 7.

1.1.3 Éléments génétiques mobiles

Deux notions importantes en écologie pour comprendre la dissémination des gènes de résistances aux antibiotiques sont illustrées dans la Figure 5 : (i) la notion d'émergence lorsqu'un gène de résistance passe, par exemple, d'une bactérie environnementale à une autre bactérie dans un écosystème humain (flore digestive). Pour illustrer cela, Tacao *et al.*, montrent que les gènes de résistance comme ceux codant pour les bêta-lactamases à spectre élargi (BLSE) épidémiques de type CTX-M actuellement retrouvés chez *E. coli* proviennent de bactéries de l'environnement (ex : *Kluyvera*) non en contact par des antibiotiques humains ou animaux ¹³ (ii) la notion de transmission de gènes de résistance aux antibiotiques entre deux bactéries au sein du

même écosystème (ex : flore digestive humaine). Le transfert de gènes de résistance de bactéries commensales à des bactéries pathogènes a été décrit dès 1960 par des chercheurs japonais entre *Escherichia coli* et *Shigella dysenteriae*, un *E. coli* pathogène obligatoire responsable de la shigellose particulièrement mortelle dans les pays en développement ¹⁴.

Par quels mécanismes les gènes de résistance circulent-ils entre les bactéries ? Ces gènes de résistance peuvent être transmis à des bactéries humaines par transferts horizontaux (Figure 6) :

- Grâce à des éléments génétiques mobiles conjugatifs, comme les plasmides ou des éléments intégratifs conjugatifs.
- Par transformation naturelle. Il faut ainsi que la bactérie receveuse soit en état de compétence, cet état peut être naturel ou acquis (induit en laboratoire).
- Par l'action des bactériophages ^{15,16}. La transduction correspond au transfert de matériel génétique entre deux bactéries via un phage. Ce mécanisme sera précisément détaillé dans la suite de l'introduction.

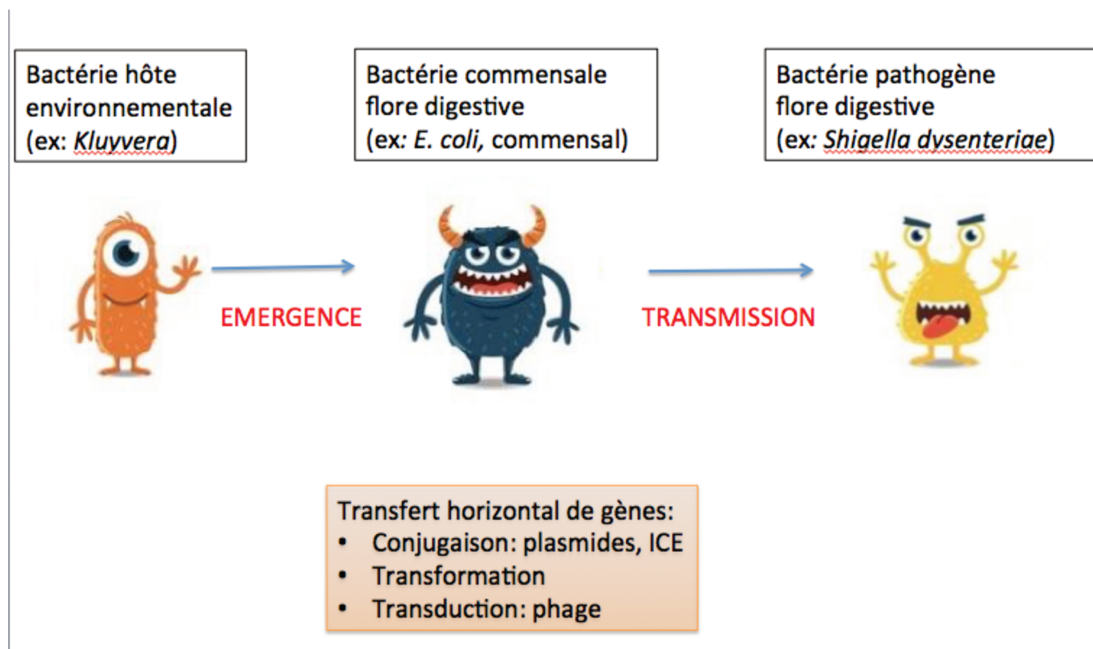


Figure 5: Schéma récapitulatif des notions d'émergence et de transmission de gènes de résistance aux antibiotiques.

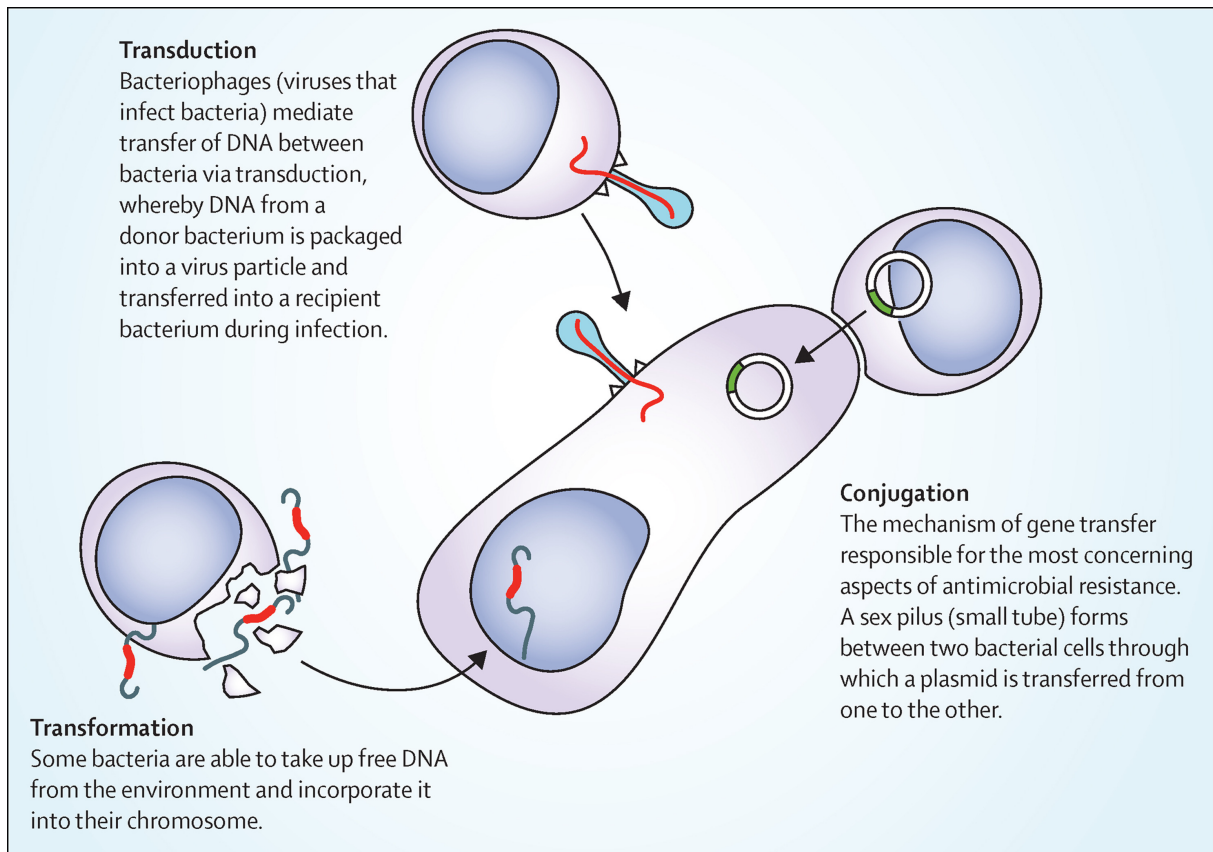


Figure 6: Schéma explicatif des différents mécanismes de transferts horizontaux de matériels génétiques entre deux bactéries d'après 7.

1.2 Le microbiote intestinal

1.2.1 Présentation et composition

Un microbiote est un ensemble de micro-organismes (bactéries, archaea, virus eucaryotes et procaryotes, parasites et champignons) qui vivent dans un environnement spécifique. Il peut être apparenté à une communauté. Chez l'être humain il existe différents microbiotes : par exemple oropharyngé, pulmonaire, vaginal, cutané et intestinal. Le microbiote intestinal est celui qui renferme le plus de micro-organismes : 10^{12} à 10^{14} , soit le même ordre de grandeur que le nombre de cellules qui constituent notre corps, pour un poids de 2 Kg. La composition du microbiote intestinal évolue en fonction de la localisation : estomac, duodenum, jéjunum, iléon, colon. En effet la différence de pH et d'oxygénation dans ces différentes parties du tube digestif entraîne une différence de composition du microbiote. Tout au long du tractus digestif, il existe un gradient de concentration et de richesse en bactéries: il est retrouvé moins de 10^1 bactéries/g dans l'estomac (en raison du pH acide), 10^3 bactéries/g dans le duodénum,

10^4 - 10^7 bactéries/g dans le jéjunum et l'ileum et jusqu'à 10^{11} - 10^{12} bactéries/g dans le caecum et le colon ¹⁷.

Pendant longtemps il a été très difficile de connaître la composition du microbiote intestinal (au niveau bactérien) car seule une minorité d'espèces bactériennes du microbiote pouvait être cultivée *in vitro*. Cela fait seulement une dizaine d'années que l'apparition de techniques de séquençage à haut débit a permis de décrire avec plus de précision la composition du microbiote intestinal. La métagénomique a fait son apparition : c'est une méthode d'étude du contenu génétique d'échantillons complexes (tel que le microbiote intestinal) via un séquençage direct de l'ADN présent dans l'échantillon. De grands programmes d'analyse du métagénome intestinal humain ont été entrepris ces 10 dernières années tel que MetaHIT ¹⁸. La partie la plus documentée est la fraction bactérienne du microbiote intestinal : elle comprend plus de 500 espèces, dont la majorité est anaérobie stricte et où le nombre de gènes bactériens a été récemment estimé à plus de 3 millions ¹⁸. Ces espèces bactériennes se répartissent en 4 phyla majeurs : Firmicutes, Bacteroidetes, Actinobacteria et Proteobacteria. Cette faible diversité au niveau des phyla est compensée par une importante variabilité de genres et d'espèces au sein de ces groupes.

Concernant les autres micro-organismes présents dans le microbiote intestinal, voici une vision d'ensemble des connaissances actuelles (Figure 7) :

-Deux archaea ont été caractérisées comme colonisant le colon humain : *Methanobrevibacter smithii* ¹⁹ et *Methanosphaera stadtmanae* ²⁰.

- Plusieurs familles de virus eucaryotes ont été décrites : Picobirnaviridae, Adenoviridae, Anelloviridae and Astroviridae ainsi que des espèces telles que bocavirus, enterovirus, rotavirus and sapovirus ²¹. Concernant les virus procaryotes (bactériophages), ils sont présents en grande quantité. La composition en bactériophages du microbiote intestinal sera détaillée ultérieurement dans un autre chapitre dédié à cette question.

- Les champignons représentent une faible part estimée entre 0.01% et 0.1% des gènes totaux retrouvés dans les selles ²². Dans la plupart des études, Ascomycota est de loin le phylum le plus prédominant avec l'espèce *Candida tropicalis* majoritairement présente, suivie par Zygomycota et Basidiomycota ²³.

- Les protistes représentent une très faible part de la composition du microbiote intestinal.

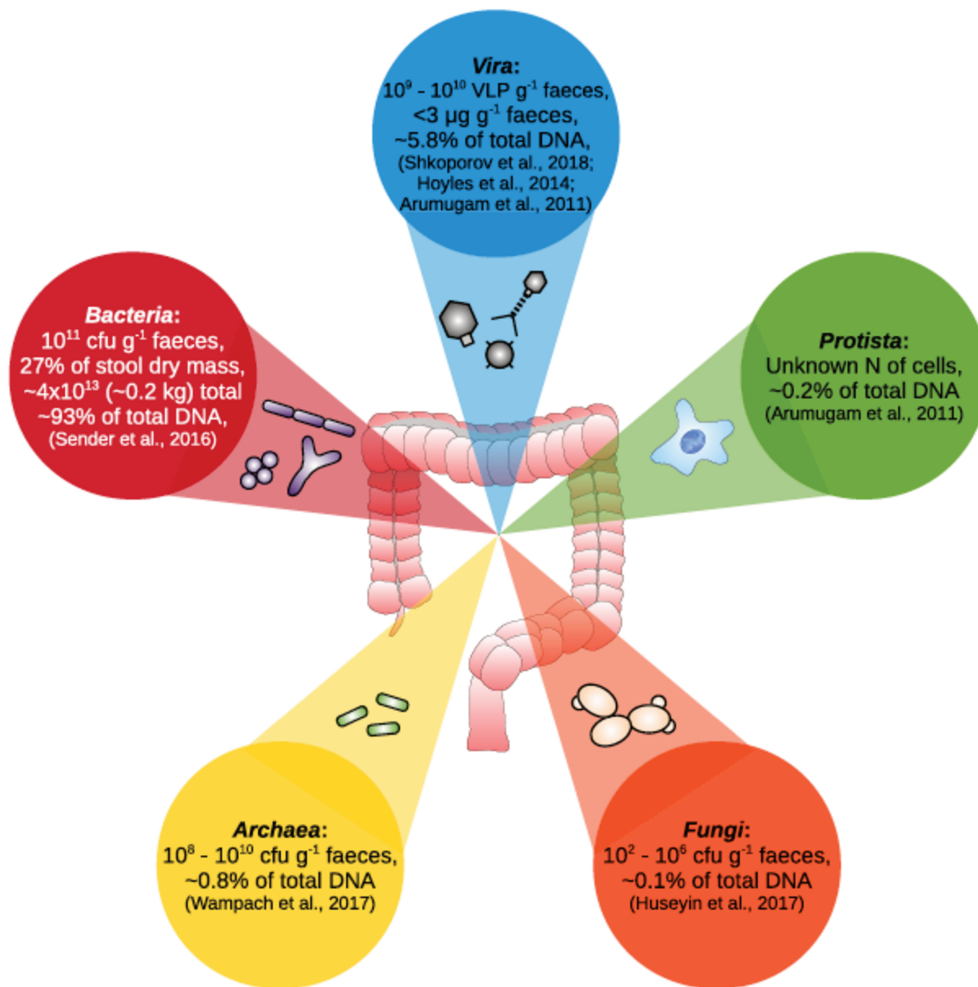


Figure 7: Principaux groupes taxonomiques présents dans le microbiome intestinal humain d'après ²⁴.

Quels sont les rôles du microbiote intestinal ?

Le microbiote intestinal est donc une communauté majeure qui :

- est indispensable à l'acquisition et au maintien des fonctions digestives.
- joue un rôle essentiel au niveau des systèmes immunitaire, intestinal et périphérique.
- maintient la barrière naturelle que sont nos muqueuses.
- agit contre la colonisation du tube digestif par des germes pathogènes.

Une dysbiose ou rupture de l'équilibre du microbiote intestinal peut être due à différents facteurs tels que : une alimentation déséquilibrée, une prise de médicaments (ex : antibiotiques), du stress ou encore un régime alimentaire inapproprié. Une

dysbiose peut avoir des conséquences fonctionnelles importantes et peut-être impliquée dans de nombreuses pathologies digestives (maladies inflammatoires chroniques intestinales, cancer colorectal, etc.) mais aussi dans l'obésité et l'autisme. Ces observations ont conduit à l'émergence de nombreuses études sur les traitements visant à restaurer l'équilibre du microbiote intestinal comme les probiotiques ou la transplantation du microbiote fécal ²⁵.

1.2.2 Méthode d'analyse du microbiote intestinal

Afin de pouvoir étudier la composition du microbiote intestinal il est nécessaire de disposer de techniques permettant de rechercher et d'analyser les différents micro-organismes.

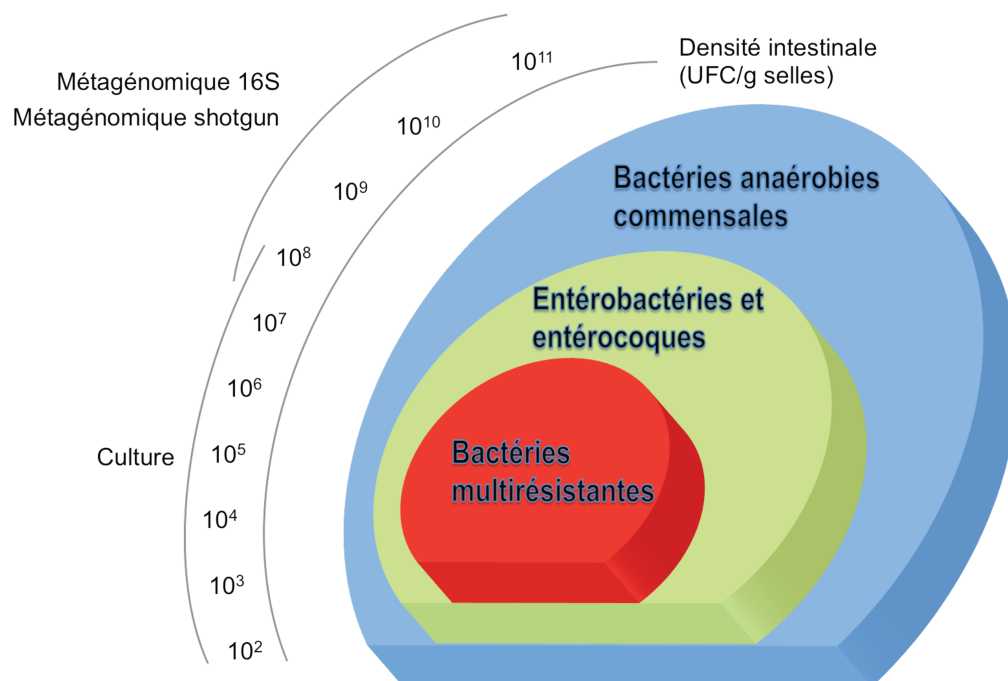


Figure 8: Illustration des différentes techniques d'analyse du microbiote intestinal, de leur sensibilité et du type de bactéries détectables d'après ²⁶.

La culture bactérienne

La culture bactérienne est la méthode la plus ancienne. Elle a l'avantage d'être simple, peu coûteuse et de donner accès à l'ensemble du potentiel génétique des bactéries isolées. Elle est encore très utilisée dans les laboratoires de bactériologie afin de diagnostiquer le portage dans la flore digestive de bactéries résistantes aux

antibiotiques appelés BMR (bactérie multi résistante). L'utilisation de milieux spécifiques enrichis en antibiotiques permet la détection de ces bactéries pouvant être présentes en faible quantité. En effet, cette technique est très sensible et peut permettre la détection de bactéries sous dominantes (ex : bactéries multirésistantes présentes en faible quantité dans un échantillon de selle). Ceci est illustré dans la Figure 8. Cette méthode reste utilisée pour analyser l'impact des antibiotiques sur les flores afin de comparer l'abondance relative de la présence de bactéries résistantes avant et après le traitement. Cependant, cette méthode se limite aux espèces dont la culture est possible au laboratoire, alors que le nombre d'espèces bactériennes de la flore intestinale non cultivables est estimé à 80% ²⁷. Récemment, la culture bactérienne connaît un renouveau grâce aux études de Browne *et al.* ²⁸ et Lagier *et al.* ²⁹ qui s'intéressent à la culturomique. Le principe est de combiner de la culture bactérienne dans des conditions particulières d'atmosphère (ex : anaérobie, microaérophilie) ou sur des milieux spécifiques (ex : milieux avec de l'éthanol afin de sélectionner les bactéries sporulantes) avec du séquençage de génome entier. Cette technique a permis de mettre en évidence de nombreuses nouvelles espèces bactériennes.

La PCR ciblée 16S

Cette approche moléculaire est basée sur l'étude du gène 16S codant pour la sous-unité 16S de l'ARN ribosomique (ARNr). L'ARNr 16S est une macromolécule ubiquitaire et sa structure est très conservée chez toutes les bactéries. D'une longueur d'environ 1500 nucléotides, l'ARNr 16S possède une structure mosaïque alternant régions conservées (identiques pour toutes les bactéries) et régions variables et hyper-variables (spécifiques de chaque espèce bactérienne). Les amorces de la PCR sont donc choisies dans des régions conservées encadrant une région variable, région qui sera alors séquencée (V1-V9) (Figure 9). Le choix de la région variable séquencée varie en fonction des études. Cette technique est donc utilisée pour connaître la composition bactérienne du microbiote intestinal. Après l'amplification et le séquençage haut débit de la région codant pour l'ARN 16S, s'en suit un travail bio-informatique afin de regrouper les séquences en unités taxonomiques opérationnelles (OTU), ou taxons, sur la base de leur homologie de séquence (seuil à 97%) puis l'attribution taxonomique peut être faite par comparaison avec les séquences disponibles dans des bases de données. Le pouvoir résolutif de l'analyse de la séquence de cette molécule permet donc l'étude du règne Procaryote dans sa globalité (richesse, diversité etc.) et à tous les niveaux

taxonomiques (composition).

Les méthodes basées sur l'analyse du gène codant l'ARNr 16S font intervenir une étape d'amplification de l'ADN (par PCR) qui ne permet pas de détecter des bactéries présentes en dessous de 10^{5-7} CFU/g de selles de l'échantillon de départ. De plus, il n'est pas possible de faire une quantification bactérienne avec cette méthode car les gènes codant pour l'ARNr peuvent être en plusieurs copies par génome et le nombre de copies varie en fonction des espèces bactériennes. La composition taxonomique de chaque échantillon n'est possible qu'en abondance relative ³⁰. Enfin, certaines espèces restent difficiles à caractériser uniquement par la séquence ADNr 16S, comme les espèces des genres *Streptococcus* ou *Enterobacter*, car les divergences des régions ADNr 16S sont inférieures à 0,5% entre les espèces de ces genres ³¹.

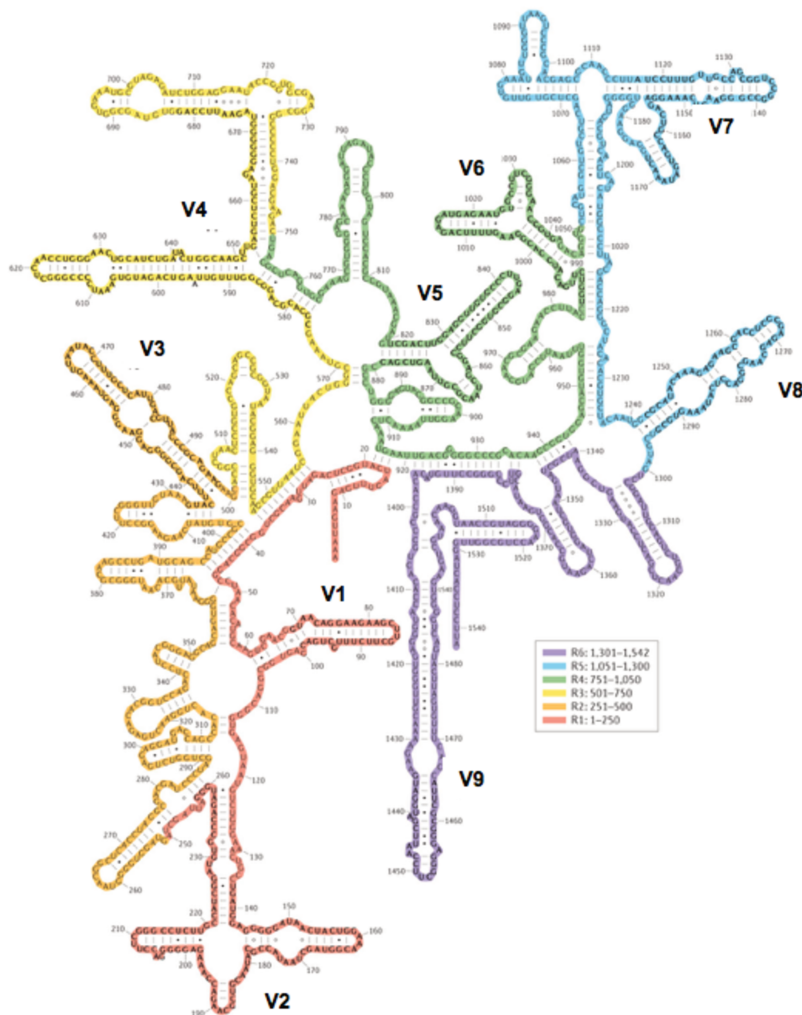


Figure 9: Structure secondaire de l'ARNr 16s d'*Escherichia coli* d'après ³². Il est constitué de 9 régions hypervariables (V1-V9) propre aux espèces bactériennes séparées par des régions conservées permettant la fixation d'amorces universelles.

L'analyse métagénomique « shotgun »

La métagénomique permet l'analyse du potentiel génétique d'un écosystème dans sa globalité en séquençant massivement tout l'ADN présent dans un échantillon sans se focaliser sur une région particulière. Il y a donc ensuite un important travail de bio-informatique afin de nettoyer les données et de reconstruire les séquences de gènes. Dans un premier temps, cette approche a été appliquée à l'analyse des sols. L'intérêt grandissant pour le microbiote intestinal a favorisé l'adaptation de la métagénomique à l'analyse de l'écosystème digestif humain dans le cadre de deux projets de grande envergure : aux Etats-Unis avec le « Human Metagenome Project » (HMP) et en Europe avec le projet « Metageno-mics of the Human Intestinal Tract » (MetaHIT). L'étude MetaHIT a donné lieu à la création d'un catalogue de gènes non redondants potentiellement présents dans le microbiote intestinal (fait à partir de 1267 individus issus de trois continents) comprenant 10 millions de gènes ¹⁸.

Les études de cohortes indiquent également que la richesse (marqueur de diversité) en gènes bactériens (et donc en espèces) caractérise l'écosystème intestinal. Certains individus ont un microbiome pauvre ou atrophié, comptant moins de 300 000 gènes, tandis que d'autres, au contraire, ont un microbiome riche de plus de 800 000 gènes ³³.

La métagénomique est néanmoins pourvue de biais méthodologiques liés à la préparation de l'échantillon initial (ex : extraction de l'ADN) et à la méthode de séquençage (ex : technologie employée, taille des reads). Ces étapes sont primordiales et il est nécessaire de les homogénéiser afin de pouvoir comparer les études entre-elles. Ainsi l'étude de Costea *et al.* propose des standards pour l'extraction et l'analyse d'ADN bactérien ³⁴.

1.2.3 Effet des antibiotiques sur le microbiote intestinal

Les traitements antibiotiques perturbent l'équilibre entre les différents groupes bactériens retrouvés dans le microbiote intestinal. Les modifications qualitatives et quantitatives de la composition taxonomique du microbiote intestinal suite à l'administration d'antibiotiques ont été régulièrement rapportées ^{35,36}. Cela contribue à

un déséquilibre, une dysbiose, pouvant entraîner des effets négatifs tels que l'infection à *Clostridioides difficile* ou l'émergence de bactéries résistantes aux antibiotiques.

Infection à *Clostridioides difficile*.

C. difficile est un bacille à Gram positif anaérobie strict, sporulant, produisant des toxines, isolé pour la première fois par Hall & O'Toole ³⁷ dans l'intestin d'un nouveau-né sain. L'intérêt pour cet organisme s'est accru lorsqu'il a été démontré qu'il était l'agent causal de la plupart des cas de colite infectieuse post-antibiotique.

Les antibiotiques ont des effets significatifs et durables sur le microbiote intestinal et réduisent la résistance à la colonisation contre les pathogènes. La physiopathologie de l'infection à *C. difficile* a été abondamment étudiée à l'aide de modèles animaux, notamment chez le hamster doré. L'idée principale est qu'en l'absence de traitement antibiotique, les acides biliaires primaires sont transformés en acides biliaires secondaires et inhibent la sporulation et la croissance de *C. difficile* au sein du microbiote intestinal. En cas de traitement antibiotique, il y a une modification de la structure du microbiote intestinal conduisant à la perte du métabolisme secondaire des acides biliaires, cela favorise la croissance du *C. difficile*, la production de toxines et la maladie ³⁸.

Le taux de mortalité lié à cette infection a été estimé entre 15 et 20% et le poids micro-économique est lourd avec une durée de séjour à l'hôpital 33% plus longue en cas d'infection, durée pouvant être multipliée par quatre en cas de récurrence infectieuse ³⁹.

Emergence de bactéries résistantes aux antibiotiques

Un effet majeur des antibiotiques est d'altérer l'effet barrière principalement assuré par les bactéries anaérobies du microbiote intestinal. Cela conduit à une augmentation de densité de bactéries normalement sous-dominantes comme des entérobactéries résistantes aux bêta-lactamines ou les entérocoques résistants à la vancomycine ⁴⁰. Cette augmentation conduit (i) à un risque plus élevé de translocation digestive (passage d'une bactérie de la flore intestinale à la circulation sanguine) et ainsi d'infection ⁴¹, (ii) à une probabilité supérieure de contracter une infection urinaire ⁴² (iii) ou encore à une diffusion de bactéries résistantes dans l'environnement. Perez *et al.* en 2011 montrent que l'implantation de souches bactériennes productrices de carbapénemases est favorisée par les antibiotiques actifs sur les bactéries anaérobies à

forte élimination biliaire comme la piperacilline-tazobactam ou la clindamycine ⁴³.

Les récents progrès spectaculaires d'analyse des flores commensales ont permis une meilleure visualisation des modifications du microbiote intestinal induit par les antibiotiques. Plusieurs études se sont intéressées à cette thématique, chez la souris et chez l'humain avec des analyses par culture classique ou par métagénomique :

- En 1990 une première étude chez la souris met en évidence, par culture bactérienne classique, l'émergence de la résistance au triméthoprimé chez les *E. coli* après administration orale de ce même antibiotique ou de l'association triméthoprimé - sulfaméthoxazole ⁴⁴

- Dethlefsen et *al.* en 2011, grâce à une approche métagénomique, ont montré que chez des sujets sains traités par ciprofloxacine (quinolone) pendant 5 jours il était observé une modification du microbiote intestinal avec une réduction de la diversité bactérienne ⁴⁵. Cette réduction de la diversité est observée au long cours (résilience incomplète) de même qu'une réduction de la résistance à la colonisation.

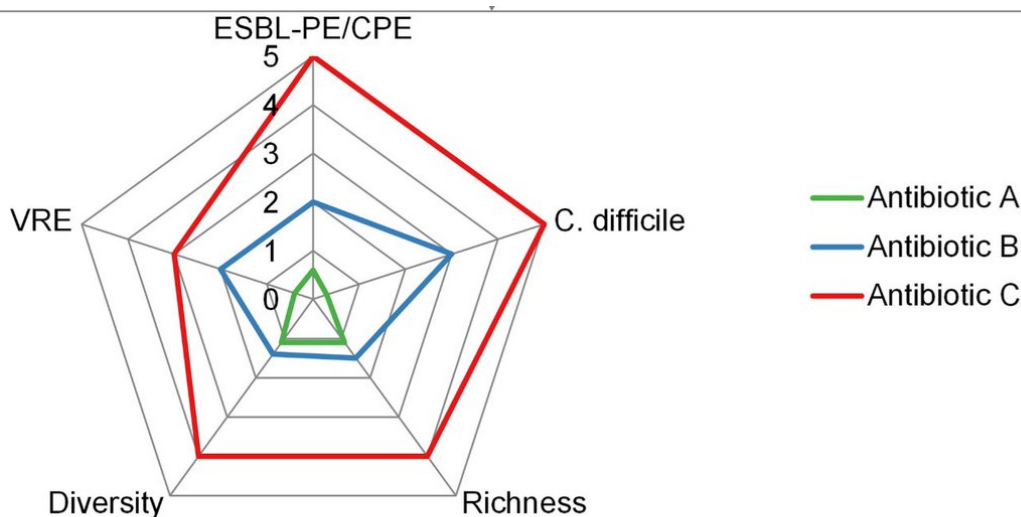
- Une étude menée chez l'humain auprès de 48 sujets sains traités pendant 14 jours par une fluoroquinolone, la ciprofloxacine, a montré qu'un tiers des sujets traités étaient colonisés avec des entérobactéries résistantes à la ciprofloxacine un mois après la fin du traitement, contre 12.5% avant traitement ¹².

Actions proposées

La préservation de l'effet barrière du microbiote est donc indispensable pour permettre une résistance à la colonisation par des bactéries multi-résistantes ou des bactéries pathogènes (ex : *C. difficile*). C'est pourquoi il faut limiter au maximum la consommation d'antibiotique et lorsque celle-ci est nécessaire il faut ajuster au mieux la prescription. Ainsi à l'hôpital, l'application de la désescalade antibiotique est primordiale : il s'agit de réévaluer chez les patients l'antibiothérapie prescrite initialement (antibiothérapie probabiliste, avant documentation du germe, souvent à large spectre) par une antibiothérapie ciblée grâce la documentation de la bactérie et de son antibiogramme. Actuellement ces procédures prennent 48h mais les révolutions technologiques qui s'opèrent dans les laboratoires de bactériologie laissent à penser/espérer que ce délai va drastiquement se raccourcir dans les années à venir.

L'impact d'un antibiotique sur le microbiote intestinal dépend de son spectre antibactérien et de sa voie d'élimination (et ainsi de sa concentration dans le tube digestif). Comme discuté dans l'article de Ruppé *et al.* ⁴⁶ la notion d'impact reste à définir

précisément et a besoin d'être standardisée afin d'avoir des études comparables pour les antibiotiques disponibles et ceux à venir. Ainsi, il est proposé une analyse de l'impact d'un antibiotique sur le microbiote intestinal défini par son action (i) sur la richesse bactérienne, (ii) sur la diversité bactérienne, (iii) sur l'émergence de bactéries hautement résistantes (BHR) tel que les entérocoques résistants à la vancomycine ou les entérobactéries productrices de carbapénemases, (iv) sur la colonisation a *C. difficile*, avant et après la prise de traitement (Figure 10).



Ranking antibiotics according to their impact on the intestinal microbiota

Figure 10: Proposition pour l'évaluation de l'impact des antibiotiques sur le microbiote intestinal d'après ⁴⁶. VRE : vancomycin-resistant enterococci, ESBL-PE : extended-spectrum beta-lactamase-producing *Enterobacteriaceae*, CPE : carbapenemase-producing *Enterobacteriaceae*, NDRB : multidrug-resistant bacteria.

Ces dernières années, afin de restaurer l'équilibre de la flore fécale chez certains patients très altérés et pénalisés, le traitement par la transplantation de flore fécale a émergé. Ce traitement est déjà utilisé pour les patients avec des infections récurrentes à *Clostridioides difficile* avec d'excellents résultats ⁴⁷ et montre des premiers résultats encourageants pour les patients porteurs de BHR, surtout pour la décolonisation des *Pseudomonas auruginosa* toto-résistants ⁴⁸.

1.3 Les phages

1.3.1 Généralités sur les phages : définition, composition, classification

Un phage est un virus infectant des bactéries, il est ainsi appelé bactériophage. Ils ont été découverts par Frederick William Twort en 1915 puis redécouverts en 1917 par

Félix d'Hérelle qui leur donna ce nom. Le terme bactériophage est formé à partir de deux mots d'origine grecque : baktêria qui signifie «bâton» et phagos qui signifie «manger»; la notion de « bâton » faisant référence, pour l'époque, aux bactéries.

Les phages sont présents dans l'ensemble de la biosphère : dans les eaux usées ($>10^9$ « virus-like particles » VLP/ml) ⁴⁹, le sol (10^7 - 10^8 VLP/ml) ⁵⁰ ou encore les selles (10^7 à 10^{10} VLP/ml) ⁵¹. Le phage a un rôle important dans la dynamique des populations bactériennes. On estime le nombre de bactériophages dans la biosphère à environ 10^{31} particules soit dix à cent fois plus que le nombre total de bactéries⁵². L'ordre de grandeur de la taille de ces bactériophages est de 20 à 200 nanomètres soit dix à cent fois plus petits que les bactéries.

Le bactériophage est constitué systématiquement d'un seul type d'acide nucléique porteur de l'information génétique (ADN ou ARN) et d'une capsidie formée de protéines qui sont arrangées selon un motif précis permettant de protéger l'acide nucléique. On retrouve dans certains cas une enveloppe constituée de lipides et de protéines virales qui entoure la nucléocapsidie. On parle alors de virus nus (résistants) ou de virus enveloppés (fragiles) ⁵³. Les bactériophages sont classés en fonction de la nature de leur acide nucléique (ADN ou ARN) et de leur structure générale (symétrie de la capsidie, présence ou non d'enveloppe). A l'heure actuelle, la dernière édition de taxonomie virale proposée par l'ICTV (International Committee on Taxonomy of Viruses) ⁵⁴ permet de classer environ 5500 phages différents (Figure 11).

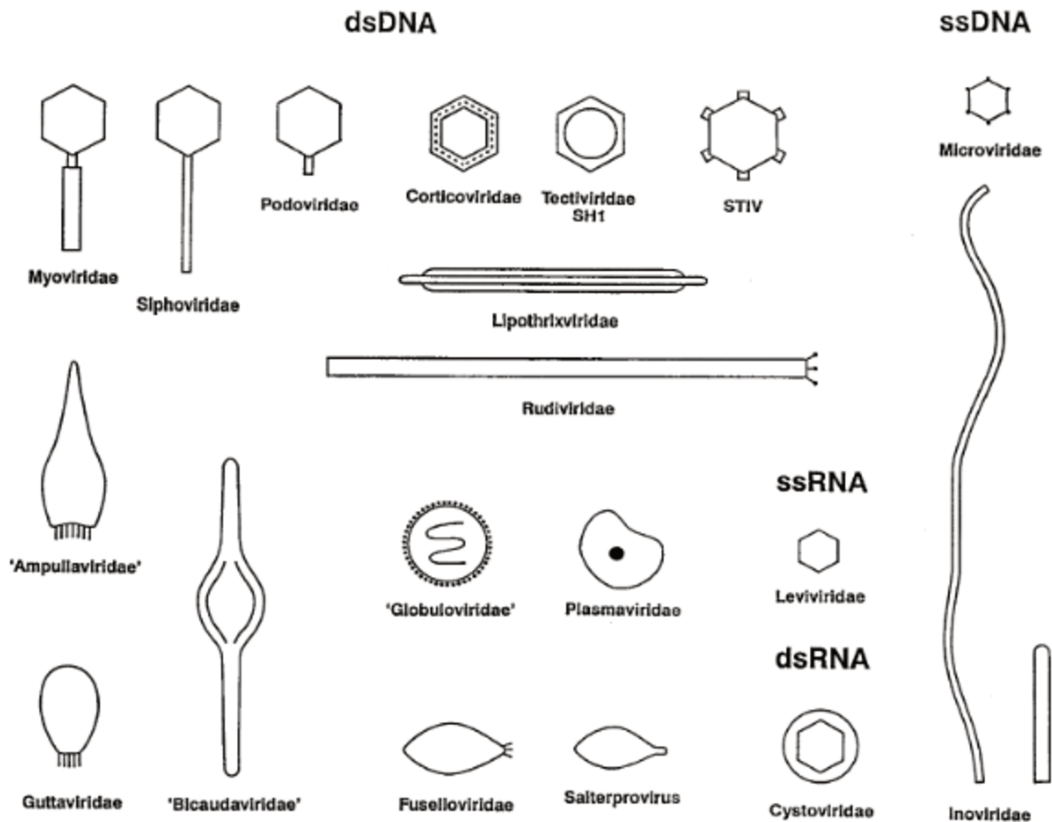


Figure 11: Morphotypes des différentes familles de bactériophages, d'après ⁵⁵.

Phage à ADN double brin

Dans la catégorie des phages ADN double brin (dsDNA) il y a deux grands ordres (les Caudovirales et les Ligamenvirales) qui se distinguent par la symétrie de leurs capsides respectivement icosaédriques et hélicoïdales. Les autres familles (Corticoviridae, Tectiviridae, Ampullaviridae, Guttaviridae, Bicaudaviridae, Globuloviridae, Plasmiviridae, Fuselloviridae, Salterprovirus) ne sont pas rattachées à un ordre et c'est la géométrie de leur capside qui permet de les classer ainsi ⁵⁴.

- Ordres des Caudovirales

L'ordre des Caudovirales regroupe la majorité des bactériophages connus (96%). Les caractéristiques de cet ordre sont des phages avec un ADN double brin linéaire encapsidé dans une tête à symétrie icosaédrique et la présence d'une queue à symétrie hélicoïdale (Figure 12).

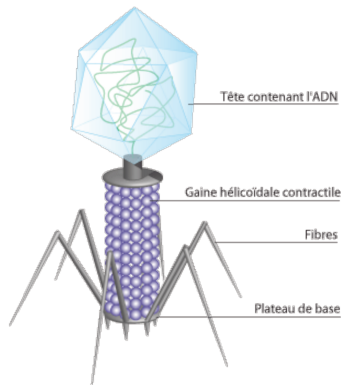


Figure 12: représentation schématique typique d'un bactériophage de la famille des Myoviridae.

Cet ordre est divisé en 3 familles (Figure 13):

- les Myoviridae : caractérisées par la présence d'une queue contractile (exemple de phage dans cette famille : phage T4).
- les Siphoviridae : caractérisées par la présence d'une queue longue mais non contractile (exemple de phage dans cette famille : phage lambda).
- les Podoviridae : caractérisées par la présence d'une queue courte et non contractile (exemple de phage dans cette famille : phage T7).

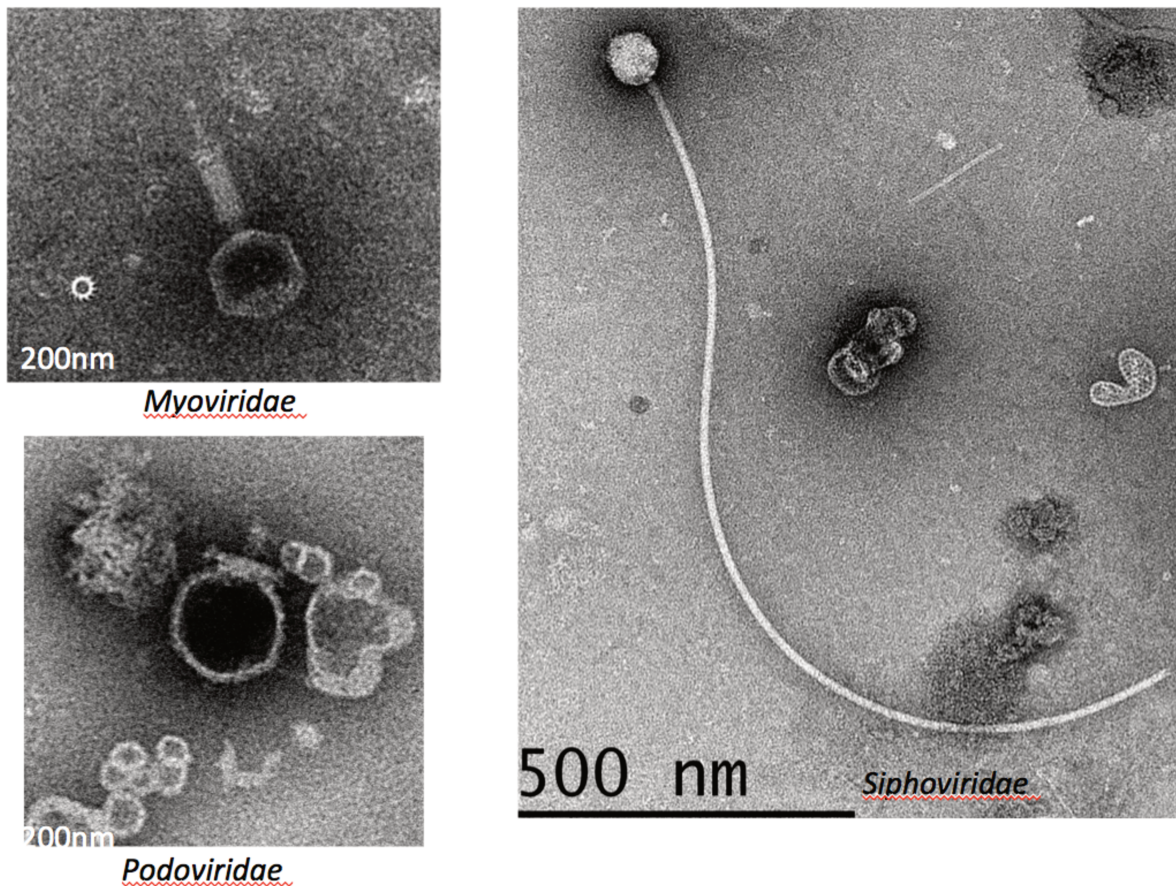


Figure 13: Photos prise au microscope électronique (Institut Pasteur). Crédits: C. d'Humières, L. Debarbieux

- Ordres des Ligamenvirales

Cet ordre se caractérise par une capsid à symétrie hélicoïdale, il regroupe la famille des Lipothrixiviridae et des Rudivirodae.

Phage à ADN simple brin

Deux familles de phages dans cette catégorie ⁵³ :

- les Inoviridae : phages à ADN monocaténaire, circulaires et petits (27 nm de diamètre). Ils sont dépourvus d'enveloppe et de queue. Leur capsid est à symétrie hélicoïdale, ils sont également appelés phages filamenteux (ex : phage M13)
- les Microviridae : phages à ADN monocaténaire, circulaires et possédant une petite capsid cubique non enveloppée. Le petit phage ϕ X174 dont le génome comprend 5386 paires de bases appartient à cette famille, il a été le premier génome séquencé en 1977.

Phage à ARN double brin

Une seule famille de phages dans cette catégorie, les Cystoviridae⁵³. Ce sont des phages à ARN bicaténaire, linéaire, segmenté et possédant une enveloppe lipidique (ex : le phage $\phi 6$).

Phage à ARN simple brin

Une seule famille de phages également dans cette catégorie, les Leviviridae⁵³. Ce sont des phages à ARN monocaténaire, linéaire et ne possédant pas d'enveloppe (ex : le phage MS2).

1.3.2 Etape d'adsorption : spécificité d'hôte des phages

La première étape de l'infection est la reconnaissance d'un récepteur spécifique à la surface de la bactérie hôte par la machinerie d'adsorption du phage. L'adsorption bactériophagique est donc non seulement une étape cruciale dans le processus d'infection, mais elle représente aussi le point de contact initial entre le virus et l'hôte et dicte la spécificité de la gamme d'hôtes. L'adsorption du phage se compose généralement de trois étapes : le contact initial, la liaison réversible et l'attachement irréversible⁵⁶. La première étape consiste en des collisions aléatoires entre phage et hôte causées par le mouvement, la dispersion, la diffusion ou l'écoulement brownien. Dans l'étape réversible, la liaison aux composants de la surface bactérienne n'est pas définitive et le phage peut se détacher de l'hôte. Puis l'étape irréversible permet la fixation définitive du phage à sa bactérie hôte spécifique.

De nombreuses études ont mis en évidence l'étendue des récepteurs associés à l'hôte (protéines, sucres, structures de surface des cellules, pili) que les bactériophages ciblent pendant l'adsorption et une revue récente détaille tous les récepteurs des phages actuellement décrits⁵⁷.

1.3.3 Cycles de reproduction phagique

On distingue à ce jour quatre types de cycles infectieux chez les phages. Deux majoritairement présents : le cycle lytique et le cycle lysogénique. Les deux autres sont plus rares et moins étudiés : le cycle pseudolysogénique et l'infection chronique⁵².

Le cycle lytique

Il s'agit du cycle réalisé par les bactériophages ainsi qualifiés de **virulent**. Il y a en premier lieu une étape d'absorption entre le phage et la bactérie nécessitant une complémentarité étroite (le phage est hôte spécifique). Ensuite le phage injecte son génome dans le cytoplasme de la bactérie. Les premiers gènes précoces sont transcrits puis traduits par la machinerie cellulaire de l'hôte aboutissant à la synthèse de protéines virales dont la tâche est de détourner le métabolisme cellulaire de la bactérie au profit de la synthèse protéique virale. Les protéines virales sont ainsi produites en grande quantité puis assemblées et encapsidées pour former des virions matures. La dernière étape est une étape de lyse de la cellule bactérienne par destruction de l'architecture de la membrane permettant de libérer les nouvelles particules virales pouvant alors réinfecter de nouvelles bactéries hôtes ⁵⁸ (Figure 14, Figure 15)

Le cycle lysogénique

Il s'agit d'un cycle plus complexe propre aux phages **tempérés** (Figure 14). Les premières étapes (adsorption et introduction du génome dans la bactérie) sont identiques à celles décrites pour le cycle lytique. Mais dans le cas du cycle lysogénique le matériel génétique du phage s'intègre, ou non, dans le génome bactérien sous forme de prophage à l'aide le plus souvent d'une intégrase. Le prophage se réplique avec le génome bactérien et est transmis aux futures générations lors des divisions successives. La bactérie est dite lysogène et ne peut, alors, pas être infectée par le même phage (phénomène d'immunité). Cet état de lysogénie est maintenu par la synthèse de gènes répresseurs des fonctions lytiques. Le prophage peut être induit de façon spontanée mais l'induction résulte le plus souvent d'un stress (ex : antibiotiques), d'une carence, et ainsi, le prophage s'excise encapsidant son ADN et/ou l'ADN de la bactérie hôte et enclenche un cycle lytique.

A l'état de prophage, certains gènes du phage peuvent être exprimés (conversion lysogénique). Des facteurs de virulence essentiels d'un grand nombre de pathogènes sont codés dans les prophages ⁵⁹. Rocha *et al.* ont particulièrement étudié les prophages, leur mode d'intégration dans le chromosome, leur machinerie de recombinaison ⁶⁰ ainsi que les processus de domestication de leur contenu génétique par l'hôte bactérien ⁶¹. Il a été montré que 46% des génomes bactériens séquencés hébergeaient au moins un prophage de 30 Kb ⁶² et chez certaines souches bactériennes (*E. coli*) 18 génomes de bactériophages peuvent être retrouvés représentant 10% du génome bactérien ⁶³.

Le cycle pseudolysogénique

Le cycle pseudolysogénique est un état intermédiaire entre le cycle lytique et le cycle lysogénique ⁶⁴. Après injection du matériel génétique du phage dans la cellule hôte, celui-ci reste sous forme extra-chromosomique sans intégration dans le génome de la cellule hôte. Lors de la descendance bactérienne, une seule des deux cellules filles héritera du phage car il n'y aura pas eu répllication de celui-ci. Dans certain cas (ex : phage P1) une répllication est possible aboutissant à une distribution équitable lors de la division cellulaire.

Le cycle pseudolysogénique est généralement causé par des conditions de croissance défavorables pour la cellule hôte (ex : la famine) et se termine par une véritable lysogénisation ou une croissance lytique lorsque les conditions de croissance s'améliorent. La pseudolysogénie est connue depuis des dizaines d'années, mais son rôle a souvent été sous-estimé. Actuellement, elle est considérée comme un aspect important des interactions phage-hôte.

L'infection chronique

L'infection chronique est réalisée par les phages dits filamenteux (ex : M13, fd, f1). Lorsque l'acide nucléique (en général simple brin) est injecté dans la bactérie hôte, les protéines de capsidie sont insérées dans la membrane cytoplasmique de l'hôte. Il est supposé que cela sert à maintenir l'intégrité de la membrane cellulaire externe pendant l'injection du génome. Le génome, une fois dans la cellule hôte, est répliqué de nombreuses fois et les gènes nécessaires à la synthèse des protéines structurales sont exprimés. Les phages nouvellement synthétisés sont alors sécrétés par bourgeonnement de l'enveloppe bactérienne grâce à un canal spécifique. Contrairement aux autres types de phages, les phages filamenteux ne tuent pas leur hôte, mais sont libérés au fur et à mesure qu'ils se répliquent. Les bactéries infectées continuent à se diviser, transmettent le phage à la descendance et sont peu altérées par ce parasitisme chronique. Cette intéressante caractéristique en fait des outils de choix en biologie moléculaire ⁶⁵.

Phage virulent

Phage tempéré

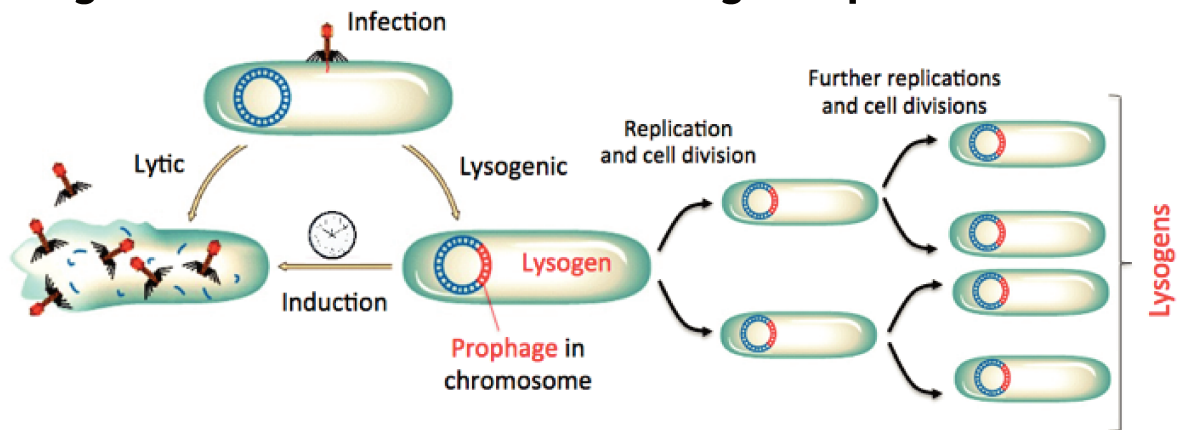


Figure 14: Schéma explicatif des différents types de phages: virulents et lytiques. M. Touchon, non publié.

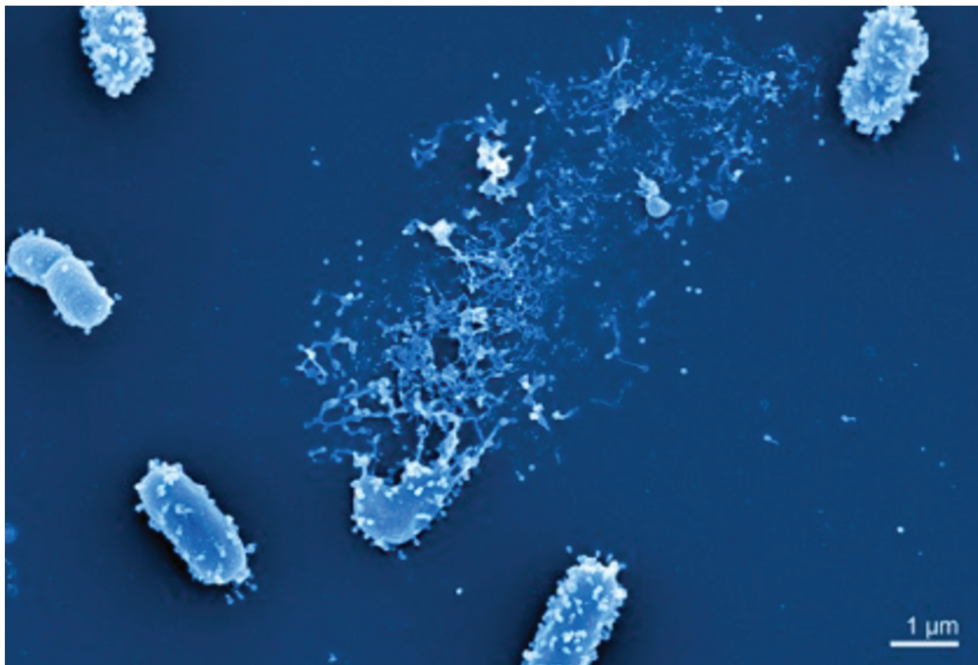


Figure 15: Photographie d'une lyse bactérienne par un phage, d'après ⁶⁶

1.3.4 Transduction

Comment les phages peuvent-ils véhiculer du matériel génétique bactérien ?

En 1952 l'expérience de Zinder et Lederberg ⁶⁷ met en évidence le transfert de matériel génétique entre deux bactéries séparées par un verre poreux empêchant le contact mais laissant passer les virus. En effet, comme nous pouvons l'observer sur la Figure 16, il y a d'un côté du tube en « U » une bactérie *Salmonella enterica* porteuse du

prophage 22A initialement auxotrophe pour le tryptophane mais autotrophe pour l'histidine et de l'autre côté il y a également une bactérie *Salmonella enterica* porteuse du prophage 2A initialement auxotrophe pour l'histidine et autotrophe pour le tryptophane. L'expérience a montré que la bactérie *S. enterica* souche 22A est devenue autotrophe pour le tryptophane et la *S. enterica* souche 2A autotrophe pour l'histidine, témoignant d'un transfert de gène à travers cette membrane. La transduction : transfert d'ADN bactérien par l'intermédiaire d'un vecteur viral, a ainsi été mise en évidence.

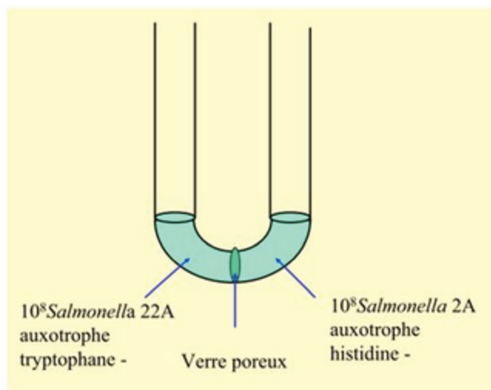


Figure 16: Illustration de l'expérience de Zinder et Lederberg

Il existe deux grand types de transduction : la transduction généralisée et la transduction spécialisée.

La transduction généralisée est une erreur d'encapsidation : ce n'est pas de l'ADN viral qui est encapsidé mais un fragment d'ADN bactérien chromosomique ou plasmidique. Le virion est alors capable de délivrer ce fragment à une autre bactérie hôte mais sans entraîner de cycle infectieux car il s'agit d'un virion défectif. L'ADN bactérien délivré peut alors s'intégrer dans la bactérie hôte par recombinaison homologue (Figure 17).

La transduction spécialisée concerne les phages tempérés uniquement. Lorsque le génome du prophage s'extrait du chromosome bactérien (excision), des erreurs peuvent survenir aboutissant à l'excision d'une molécule d'ADN « hybride » constituée d'un fragment d'ADN phagique et d'un fragment d'ADN bactérien. Le fragment d'ADN bactérien provient d'une zone adjacente à la zone d'intégration du prophage. La nouvelle particule virale, fonctionnelle, peut ainsi infecter une nouvelle cellule hôte, les gènes bactériens du virion seront alors intégrés dans le génome bactérien de la cellule hôte (Figure 17).

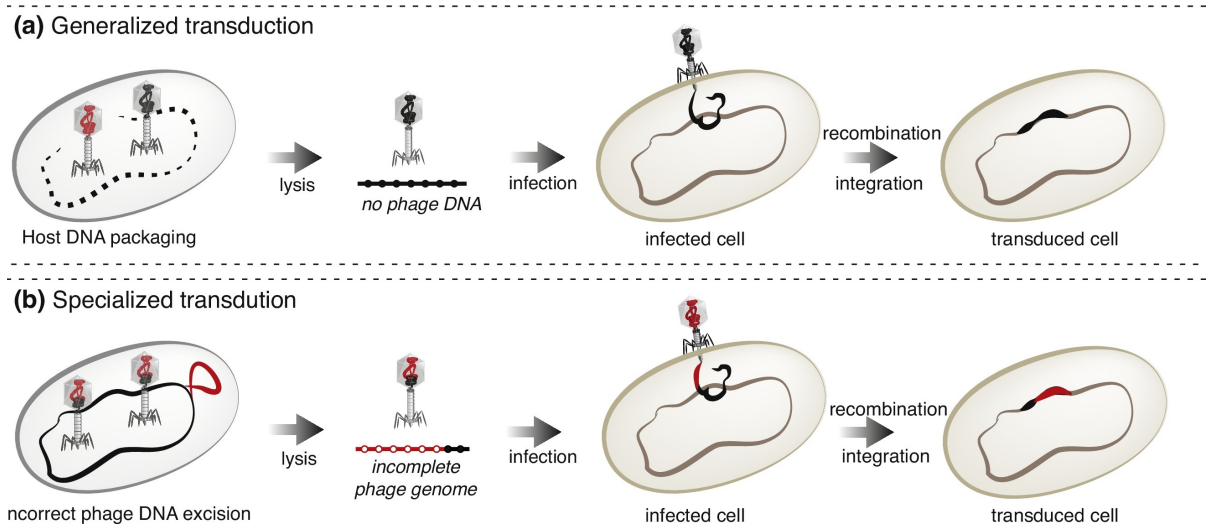


Figure 17: Schéma explicatif des deux types de transduction: généralisée et spécialisée. D'après ⁶⁸

Chez le *Staphylococcus aureus* il a été récemment décrit deux nouveaux modes de transductions : l'autotransduction et la transduction latérale, s'appuyant chacune respectivement sur la transduction généralisée et la transduction spécialisée :

L'autotransduction a été décrite chez *Staphylococcus aureus* ⁶⁹. Voici le concept détaillé dans la Figure 18 : la population lysogène de *S. aureus* libère spontanément des phages dans l'environnement, ces phages vont pouvoir infecter une des bactéries sensibles au phage (porteur d'un gène de résistance, par exemple). Au cours du cycle lytique, il y aura une production de certaines particules phagiques avec uniquement de l'ADN bactérien (transduction généralisée). Ces particules transduites vont pouvoir infecter la population lysogène de départ et intégrer l'ADN bactérien au sein de la bactérie alors que les autres particules phagiques ne pourront pas infecter et lyser la bactérie de départ car celle-ci est immunisée (elle porte déjà le prophage, pas de surinfection possible). L'étude montre l'acquisition chez *S. aureus* d'un gène de résistance aux antibiotiques par auto-transduction *in vitro* et *in vivo* (wax moth larvae).

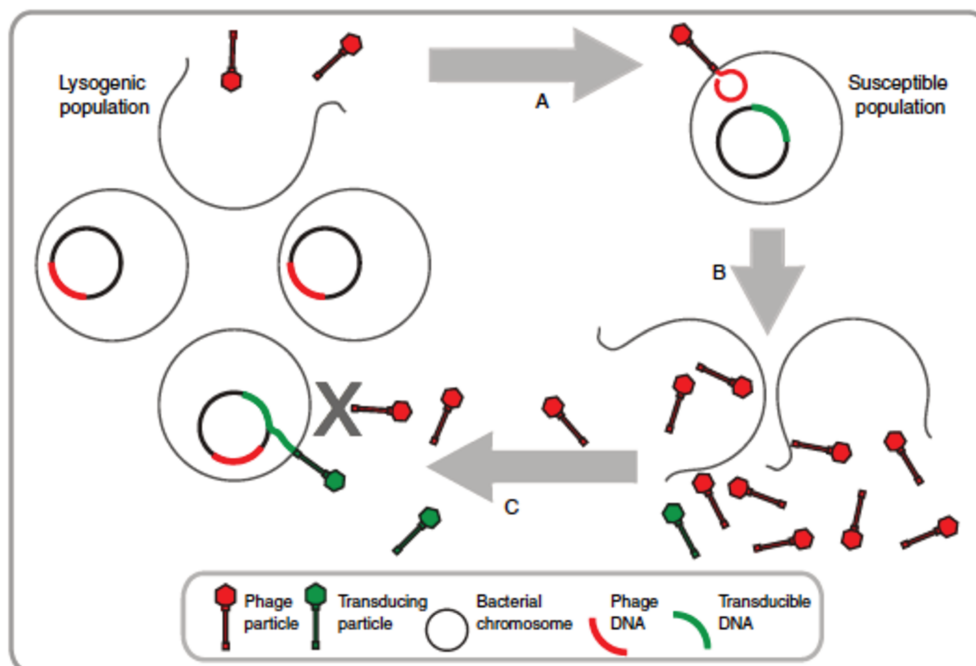


Figure 18: Présentation du modèle d'autotransduction d'après ⁶⁹

La transduction latérale

Une nouvelle étude publiée en 2018 introduit également la notion de « transduction latérale » chez *S. aureus* ⁷⁰. Elle met en évidence que certains prophages portés par *S. aureus* ne suivent pas la voie habituelle « excision-réplication-encapsidation ». Les prophages sont répliqués *in situ* dans le chromosome, au moment de l'encapsidation il y a de grandes portions d'ADN (plusieurs centaines de kilobases) bactérien adjacent au prophage qui peuvent être encapsidées. Certaines régions du chromosome de *S. aureus* proche du prophage sont donc choisies préférentiellement et deviennent alors des plateformes hypermobiles de transfert génétique avec la formation de très nombreuses particules transductrices.

1.3.5 Transduction et transfert génétique

La transduction permet ainsi des échanges génétiques entre les bactéries, contribuant à leur évolution. Cependant la spécificité du phage pour un hôte bactérien est une variable clé qui affecte la capacité de ces processus à disséminer les gènes dans les communautés. L'analyse des réseaux d'homologie génétique entre bactéries et phages (réseaux dits de transduction) suggère que la plupart des transferts ont lieu entre taxons étroitement apparentés⁷¹, en accord avec l'opinion traditionnelle selon

laquelle les phages ont une gamme d'hôtes étroite. Pourtant, ce point de vue est remis en question par la découverte de bactériophages capables de traverser l'espèce ou même la barrière des genres ^{72,73}, c'est à dire ayant des hôtes d'espèces voir même de genres bactérien différent.

Le transfert horizontal de gènes entre les bactéries via les phages (transduction) dépend de la capacité de la bactérie à se défendre contre l'invasion de phages (revue de ⁷⁵) et la capacité des phages à contourner les systèmes de défense bactérien (revue de ⁷⁶).

1.3.6 Les phages du microbiote intestinal

La communauté des bactériophages dans le microbiote intestinal est appelée phageome. Au cours des dernières années, de nombreuses études métagénomiques ont révélé que le phageome est une composante peu documentée de l'écosystème microbien intestinal. La grande majorité des séquences de phages nouvellement caractérisées n'a pas d'homologue dans les bases de données génétiques publiques. De plus, de nombreux phages détectés dans les études métagénomiques ne sont pas liés aux phages actuellement identifiés ou présentent une diversification considérable par rapport aux phages connus ²¹, ce qui rend l'analyse du phageome intestinal difficile.

Grâce à des observations microscopiques il a été observé entre 10^8 et 10^{12} particules phagiques par gramme de selles ⁷⁷. Les phages les plus abondants seraient des phages à ADN double brin de l'ordre des Caudovirales (famille des Myoviridae, Podoviridae et Siphoviridae) et des phages à ADN simple brin tels que les phages de la famille des Microviridae ⁷⁸. Ces familles de phages infectent divers taxons bactériens dans l'intestin tels que les Firmicutes, Bacteroidetes, Proteobacteria et Actinobacteria. En plus des phages à ADN, l'intestin peut aussi contenir des phages à ARN, qui semblent être des membres transitoires qui proviennent de sources alimentaires⁷⁹.

Plusieurs études mettent en évidence que dans la fraction identifiable des phages intestinaux présents dans les selles, ce sont les phages tempérés qui sont majoritairement présents ^{80,81}. Cette prédominance pourrait dépendre de la localisation de la flore intestinale. En effet, il a été observé des taux de croissance bactérienne plus élevés à la surface du biofilm intestinal et les taux de croissance de la bactérie sont corrélés au nombre de prophages retrouvés ^{62,82}. Ainsi, il existerait un gradient de

stratégies de réplication des phages, de lysogénique à lytique, le long de la profondeur du mucus intestinal⁸³.

Phageome individu spécifique

Ces dernières années, un nombre croissant d'études a été réalisé pour comparer les phageomes de différentes personnes. Ces études ont révélé qu'il existe une variabilité très élevée du virome bactérien intestinal chez les adultes ^{84,85}. Cette grande variabilité débute dès la naissance et s'accroît ensuite via les différents facteurs environnementaux (ex : alimentation). Malgré une grande diversité, un ensemble de phages communs est retrouvé chez de nombreux individus.

Phageome commun

L'analyse des CRISPRs (clustered regularly interspaced short palindromic repeats) dans le microbiome intestinal humain d'individus européens a démontré la similarité d'un grand nombre de phages entre plusieurs personnes. Un certain nombre de ces phages ont été détectés dans des échantillons fécaux prélevés sur des individus japonais et américains ⁸⁶. Récemment Manrique *et al.* ont introduit le terme de « healthy gut phageome » ⁸⁷ ; il a été rapporté un « core » phageome (ou phageome commun) de 23 phages présent chez plus de 50% des individus du monde entier. Et un phageome « commun » de 132 phages présents chez 20 à 50% des individus. Les auteurs appellent « healthy gut phageome » la combinaison des phages « core » et « commun ». Ces observations ont été faites chez des individus bien portants. Les phages composant le « healthy gut phageome » n'ont pas été retrouvés chez des patients atteints de maladies inflammatoires de l'intestin.

En 2014, une étude ⁸⁸ a mis en évidence un phage très fréquemment retrouvé dans les métagénomomes viraux. Ce phage a été retrouvé dans 22% de l'ensemble des données métagénomiques du microbiote intestinal disponibles dans les banques publiques et dans 90% de ces données en ne sélectionnant que les métagénomomes viraux du microbiote intestinal. Ce phage a été appelé le crAssphage (cross-assembly phage). La majorité des protéines codées par le crAssphage ne correspond à aucune séquence connue dans les bases de données. Le génome de ce phage fait 97 000 paires de base, il appartient à la famille des Podoviridae ⁸⁹ et son hôte probable est *Bacteroides intestinalis* ⁹⁰.

D'après l'étude de Lim *et al.*⁹¹, le crAssphage n'a été détecté que chez un seul enfant sur 8 à l'âge de 24 mois, suggérant que ce phage n'est pas acquis dans la petite enfance. Mais

récemment une nouvelle étude a montré que le crAssphage a été retrouvé chez l'ensemble des bébés nés par voies basses ou par césarienne ⁹², démontrant une acquisition très précoce de ce phage.

Dans l'étude de Norman *et al.*⁹³ en 2015 le crAssphage a été retrouvé chez 71% des individus sains et ce phage appartient au « core » phageome du healthy gut phageome de Manrique *et al.*⁸⁷.

Phageome intestinal à travers les âges

L'analyse du méconium (première excrétion fécale d'un nouveau-né) par microscopie à fluorescence n'a détecté aucune particule de phage ⁹⁴. Il a été démontré que l'intestin des nouveau-nés est rapidement colonisé par des bactéries provenant de la flore vaginale maternelle et de l'environnement. Étant donné que des phages apparaissent à chaque endroit où des bactéries sont présentes, la colonisation bactérienne de l'intestin du nourrisson est suivie d'une colonisation par des phages. L'évolution du phageome dans l'intestin du nourrisson est rapide : une semaine après la naissance, 10^8 VLP de type viral (dont la majorité sont des phages) ont été détectées dans chaque gramme de fèces. De plus, on observe un renouvellement des principaux phages au cours de la deuxième semaine suivant la naissance ⁹⁴. Une autre étude ⁸⁰, s'intéressant aux jumeaux à la naissance, a comparé le microbiote d'une mère et de ses jumeaux. Elle a révélé une homogénéité bactérienne au sein des microbiotes intestinaux des vrais jumeaux mais une grande diversité virale. Les phageomes de jumeaux seraient quand même plus proches que les phageomes de deux bébés sans lien génétique ⁹¹. Pendant l'enfance, la proportion relative de phages appartenant à la famille Microviridae augmente tandis que les phages Caudovirales diminuent ⁹¹.

Avec le temps et la maturation de l'intestin, des changements importants se produisent dans la composition du phageome qui semblent se stabiliser à l'âge adulte. Une étude a suivi pendant 2 ans et demi l'évolution du phageome du microbiote intestinal chez l'adulte et a montré que la population des phages tempérés est très abondante et stable alors que la population des phages lytiques (ex : Microviridae) présente une plus grande diversité au cours du temps ²¹. Des facteurs extérieurs comme l'alimentation peuvent influencer sur la composition du phageome. Minot *et al.*, en 2011 ont mis en évidence une convergence des communautés virales intestinales en fonction du régime alimentaire ⁸⁴. De même, chez la souris subissant un régime riche en graisses et sucres, le phageome intestinal s'est modifié et enrichi en Caudovirales ⁹⁵.

Phageome, marqueur potentiel de santé

Il existe des rapports dans la littérature qui décrivent des changements dans la structure du phageome intestinal dans différents états pathologiques. Par exemple une étude a exploré le phageome chez des patients atteints de la maladie de Crohn (MC) ou de rectocolite hémorragique (RCH), observant un enrichissement des phages de l'ordre des Caudovirales dans ces populations ⁹³. Une autre étude récente (2018) a montré une différence entre les phageomes intestinaux d'adultes sains et d'adultes atteints de diabète de type 2 : il a été observé une augmentation significative du nombre de phages intestinaux dans le groupe des patients diabétiques avec la mise en évidence de 7 phages spécifiques à ce groupe ⁹⁶.

Cela a conduit à suggérer que le phageome pourrait représenter un marqueur indirect de la santé ou de la maladie. Cette idée est également soutenue par le fait que les phages peuvent moduler le microbiome intestinal et ont des connexions étendues avec le système immunitaire de l'hôte, principalement en traversant la muqueuse intestinale et en se rendant dans les ganglions lymphatiques locaux et les organes ⁹⁷.

1.4 Phages et gènes de résistance aux antibiotiques

1.4.1 Phages et gènes de résistance aux antibiotiques : observations

Des gènes de résistances aux antibiotiques ont été retrouvés chez des prophages intégrés à des bactéries ou dans des particules phagiques.

Dans des bactéries sous forme de prophages

Dans la littérature, on retrouve de rares exemples de gènes de résistance aux antibiotiques identifiés dans des prophages insérés dans des bactéries. En voici plusieurs exemples : la bactérie *Salmonella enterica* a été décrite avec un prophage P22 dans son génome portant une beta-lactamase (*TEM-1*)⁹⁸, un prophage avec un gène de résistance à la tétracycline a été décrit chez Enterocoque⁹⁹, il a été retrouvé chez *Bacillus anthracis*, un prophage avec un gène conférant une résistance à la fosfomycine¹⁰⁰. De plus, en 2014, il a été décrit pour la première fois un phage P1-like porteur de Beta-Lactamase à Spectre élargi (BLSE de type *SHV-2*) dans un *E. coli* isolé lors d'une infection urinaire¹⁰¹. Le phage P1-like lysogénise la bactérie sous forme d'un élément plasmidique non intégré dans le génome de la bactérie hôte. Ce prophage est défectif et n'a pas pu être induit pour la formation de nouvelle particule virale, ainsi la lysogénisation du gène de la BLSE n'a pas pu être observée.

Dans les particules phagiques

Dans la littérature, de nombreux exemples de gènes de résistance aux antibiotiques retrouvés dans des particules phagiques isolées sont décrits. Dans ces différentes études, la fraction phagique du milieu étudié a d'abord été isolée, l'ADN contenu dans les capsides extraits, puis les gènes de résistance aux antibiotiques ont pu être mis en évidence. Plusieurs techniques ont été utilisées pour détecter les gènes de résistances (i) soit par PCR ciblée et quantitative (qPCR) sur certains gènes (ex : beta-lactamases tel que *blaTEM* ou *blaCTXM*, *mecA*) (ii) soit par analyses métagénomiques. Ainsi, il a été retrouvé dans des phages isolés des eaux usées des gènes de résistance aux antibiotiques soit par qPCR^{102,103} soit par métagénomique¹⁰⁴. Chez l'homme également, au sein des microbiotes intestinaux^{84,105,106} et pulmonaires¹⁰⁷, il a été retrouvé des gènes de résistance aux antibiotiques dans la fraction phagique. Ces dernières études

métagénomiques sont controversées car il pourrait s'agir de contamination bactérienne et non de contenu phagique. Ce point sera repris de façon plus détaillée dans la suite de l'introduction.

Ces gènes de résistance retrouvés dans les phages proviennent très probablement des bactéries, mais a-t-on déjà formellement observé des transferts de gènes de résistance aux antibiotiques entre les bactéries et les phages ?

1.4.2 Transfert de gènes de résistance aux antibiotiques entre bactéries et phages

La transduction, mécanisme décrit précédemment, permet le transfert de matériel génétique entre deux bactéries via un phage. Le matériel génétique en question peut être un gène de résistance aux antibiotiques. Ce phénomène a été mis en évidence dès 1986¹⁰⁸ avec le transfert du gène de la résistance à la méticilline entre deux staphylocoques (*S. epidermidis*) après des irradiations UV stimulant la fréquence de transduction. En 1997, trois staphylocoques dorés lysogéniques résistants à la tétracycline ont été isolés chez des bovins au Brésil et le transfert de la résistance via le phage à d'autres souches de *S. aureus* initialement sensibles a pu être observé en culture¹⁰⁹.

D'autres approches mettent en évidence indirectement des phénomènes de transduction de gènes de résistance aux antibiotiques. En 2014, Colomer-Lluch *et al.*¹⁰⁶ ont recherché des gènes de résistance aux antibiotiques dans les fèces et spécifiquement dans la fraction phagique. La quantité de gènes de résistance détecté dans la fraction phagique s'est avérée plus faible que dans l'ensemble des fèces mais elle suit la même fluctuation au cours du temps. Cela impliquerait une étroite corrélation entre les gènes de résistance retrouvés chez les bactéries et chez les phages d'un même environnement. Mais la question de la contamination bactérienne est toujours au cœur du débat ; si la fraction phagique est contaminée par de l'ADN bactérien, le résultat de l'étude n'est alors plus une preuve de transfert de gène de résistance entre le monde bactérien et les phages.

Chez *Staphylococcus aureus* il a été récemment décrit deux mécanismes de transduction : l'autotransduction et la transduction latérale, déjà introduit et définis dans ce travail. Le travail d'Haaber *et al.*⁶⁹ montre l'acquisition chez *S. aureus* d'un gène de résistance aux antibiotiques par auto-transduction *in vitro* et *in vivo* et le travail de

Chen *et al.* ⁷⁰ met en avant la formation d'un grand nombre de virions avec des gènes bactériens de résistance aux antibiotiques lors du processus de transduction latérale. Ces deux nouvelles découvertes sont donc en faveur du transfert de gènes de résistance entre deux bactéries hôtes du même phage.

Afin de savoir si la transmission de gènes de résistance aux antibiotiques entre des bactéries via la transduction est un phénomène important, il faut en estimer la fréquence. La question est complexe car il faudrait évaluer les différents types de transduction pour chaque couple bactérie/phage et dans différentes conditions (*in vitro* ou *in vivo*).

Pour la transduction spécialisée, l'emballage « par erreur » d'ADN bactérien au moment de l'excision du prophage dépend du contexte génétique dans lequel est intégré le prophage. Par exemple la transduction spécialisée est favorisée si le prophage est intégré dans une zone de séquences répétées ¹¹⁰. La transduction spécialisée a lieu à très basse fréquence, elle a été estimée à 10^{-6} par rapport à une excision normale pour le phage lambda. *In vivo*, dans un environnement complexe comme le microbiote intestinal de la souris, il a été montré pour l'espèce *E. coli* et le phage lambda que le nombre de prophages recombinant avec un gène de résistance aux antibiotiques était de 5.10^{-8} parmi les nouveaux phages lysogènes ¹¹¹. De plus, le volume de la capsid du virion ne permet pas d'intégrer beaucoup plus d'ADN que celui du phage prévu. Pour le phage lambda il est décrit une capacité d'encapsidation de 108% ¹¹² et des valeurs similaires sont trouvées pour d'autres phages ¹¹³. Cela n'est pas en faveur d'une place majeure de ce type de transduction dans le transfert de gène bactérien. Le rôle de la transduction spécialisée dans l'évolution bactérienne et plus précisément dans le transfert de gène de résistance est difficile à quantifier et la littérature est pauvre à ce sujet.

Concernant la transduction généralisée, des études ont montré que beaucoup de phages étaient capables de transduire de l'ADN bactérien. Par exemple, 99% des prophages de salmonelle sont capables de faire ce type de transduction ¹¹⁴. Les taux de transduction généralisée estimés *in vitro* sont plutôt faibles. Par exemple, une étude *in vitro* chez *S. aureus* USA 300 a montré un taux de transduction généralisée avec des particules phagiques contenant un gène de résistance aux antibiotiques de 1/1700 particules phagiques ¹¹⁵. Mais un article récent a mis en évidence un taux de transduction de 1% par unité formant colonie lorsque des microbiotes naturels ont été

utilisés comme bactéries hôtes ¹¹⁶. L'effet génétique de l'environnement peut donc être un élément majeur dans la fréquence de transduction.

1.4.3 Phages : réservoir de gènes de résistance aux antibiotiques au sein du microbiote intestinal

Le résistome intestinal est l'ensemble des gènes de résistance aux antibiotiques présents dans le microbiote intestinal. Le résistome des bactéries aérobies pathogènes est bien mieux connu que celui des bactéries anaérobies en raison de la facilité à les cultiver et de leur intérêt médical.

La question de la place des phages dans la dynamique de ce résistome est très intéressante. Leur rôle dans le transfert de l'antibiorésistance dans les milieux naturels est beaucoup moins étudié et n'est apparu dans la littérature que récemment.

Cette section est une revue de la littérature des articles où il est mis en évidence un rôle des phages dans la diffusion de la résistance chez les animaux, chez les humains sans pression de sélection antibiotique, chez les humains sains sous antibiotiques et chez les patients hospitalisés. Les conclusions de ces articles seront ensuite remises en question dans la section suivante.

Etudes dans le microbiote intestinal des animaux

En 2013, Modi *et al.*, ont publié une étude sur l'impact des antibiotiques sur les phages présents dans le microbiote intestinal de souris ¹⁰⁵. Cette étude est d'une importance majeure et relance cette question passionnante. Les auteurs ont montré qu'un traitement antibiotique (béta-lactamine et quinolone) conduit à l'augmentation du nombre de gènes de résistance codés par les phages, conférant ainsi une résistance à l'antibiotique administré mais aussi à d'autres antibiotiques.

En 2018, Wang *et al.* ont analysé le microbiote intestinal de cochons de fermes commerciales en Chine ¹¹⁷. Les gènes de résistance aux antibiotiques ont été recherchés par qPCR dans les bactéries et dans la fraction phagique. Dans les phages, l'ensemble des gènes de résistance recherchés (12 gènes) sont retrouvés chez 35,5% des échantillons et les gènes suivant sont retrouvés chez 100% d'entre eux : gène de résistance aux sulfamides (*sul1*), gène de résistance aux beta-lactamines (*blaTEM*) et gène de résistance aux macrolides (*ermB*). La corrélation entre la quantité de gènes de résistance dans la fraction phagique et dans la fraction bactérienne montre un fort ratio de transférabilité du gène *qnrA* conférant la résistance aux quinolones. A noter que, pour la plupart des

gènes de résistance, il n'y a pas de corrélation directe entre les deux entités. Cela met en évidence un rôle potentiel des phages dans la dissémination des gènes de résistance aux antibiotiques entre les bactéries.

Etude dans le microbiote intestinal de l'homme sans pression de sélection antibiotique

L'étude de 254 métagénomés intestinaux humains, chez des individus en bonne santé a mis en évidence la présence de prophages dans les bactéries ⁸¹. Certains de ces prophages contiennent des gènes de résistance aux antibiotiques. La famille des gènes de résistance à la streptomycine est particulièrement représentée au sein de ces prophages alors que les gènes de résistance à l'érythromycine sont très peu présents.

En 2018, Brown-Jaque *et al.*¹¹⁸ ont analysé 150 échantillons de selles humaines provenant d'individus sains qui n'avaient pas reçu d'antibiotique et qui n'avaient pas voyagé dans les trois mois précédant le recueil. La fraction phagique a été isolée et l'ADN extrait. Neuf gènes de résistance aux antibiotiques ont été recherchés par qPCR. Soixante-dix pourcents des échantillons contenaient au moins un gène de résistance. Les gènes de résistance les plus prévalent étaient : *blaTEM* et *blaCTX-M-9* (gènes de résistance aux beta-lactamines), *sul1* (gène de résistance aux sulfamides) et *qnrA* (gène de résistance aux quinolones). Une sélection de *E. coli* et *K. pneumoniae* porteurs d'un gène de résistance aux antibiotiques a été réalisée. L'induction des potentiels prophages a été réalisée grâce à la mytomycine C. Les phages du surnageant ont été extraits avant et après induction et il a été montré une augmentation du nombre de copies des gènes de résistance dans 62% des échantillons. Ce surnageant a été observé au microscope électronique et des virions ont été identifiés. L'hypothèse de la transduction généralisée est mise en avant dans la discussion de ce papier.

Etude du microbiote intestinal humain chez des patients hospitalisés recevant des antibiotiques

En 2015, Abeles *et al.* analysent le phageome fécal de 4 patients ayant reçu une antibiothérapie par voie intraveineuse pendant 6 semaines (chaque patient a reçu un traitement différent) à trois temps différents (3^{ème} jour, deuxième semaine et sixième semaine) versus 5 sujets contrôles ¹¹⁹. Il est montré que le phageome intestinal des patients sous antibiotiques présente la même diversité que les sujets non traités mais il est observé une plus grande quantité de contigs porteurs d'un gène de résistance aux

antibiotiques (famille des beta-lactamines, vancomycine, macrolides, tétracyclines) dans le groupe des patients traités. Ce résultat n'est cependant pas statistiquement significatif.

L'émergence d'une résistance aux antibiotiques chez les patients avec une antibiothérapie au long cours suggère que les phages jouent un rôle important dans la résilience des communautés microbiennes humaines face aux antibiotiques.

Etude dans le microbiote intestinal de l'homme sain recevant des antibiotiques

Afin de limiter le nombre de biais dans l'étude de l'implication des phages du microbiote intestinal dans la dissémination de gènes de résistance aux antibiotiques, il est préférable de travailler sur des cohortes de volontaires sains soumis à un traitement antibiotique plutôt que sur des patients hospitalisés avec déjà de lourds antécédents et un microbiote intestinal altéré.

En 2018, l'équipe de Gorska *et al.* a analysé le microbiome et le phageome intestinal de deux individus en bonne santé avant et après un traitement de 6 jours à la ciprofloxacine (fluoroquinolone)¹²⁰. Il est intéressant de voir que sur les mêmes échantillons de selles, les résultats de l'étude du phageome et les résultats de l'étude sur les phages retrouvés dans le microbiome ne sont pas corrélés. Les phages identifiés dans le microbiome ne sont pas systématiquement retrouvés dans les données de phageomes. Cela met en valeur la difficulté d'analyse de ce type de données. Chez les deux volontaires, il est observé une augmentation de l'abondance des contigs contenant un gène de résistance au sein du phageome 28 jours après la prise d'antibiotique.

En 2019, Fernandez-Orth *et al.* ont étudié le phageome intestinal d'un seul volontaire sain avant et après un traitement antibiotique de 5 jours par ciprofloxacine (fluoroquinolone)¹²¹. Dans les prélèvements de selles, les gènes de résistance aux antibiotiques ont été détectés à la fois bio-informatiquement, grâce au séquençage du phageome, mais aussi par qPCR ciblée, directement sur l'ADN phagique. Les résultats obtenus par ces deux techniques différentes ne coïncident pas toujours, mettant en évidence que la métagénomique ne détecte que les gènes de résistance aux antibiotiques les plus abondants. Il est intéressant de noter qu'aucun gène de résistance aux antibiotiques n'a été détecté (ni par qPCR ni par métagénomique) dans l'échantillon avant la prise d'antibiotique alors que des gènes de résistance aux antibiotiques

(notamment aux quinolones) ont été mis en évidence dans l'échantillon qui suit la prise d'antibiotique. Un échantillon prélevé un an après la prise d'antibiotique n'a pas permis de retrouver de gène de résistance.

Les deux études décrites ci-dessus témoignent de l'intérêt d'analyser les gènes de résistance aux antibiotiques dans les phages avant et après la prise d'un traitement antibiotique. Cependant, elles ont très peu de puissance car il y a très peu d'individus étudiés, cela ne permet donc pas de mettre en évidence un véritable effet des phages dans le transfert de gènes de résistance sous pression antibiotique.

1.4.4 Analyse bio-informatique des gènes de résistance : bases de données disponibles et méthodologies utilisées dans la littérature

Base de données de gènes de résistance aux antibiotiques

Il y a actuellement plusieurs bases de données de gènes de résistance aux antibiotiques disponibles, ces informations sont résumées dans le Tableau 1

Tableau 1 : Présentation des différentes bases de données de gènes de résistance

Nom de la base de données, référence, date (dernier update)	Description	Avantages	Limites
CARD ¹²² , 2013 (dernier update en 2017 puis revu tous les mois)	2402 gènes (dont 1440 beta-lactamases). Tous les gènes de ARDB sont dans CARD	Bonne spécificité	Comprend surtout des gènes de résistance de bactéries pathogènes
Resfam ¹²³ , 2015 (dernier update en mars 2017)	173 profils HMM construits en utilisant les bases de données suivantes : CARD, LacED, Jacoby and Bush's collection, PFAM, TIGRFAMs	Bonne sensibilité	Comprend surtout des gènes de résistance de bactéries pathogènes
ARG-ANNOT ¹²⁴ , 2013 (dernier update : mai)	2024 gènes (dont 1300 beta-lactamases) Fait à partir de plusieurs bases de données existantes : ARGO,	Bonne spécificité	Comprend surtout des gènes de résistance

2018)	MvirDB, ARDB, and Resfinder		de bactéries pathogènes
Resfinder ¹²⁵ , 2012 (dernier update : 2018)	3100 gènes (dont 2000 beta- lactamases) Fait à partir de plusieurs bases de données existantes : ARDB, https://faculty.washington.edu/marilynr/ , http://www.lahey.org/Studies/	Bonne spécificité	Comprend surtout des gènes de résistance de bactéries pathogènes
ResfinderFG (Disponible sur le site de Center of Génomique Epidémiologie) ¹²⁶ , 2018	2 282 gènes de résistance identifiés grâce à 4 études de métagénomique fonctionnelle (microbiote intestinal, eaux usées, sol)	Gènes de résistance retrouvés dans des données de métagénomiques	Très récente, encore peu de recul
Mustard ¹²⁷ , 2019	6 095 gènes de résistance provenant du microbiote intestinal et prédit par modélisation comparative	Base spécifique du microbiote intestinal	Très récente, encore peu de recul
FARME ¹²⁸ , 2017	Fait à partir de données fonctionnelles de métagénomiques environnementales (20 études). Inclus des éléments génétiques mobiles 2 bases de données: - FARME DNA : 11 057 gènes de résistance. -FARME Protein : 26 253 séquences protéiques	Genès de résistance retrouvés dans des données de métagénomiques environnementales	Très récente, encore peu de recul.

Les nouvelles bases de données de gènes de résistance publiées récemment (ResfinderFG, Mustard et FARME) sont très intéressantes car elles sont riches en gènes de résistance présents dans l'environnement ou dans le microbiote intestinal. Cela diffère des bases de données classiques qui contiennent majoritairement des gènes de résistance présents chez les bactéries pathogènes. Elles n'ont pas encore été testées sur les données métagénomiques de phages mais cela sera incontournable pour les études à venir.

Méthodologie bio-informatique pour la recherche de gènes de résistance aux antibiotiques.

Dans les différentes études analysant le phageome intestinal, les stratégies bio-informatiques de recherche de gènes de résistance aux antibiotiques sont très hétérogènes. Or un récent papier ¹²⁹ a montré que le choix des outils et des seuils était primordial afin de ne pas faire de la sur-prédiction (fausse prédiction), ce point sera détaillé dans le prochain paragraphe.

Pour illustrer cela, voici une revue de la littérature des différentes approches bio-informatiques pour les recherches de gènes de résistance aux antibiotiques dans les données de phageome intestinal (Tableau 2).

Tableau 2: Revue de la littérature des différentes approches bio-informatiques pour rechercher des gènes de résistance aux antibiotiques dans des données métagénomiques de phageome intestinal.

Références	Support	Base de données utilisée	Cut-off choisi pour valider la présence d'un gène de résistance
84	Contig (protéine)	ARDB	BLASTp – e-value : 10^{-5}
105	Reads (nucléique)	ARDB + Uniprot avec les mots clé 'antibiotic' et 'breakdown'	BLASTx – e-value : 10^{-3}
119	Contig (nucléique)	CARD	BLASTx – e-value : 10^{-30}
130	Contig (protéine)	ARDB	BLASTp – e-value : 10^{-3}
81	Contig (protéine)	ARDB	BLASTp – identité à 95% et couverture 95%
121	Contig (nucléique)	Resfinder	BLASTn - identité à 90% et couverture

			60%
120	Reads (nucléique)	CARD	Diamond – couverture 80%
131	Non précisé	MG-RAST et ARDB	Non précisé
132	Contig (protéine)	Resfam	Non précisé
129	Contig (protéine)	ARDB Resfam	Critère exploratoire : BLASTp – bit score > 70 Critère conservateur : BLASTp - identité à 80% et couverture 40% / Resfam (-- cut-ga)

1.4.5 Controverse sur le transfert de gènes de résistance par les phages

Dans la majorité des études publiées, les gènes de résistance aux antibiotiques sont recherchés dans les métagénomés de phages avec des outils bio-informatiques. Or il y a une très grande hétérogénéité dans l'utilisation des stratégies bio-informatiques et des seuils choisis pour qualifier la présence d'un gène de résistance (Tableau 2). L'hypothèse de l'implication des phages dans la transmission de gènes de résistance aux antibiotiques a été remise en question par plusieurs papiers et principalement par celui de Enault *et al.* publié en 2017 ¹²⁹. En travaillant sur une banque de phages connus (les phages de RefSeq), dans lesquels il n'y a que deux gènes de résistance aux antibiotiques décrits, ils mettent en évidence l'importance du choix des seuils bio-informatiques dans la détection des gènes de résistance aux antibiotiques. Ainsi, avec des seuils trop exploratoires, la fréquence de faux positifs est très élevée (421 gènes de résistance prédits dans les phages de RefSeq versus 2 initialement décrits) (Figure 19). Les données du phageome intestinal (de l'homme ou de la souris) publiquement disponibles ont été ré-analysées (sur les reads assemblés en contigs) pour la recherche de gènes de

résistance aux antibiotiques mais avec des seuils bio-informatiques restrictifs. Ainsi sur les 465 contigs du microbiote de souris de l'étude de ¹⁰⁵, aucun gène de résistance aux antibiotiques n'a été retrouvé. Cette étude suggère qu'il y a eu dans les études publiées, une surestimation du nombre de gènes de résistance aux antibiotiques amenant donc à des conclusions possiblement erronées sur le rôle des phages dans le transfert de gène de résistance aux antibiotiques par lysogénie. Ce travail préconise donc de rechercher des gènes de résistance avec des seuils restrictifs afin de minimiser les faux positifs.

Un autre problème soulevé dans l'analyse de données métagénomiques de phage est la présence d'ADN bactérien. Il est alors impossible de savoir si l'ADN bactérien présent provient d'une transduction généralisée (ADN bactérien dans les capsides de phage) ou de réelle contamination bactérienne lors du processus d'extraction. Il convient alors d'augmenter la sensibilité de détection de cet ADN bactérien en recherchant des marqueurs spécifiques du monde bactérien sur les contigs assemblés, ce qui est une méthode plus sensible que la recherche d'ADN 16S sur les reads. Lors de prélèvements longitudinaux il faut alors surveiller le taux d'ADN bactérien présent afin d'essayer de comprendre si celui vient d'une contamination ou de la transduction généralisée. Dans la discussion du papier de Enault *et al.* ¹²⁹, l'implication de la transduction spécialisée dans le transfert de gènes de résistance est drastiquement minimisée (pas de gènes de résistance dans les contigs de phage) tandis que celle de la transduction généralisée est difficile à apprécier. Il est alors difficile d'analyser le rôle de cette dernière, lorsqu'on retrouve les gènes de résistance aux antibiotiques sur des contigs bactériens comment savoir s'il s'agit d'une contamination ou d'une transduction généralisée ?

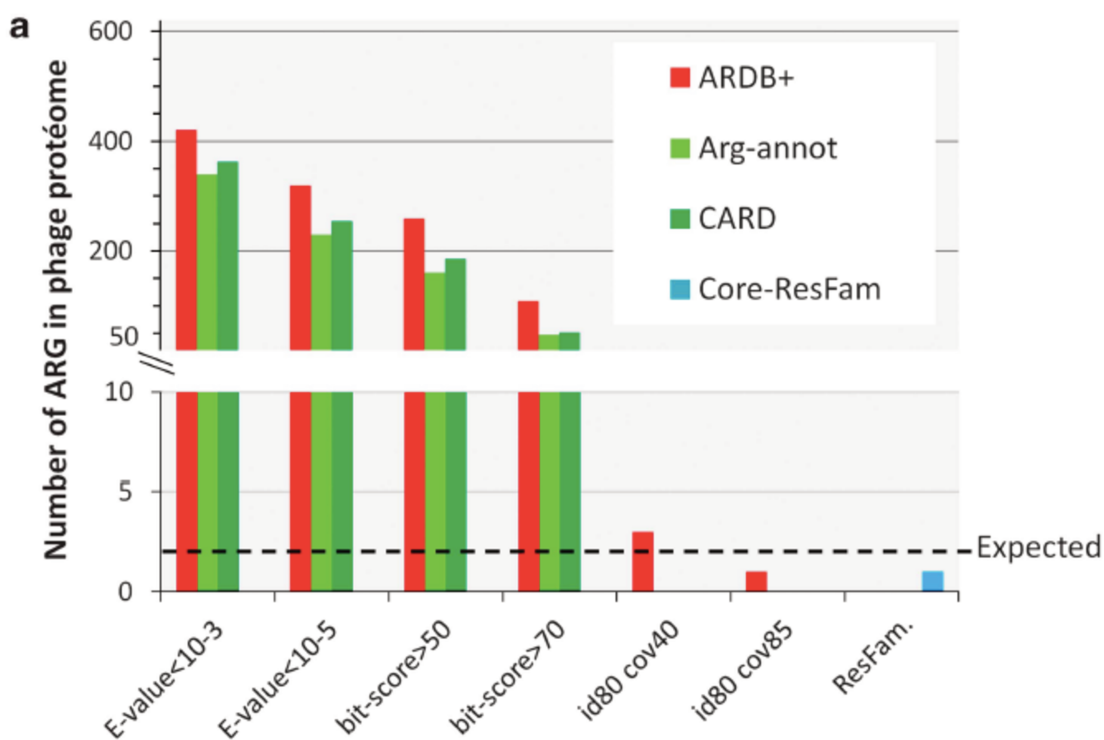


Figure 19: Analyse du protéome de phages connus pour la recherche de gènes de résistance aux antibiotiques avec différentes bases de données et différents seuils bio-informatiques. La ligne en pointillée indique le nombre de gènes de résistances attendus (n=2). D'après ¹³³

D'autres études n'ont pas mis en évidence d'implication des phages dans le transfert de gène de résistance. Deux études sur des porcs traités par antibiotiques n'ont pas retrouvé plus de gènes de résistance aux antibiotiques dans la fraction phagique post-traitement. Cependant, Allen *et al.* ¹³¹ ont montré que chez des porcs traités par cycline et pénicilline, l'abondance des phages codant une intégrase avait augmenté. Cela est probablement dû à une induction des prophages par le traitement antibiotique. En revanche, il n'a pas été retrouvé plus de gènes de résistance aux antibiotiques dans la fraction phagique post traitement. Cette même observation a été faite dans l'étude de Johnson *et al.* chez des porcs également traités par des cyclines ¹³².

1.5 Phage et microbiote intestinal : procédure d'extraction et d'analyse

Etant donnée l'hétérogénéité des procédures d'extraction et d'analyse des phages dans le microbiote intestinal, il est nécessaire de présenter ici des tableaux qui comparent et récapitulent ces différentes approches.

1.5.1 Isolement

Voici les grandes étapes, présentées et discutées, de la procédure d'isolement des phages de selles :

- Collection de selles et conditions de stockage

Le travail d'isolement des phages commence par une étape de collecte des selles suivie d'une étape de stockage. Ces toutes premières étapes sont déjà primordiales pour garantir une bonne analyse des échantillons.

Une étude récente ¹³⁴ s'est intéressée aux variations de stockage (température, nombre de cycle de congélation-décongélation, variabilité inter-opérateur) de la selle et aux conséquences que cela peut avoir sur les phages fécaux étant donné que des rapports récents avaient montré une influence des conditions de collecte et de stockage sur la population bactérienne ¹³⁵. La beta diversité (distance de Spearman) d'un même échantillon de selle traité dans différentes conditions a été étudié. Le phageome et le bactériome présentent tous deux des variations modérées lorsqu'ils sont conservés à + 4 °C ou à température ambiante. Les profils des phageomes sont plus stables face aux multiples congélations-décongélation que les profils des bactériomes. Concernant la variabilité de l'effet « opérateur » (personne qui effectue les techniques) le phageome y est très sensible et cela s'explique aisément car il y a de multiples étapes manuelles dans la procédure d'isolement des phages et de leur ADN.

- Quantité de selles prises pour l'extraction

Selon les études, il y a une très grande hétérogénéité dans la quantité de selles prise pour procéder à l'extraction, allant de quelques milligrammes à plusieurs grammes (Tableau 3). Le principal avantage de prendre une grande quantité de selles est de retrouver potentiellement des phages très peu représentés mais cela complique l'étape de filtration à cause des nombreux débris et du mucus.

- Homogénéisation sur prélèvement dans un tampon approprié

La selle est ensuite homogénéisée dans un tampon tel que le phosphate buffered saline (PBS) ou le tampon SM (NaCl 100 mM, MgSO₄.7H₂O 8 mM, Tris-Cl 50 mM) (Tableau 3). Ces deux tampons semblent appropriés pour cette étape.

- Centrifugation

Après l'homogénéisation, l'étape de centrifugation permet d'éliminer les débris. Il peut y avoir une ou plusieurs centrifugations avec des vitesses très différentes mais préférentiellement à basse vitesse (Tableau 3). L'étude de Conceição-Neto *et al.* ¹³⁶ a comparé plusieurs vitesses de centrifugation et préconise une centrifugation de 3 min à 17000 g même si il y a une perte de virus géants comme les Mimivirus.

- Filtration afin d'éliminer les cellules humaines et bactériennes

Suite à la centrifugation, le surnageant est récupéré et une étape de filtration est alors effectuée pour éliminer les cellules humaines et bactériennes. Selon les études, plusieurs tailles de filtres peuvent être utilisées : 0,8, 0,45 ou 0,22 μm dans le but d'éliminer les cellules humaines et bactériennes tout en conservant les particules virales plus petites (Tableau 3). Or une étude ⁷⁷ a montré que l'utilisation du filtre 0,22 μm réduisait drastiquement (de moitié) le nombre de particules virales observées au microscope. Une autre étude ¹³⁶ a montré que les filtres de 0,45 μm et de 0,22 μm éliminent les gros virus (Mimivirus) et donc les excluent de l'analyse. L'utilisation du filtre 0,8 μm uniquement permet de ne pas filtrer les gros virus mais cela entraîne une plus grande contamination par les cellules bactériennes.

- Concentration des particules virales

Après l'étape de filtration, l'étape de concentration des particules virales est effectuée. Cette étape est essentielle dans le processus et plusieurs techniques de concentration sont décrites dans la littérature : précipitation au polyéthylène glycol (PEG), ultrafiltration en flux tangentiel, ultracentrifugation simple et ultracentrifugation avec un gradient de chlorure de Césium (ClCs). Notons que dans certains articles, il n'y a pas étape de concentration (Tableau 3).

La technique historique de référence est l'ultracentrifugation avec un gradient ClCs. Cependant, cette technique est très laborieuse, prend beaucoup de temps et une étude montre qu'elle est peu reproductible et que l'on perd beaucoup en diversité de virus extraits du fait de la perte de virus aux densités atypiques ¹³⁷.

Deux articles ont récemment comparé différentes techniques d'isolation des phages. Dans le premier article, ¹³⁸ toutes les techniques comparées utilisent une étape

d'ultracentrifugation au CICs qui est donc complexe, longue et non recommandée. Le deuxième article¹³⁷ compare 5 méthodes différentes mais malheureusement l'étape avec une concentration au PEG a échoué (dû à la formation d'une masse visqueuse) et elle n'a pas pu être comparé aux autres.

- Utilisation du chloroforme

Le chloroforme, qui perturbe les membranes lipidiques, est souvent utilisé pour enlever les bactéries résiduelles ou les vésicules. Cependant l'étude de Conceição-Neto *et al.* ¹³⁶ montre que cette étape entraîne la perte de certains virus sensibles au chloroforme et donc ne préconise pas l'utilisation de celui-ci. Le Tableau 3 résume l'utilisation ou non du chloroforme selon les études.

- Elimination de l'ADN/ARN résiduel avec une DNase/RNase

L'étape d'élimination de l'ADN/ARN résiduel avec une DNase/RNase est assez consensuelle et très souvent effectuée. Le type de DNase, la concentration, le temps d'exposition et la température peuvent varier (Tableau 3).

- Extraction de l'ADN.

L'étape finale d'extraction de l'ADN des particules phagiques est réalisée. Le protocole d'extraction de l'ADN varie en fonction des études et peut être soit manuel (ex : extraction manuelle au phénol chloroforme ou cetyltriméthylammonium bromide) soit avec un kit commercial d'extraction (ex : Phage DNA extraction (Norgen biotek Corp, Ontario, Canada), QIAamp Viral RNA Mini Kit (Qiagen), DNase easyblood and Tissue Kit (qiagen), Viral DNA/RNA extraction kit (Invitrogen) (Tableau 3).

Tableau 3: Vue d'ensemble des différentes méthodologies d'isolement des phages dans les selles

Référence du papier	Matériel	Quantité / Tampon	Centrifugation	Filtration	Concentration	Chloroforme	Dnase /Rnase	Extraction de l'ADN
80	Selles humaines	2-5 g dans du SM	2 500g pour 10 min	0.45 µm + 0.22 µm	Ultracentrifugation au ClCs	Oui	Oui, 2.5 UI/mL	Extraction manuelle au phénol-chloroforme
78	Selles humaines	100-300 mg dans du SM	2 500g pour 10 min	0.45 µm + 0.22 µm	Non	Oui	Oui 3U	Extraction manuelle au phénol-chloroforme
91	Selles humaines	200 mg dans du PBS	Non	0.45 µm	Non	Non	Non	Extrait avec le Cobas Ampliprep Instrument (roche)
84	Selles humaines	500 mg dans du SM	4 700g pour 30 min	0.22 µm	Ultracentrifugation au ClCs	Oui	Oui	DNase easyblood and Tissue Kit (qiagen)
21	Selles humaines	1 gr dans du SM	no	0.22 µm	Ultrafiltration (Millipore Centricon)	Oui	Oui	DNase easyblood and Tissue Kit (qiagen)
94	Selles humaines	4 gr dans SM	3 000g pour 10 min	0.22 µm	Ultracentrifugation au ClCs	Non	Non	Extraction manuelle au formamide et bromide
93	Selles humaines	100-200 mg dans du PBS	Oui mais sans précision	0.45 µm	Ultracentrifugation au ClCs	Oui	Oui	Roche ampliprep
105	Selles de souris	3-4gr	2 500g pour 10 min	0.22 µm	Ultracentrifugation au ClCs + Ultrafiltration (Millipore Centricon)	Oui	Oui 0.2 mg/mL	DNase easyblood and Tissue Kit (qiagen)
87	Selles humaines	500 mg dans du SM	4 700g pour 30 min	0.45 µm	Ultracentrifugation au ClCs	Oui	Oui	Viral DNA/RNA extraction kit (Invitrogen)
119	Selles humaines	400 mg dans SM	4 000g pour 10 min	0.44 µm + 0.22 µm	Ultracentrifugation au ClCs + Ultrafiltration (Millipore Centricon)	Non	Oui 2UI/mL	UltraSens Virus Kit (Qiagen)

Référence du papier	Matériel	Quantité / Tampon	Centrifugation	Filtration	Concentration	Chloroforme	Dnase /Rnase	Extraction de l'ADN
131	Selles de porcs	2 gr dans SM	3 000g pour 10 min et 2 fois 10 000g pour 10min	0.45 µm	Ultracentrifugation au ClCs	Non	Oui	Extraction manuelle au formamide et bromide
139	Selles humaines	200 mg dans SM	2 000g	0.45 µm + 0.22 µm	Non	Oui	Oui 10 UI/mL	Extraction manuelle au formamide et bromide
121	Selles humaines	10 gr dans PBS	3 000g	0.22 µm	Ultrafiltration (Millipore Centricon)	Oui	Oui	PowerSoilVR DNA isolation kit (MO BIO, CA, USA + EZ1 Virus Mini Kit v2.0 (Qiagen, CA, USA)
120	Selles humaines	12 gr dans SM	2 500g	0.45 µm + 0.22 µm	Ultrafiltration (Millipore Centricon) + Ultracentrifugation au ClCs	Non	Oui	QiaAMP DNA mini kit (Qiagen)
117	Selles de porc	25 gr dans SM	3 000g pour 15 min	0.45 µm + 0.22 µm	Ultrafiltration (Millipore Centricon)	Non	Oui 100UI/mL	Extraction manuelle au phenol chloroforme
134	Selles humaines	500 mg dans SM	5 000 rpm 10 min	0.45 µm	PEG	Oui	Oui 8UI/mL	DNeasy Blood & Tissue Kit (Qiagen)
136	Mock community	Non applicable	17 000g pour 3 min	0.8 µm	Non	Non	Oui	Modified QIAmp Viral RNA mini kit
132	Selles de porc	10 gr dans SM	Oui mais non précisé	118 µm	Ultracentrifugation au ClCs + Ultrafiltration	Non	Oui	Extraction manuelle au formamide et bromide

1.5.2 Amplification de l'ADN

Après l'étape d'isolement des phages et l'extraction de l'ADN, la plupart des méthodes trouvées dans les études de phageomes utilisent une étape d'amplification

(appelée amplification à déplacement multiple, MDA) de l'ADN avant le séquençage (Tableau 4). En effet, les kits utilisés au moment des études pour la préparation des librairies en vue du séquençage haut débit nécessitaient une grande quantité d'ADN, l'étape d'amplification était alors incontournable.

L'étape de MDA, utilisant la polymérase répliquative hautement processive du phage $\phi 29$, est connue pour introduire des biais associés à la composition des séquences (en faveur des séquences à fort GC%) ou à la composition de la communauté étudiée ^{140,141}. De plus cela favorise l'amplification des phages à ADN simple brin ¹⁴². Cela a été bien documenté pour les communautés fictives ^{143,144}, mais pas pour les échantillons d'intestin humain. A ce jour il y a donc beaucoup d'études publiées qui ont utilisé le processus de MDA (Tableau 4) et qui ont potentiellement surestimé la quantité de phages à ADN simple brin dans le microbiote intestinal.

En 2011, les kits commerciaux de préparation de librairies pour le séquençage haut débit ont évolué et il y a eu notamment l'apparition du kit Illumina appelé Nextera XT qui nécessite uniquement 1 ng d'ADN. Cela permet de séquencer directement l'ADN de phage sans amplification au préalable. Mais dans ce cas les kits ne prennent pas (ou très peu) en compte les phages à ADN simple brin. Même si les études microscopiques ont montré une grande majorité de phage à ADN double brin ⁷⁷, il serait pertinent d'avoir une technique qui permette d'obtenir une représentation complète et non biaisée de la population phagique. Une étude a proposé des solutions pour parer à ces problèmes. En 2016, Roux *et al.* ¹⁴³, en travaillant sur une « mock community » (communauté artificielle de phages), ont bien montré la sur-amplification des phages à ADN simple brin avec des méthodes de types MDA. Ils ont alors testé un kit commercial appelé « Swift Biosciences 1S Plus kit » pour la préparation des libraires (compatible avec Illumina) et qui prend en compte l'ADN double et simple brin. Ce kit présente de très bons résultats avec une bonne représentativité post-séquençage de la communauté artificielle (phages à ADN double/simple brin).

Par ailleurs, dans toutes ces études il n'y a que les phages à ADN qui sont étudiés. L'étude de Conceição-Neto *et al.*, parue en 2015, propose une méthode afin de prendre en compte les phages à ARN et à ADN¹³⁶. Ils utilisent un kit d'amplification appelé Amplification Kit 2 (WTA2, Sigma Aldrich) faisant intervenir une étape de reverse transcriptase. Le nombre de cycles d'amplification a été étudié pour minimiser le nombre de biais, cela a été testé sur une communauté artificielle de phages. Ils ont

montré, en travaillant sur une communauté artificielle de virome, que 17 amplifications suffisent à amplifier les phages à ADN/ARN de façon reproductible sans créer de biais.

Tableau 4: Tableau récapitulatif des études publiées dans la littérature utilisant, ou non, une étape d'amplification avant le séquençage d'ADN de phage issus du microbiote intestinal.

	Oui	Non
Etape d'amplification de l'ADN avant le séquençage	21,78,80,84,91,93,105,119,139,142, 134,136	87,94,131, 120,121,132

1.5.3 Analyse bio-informatique

Après le séquençage, l'analyse des séquences constitue également une étape majeure. Plusieurs étapes sont nécessaires à l'analyse bio-informatique et seront détaillées ensuite : nettoyage des reads, assemblage des reads, recherche de gènes et identification des contigs phagiques.

L'analyse des métagénomiques de phages est complexe pour différentes raisons (i) Jusqu'à présent, aucune région génomique conservée, à l'image du gène d'ARNr 16S pour le monde bactérien, n'a été trouvée dans les phages. Par conséquent, il n'existe pas de marqueur génétique universel permettant d'identifier les phages dans une communauté virale (ii) Les séquences issues des métagénomiques de phages sont petites, ce sont des morceaux de génomes de phages et il est ainsi difficile de leur attribuer une taxonomie (iii) Contrairement à la communauté bactérienne qui est bien décrite, les phages dans l'intestin sont pour la plupart inconnus. Les premières études métagénomiques sur les matières fécales ont révélé que la plupart des virus bactériens présents dans l'intestin (81% - 93%) sont inconnus et ne peuvent pas se voir attribuer une annotation taxonomique. Les banques de phages sont très pauvres (1 943 phages complet dans la banque GenBank Refseq en novembre 2016), car peu de phages ont été isolés, séquencés et annotés. Cette banque n'est, de plus, pas représentative des phages présents dans le microbiote intestinal. Cela nous permet d'introduire la notion de « Viral dark matter problem »^{80,145} : dans les études métagénomiques, la grande majorité des reads ne peut pas être annotée dans une catégorie fonctionnelle ou taxonomique (Figure 20), cela reflète l'immensité de la diversité phagique du microbiote intestinal.

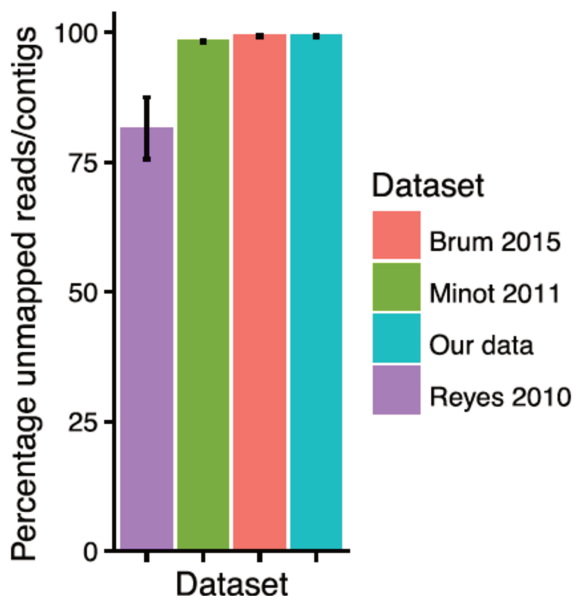


Figure 20: Illustration de la notion de "viral dark matter problem" représentant le pourcentage de reads ou contigs non mappé sur les bases de données pour différents datasets de phageome, d'après Aggarwala ¹⁴⁶

Voici les grandes étapes du traitement bio-informatique de données de séquençage de phageome intestinal :

- Nettoyage des reads

Suite à l'étape de séquençage, des « reads » ou fragment d'ADN ont été produits. Il est nécessaire de nettoyer ces reads. Il s'agit de supprimer les séquences correspondant aux adaptateurs qui ont permis le séquençage haut débit et de supprimer les reads de mauvaise qualité. Le score phred permet d'évaluer la qualité du séquençage en attribuant à chaque base un chiffre (0-100) représentant la probabilité d'une identification correcte, par exemple un score phred de 30 veut dire qu'il y a une chance sur 1000 pour que la base identifiée soit incorrecte. Des logiciels tel que Cutadapt ¹⁴⁷ combiné à UrQt ¹⁴⁸ ou Atropos ¹⁴⁹ sont utilisables pour cette étape.

- Assemblage des reads

Il existe deux grandes approches dans le traitement des données de séquençages : soit les reads sont directement mappés sur des bases de référence, soit il y a une étape d'assemblage des reads en contigs puis une annotation des contigs. Dans le cas de l'étude du phageome intestinal, l'étape d'assemblage est indispensable car il y a très peu de phages séquencés et annotés dans les banques de référence pour permettre de mapper directement la plupart des reads. Les études qui font du mapping directement

sur les reads perdent beaucoup d'information et ne détectent pas de potentiel nouveau phage.

Il existe de nombreux logiciels pour assembler des reads. Une étude très récente ¹⁵⁰ a comparé 13 assembleurs différents pour analyser des données de viromes. C'est le logiciel SPAdes (meta) qui donne les meilleurs résultats aussi bien en ce qui concerne les données simulées, les « mock community » ou les données réelles. Ce résultat est en accord avec une étude précédente ¹⁵¹ qui préconisait également l'utilisation de ce logiciel.

- Recherche d'ORF (Open Reading Frame) sur les contigs

Les reads ont été assemblés en contigs grâce à l'étape d'assemblage. Il est alors possible de rechercher des gènes sur les contigs en recherchant les phases ouvertes de lecture ou « ORF » grâce à des logiciels tel que Prodigal ¹⁵² ou MetaGeneMark ¹⁵³. En métagénomique il est souvent généré un grand nombre de contigs mais de toute petite taille, c'est pourquoi il est souvent choisi de ne garder que les contigs supérieurs à une taille précise (ex : 1 000 paires de bases (pb), 2000 pb ou 10 000 pb) ou ayant au minimum 2 ou 3 ORF détectés.

- Identification des contigs phagiques

Afin d'identifier si les contigs assemblés précédemment sont des phages ou des bouts de phages, plusieurs approches sont possibles pour comparer le contig à des bases de données de phages : (i) une approche classique par alignement de séquences nucléotidiques ou protéiques via BLAST (Blastn ou Blastp), par exemple sur des bases de données de phages connues telles que GenBank Refseq. Les seuils bio-informatiques doivent être choisis avec précaution (E-value, score, couverture, pourcentage d'identité ...) (ii) une approche par homologie protéique grâce à des profils dits HMM « Hidden Markov Model » caractéristiques des phages. Le formalisme HMM permet de générer un modèle statistique d'alignement multiple dans lequel l'apparition des acides aminés dans l'alignement suit un processus stochastique de Markov (la probabilité de l'état n dépend uniquement de l'état n-1). Le logiciel HMMER permet de trouver des homologies entre des profils HMM et les séquences protéines des contigs. Il est possible de trouver des profils HMM caractéristiques de phages dans des banques génériques type PFAM ¹⁵⁴ ou TIGRFAM ¹⁵⁵ ou dans des banques spécialisées comme pVOG ¹⁵⁶.

Nous trouvons plusieurs logiciels, dans la littérature, capables de prédire si les contigs assemblés sont des phages/prophages. Ils sont décrits et analysés dans le Tableau 5

Tableau 5: Présentation des différents logiciels de prédiction de séquences de phages

Nom du logiciel, (référence)	Description	Taille minimum des contigs acceptés	Machine learning	Comparaison de similarité avec des bases de données	Commentaires
Phage Finder ¹⁵⁷	Détection de prophages dans des génomes bactériens grâce à des profils HMM.	Génome complet	non	oui	Non adapté aux données de métagénomique mais possibilité de travailler uniquement avec les profils HMM pour chercher des marqueurs phagiques.
Virsorter ¹⁵⁸	Détection de signaux viraux (phage et prophage) sur des contigs	2 ORFs	non	oui	Attention aux phages de la catégorie 3 qui peuvent être confondus avec des éléments génétiques mobiles
Virfinder ¹⁵⁹	Détection de contigs de phage avec une approche de machine learning sur la fréquence de K-mer dans des données de métagénomique	1000 pb	Oui, régression logistique	non	Performant sur des petits contigs (1-5 kpb)
PHASTER ¹⁶⁰	Annotation des séquences de prophages dans des génomes bactériens ou des plasmides	2000 pb	non	oui	Webserver uniquement
PhiSpy ¹⁶¹	détection de prophages basée sur 5 de leurs caractéristiques distinctes : longueur des protéines, sens	non	Oui, random forest	oui	Manque de spécificité, détecte aussi les éléments génétiques mobiles.

Nom du logiciel, (référence)	Description	Taille minimum des contigs acceptés	Machine learning	Comparaison de similarité avec des bases de données	Commentaires
	transcriptionnel, AT et CG (%), l'abondance des protéines caractéristiques de phages, les points d'insertion des phages et les similarités avec des protéines de phages connus				
vCONTACT 2 162	Attribution taxonomique de contigs de phageome grâce à l'étude des networks de partage de gènes	non	non	oui	
MARVEL 163	Prédiction de phage à ADN double brin dans les données métagénomiques via du Machine learning	2000 pb	Oui, random forest	oui	
VirMiner 164	Prédiction de phage dans les données métagénomiques et analyse globale (interaction phage/bactérie)	1000 bp	Non	oui	Capable de comparer deux groupes de données métagénomiques
PPR-Meta 165	Recherche de phage et de plasmide dans des données métagénomiques via du deep learning	100 bp	oui	oui	
Prophage Hunter 166	Détection de prophages dans des génomes bactériens	Non	Oui, Régression logistique	oui	

On trouve également les formules «tout inclus» avec des pipelines bio-informatiques qui combinent différents outils de prétraitement, d'assemblage et d'annotation. Ces logiciels fournissent un portail en une seule étape pour l'analyse des

données de séquençage de virome. Plusieurs programmes existent actuellement: VirusSeeker ¹⁶⁷, Metavir2¹⁶⁸, VIROME ¹⁶⁹, ViromeScan ¹⁷⁰ et MetaGenome Annotation Pipeline (VMGAP) ¹⁷¹. Ces plateformes « tout inclus » sont indispensables pour que des personnes novices en bio-informatique puissent analyser leurs données, mais il est ensuite difficile d'avoir un esprit critique sur les résultats car l'utilisateur ne maîtrise pas la méthodologie utilisée, il faut donc être prudent dans l'interprétation des résultats.

1.5.4 Mesure de diversité et application en métagénomique

Les mesures de diversité utilisées pour analyser le phageome intestinal reposent sur des métriques initialement développées dans le champ de l'écologie et de l'évolution.

Définition

La **diversité alpha** correspond au nombre d'espèces coexistantes dans un milieu donné. Il existe plusieurs indices pour apprécier cette diversité qui donnent plus ou moins de poids aux nombres d'espèces différentes et à leur abondance relative :

- La **richesse** donne le nombre de taxons dans un assemblage choisi. C'est la mesure la plus ancienne et la plus intuitive de la diversité biologique
- **l'entropie de Shannon** ¹⁷², $H = - \sum p_i \times \log(p_i)$ avec p_i abondance relative du taxon i dans l'échantillon.
- **l'indice de Simpson** ¹⁷³, qui se rapporte à l'uniformité de la distribution des taxons dans la communauté : $D = 1 - \sum_{i=1}^n p_i^2$, avec n le nombre de taxon et p_i l'abondance relative du taxon i dans l'échantillon.

La **diversité beta** est une mesure de la biodiversité qui permet de comparer deux écosystèmes différents. Définition de la diversité beta pour un environnement donné :

$$Beta-diversity = 1 - \frac{Gamma-diversity}{(Alpha-diversity)}$$

Où la diversité gamma est le nombre de taxons communs à l'ensemble des échantillons d'un environnement donné et la diversité alpha est la moyenne des différentes richesses des échantillons de l'environnement étudié.

Afin d'apprécier la diversité beta, il est possible de calculer une distance représentant une différence de composition entre deux échantillons. Plusieurs calculs de distance existent pouvant prendre en compte ou non l'abondance des taxons et la phylogénie entre les taxons. Nous utiliserons dans ce travail la distance de Bray-Curtis ¹⁷⁴ qui prend en compte l'abondance des taxons mais pas leur phylogénie.

Définition de la **distance de Bray-Curtis (BC)** entre les échantillons j et k :

$$BC_{j,k} = 1 - \frac{2 \times \sum_{i=1}^p \min(N_{ij}, N_{ik})}{\sum_{i=1}^p (N_{ij} + N_{ik})}$$
, où N_{ij} est l'abondance brute du taxon i dans l'échantillon j , N_{ik} est l'abondance brute du contig i dans l'échantillon k et p est le nombre de taxons.

Application dans les données de phageome intestinal.

Les différentes mesures de diversité décrites précédemment ont été utilisées afin de définir la diversité bactérienne du microbiote intestinal. Dans ce cas, c'est l'espèce bactérienne (taxon) ou Operational Taxonomic Units (OTU) qui est prise comme unité fondamentale d'analyse ¹⁷⁵. Dans le cas du phageome intestinal, c'est le contig qui va remplir cette mission.

Afin de pouvoir comparer différents échantillons entre eux, il est communément accepté de procéder à une raréfaction ³⁰. En effet, les différences entre le nombre de séquences obtenues pour chaque échantillon peuvent affecter l'estimation de la diversité présente. Le nombre de séquences d'ADN contenues dans chaque échantillon est ainsi uniformisé au sein des différents échantillons en procédant au tirage aléatoire sans remise d'un nombre identique de séquences pour chaque échantillon.

Afin de procéder au calcul de diversités, il y a une étape de mapping des reads de l'échantillon analysé sur sa propre banque de contigs. Cette étape peut être réalisée avec des logiciels de mapping type Bowtie ¹⁷⁶ ou BWA¹⁷⁷. Cela permet de connaître l'abondance de chaque contig dans l'échantillon qui correspond au nombre de reads ayant mappé le contig.

2 Objectifs et questions posées

Afin d'étudier les phages du microbiote intestinal suite à une perturbation antibiotique, les objectifs de ce travail ont été (i) la mise au point d'une méthode d'isolement des phages et de leurs génomes suivi d'un séquençage et de l'analyse bio-informatique et (ii) l'application de cette méthode à une étude clinique chez des volontaires sains (CEREMI) recevant deux céphalosporines de troisième génération au même spectre d'activité mais au mode d'élimination différent, la ceftriaxone ayant une élimination biliaire supérieure au céfotaxime, qui a une élimination majoritairement urinaire.

Le premier objectif répond aux questions suivantes :

- (i) Quelle est la technique d'isolement des phages la plus performante ?
- (ii) Quels biais sont produits par l'utilisation d'une technique d'amplification de l'ADN avant le séquençage ?
- (iii) Quelle profondeur de séquençage est souhaitable pour apprécier la diversité de l'ensemble des phages ?
- (iv) Quels outils et quelle stratégie bio-informatique sont les plus efficaces pour analyser les séquences de phage ?
- (v) Quelles sont les limites méthodologiques dans l'isolement et l'analyse de cette population ?

Le deuxième objectif répond aux questions suivantes :

- (i) Quels sont les phages présents dans le microbiote intestinal ?
- (ii) Quel est l'impact de la céfotaxime et du ceftriaxone sur le phageome intestinal ?
- (iii) Les phages ont-ils un rôle dans le transfert de gènes de résistance sous traitement antibiotique ?
- (iv) Y a-t-il une cohérence entre les résultats des données bactériennes et phagiques ?

Les données du premier objectif sont présentées au chapitre 3.1 « Une méthode simple, reproductible et à moindre coût pour analyser le phageome intestinal : de l'isolement des phages à l'approche bio-informatique ».

Les données du second objectif sont présentées dans les chapitres 3.2 « Etude clinique d'impact du céfotaxime et de la ceftriaxone sur le microbiote intestinal chez des volontaires sains: résultats bactériologiques » et 3.3 « Etude clinique d'impact du céfotaxime et de la ceftriaxone sur le microbiote intestinal chez des volontaires sains: analyse du phageome ».

3 Travaux originaux

3.1 Une méthode simple, reproductible et à moindre cout pour analyser le phageome intestinal : de l'isolement des phages à l'approche bio-informatique

3.1.1 Présentation du contexte de l'étude

La connaissance du microbiote intestinal humain a considérablement augmenté ces dernières années grâce, notamment, à l'apport de la métagénomique. La composante bactérienne du microbiote a été particulièrement bien documentée. Les phages, virus infectant les bactéries, sont présents également dans le microbiote intestinal dans les mêmes proportions que les bactéries, mais ceux-ci ont été beaucoup moins étudiés. Cela est principalement dû à la difficulté d'analyse de cette population aussi bien dans les techniques d'isolement des phages des selles que dans l'analyse bio-informatique des séquences.

Dans ce travail nous avons cherché à définir une méthodologie simple, efficace et reproductible afin d'isoler les phages et leur ADN des selles humaines et de proposer une approche bio-informatique pour analyser les séquences d'ADN.

L'introduction de cette thèse a permis de faire un point bibliographique sur les différentes procédures d'extraction et d'analyse des selles. Actuellement, il n'y a aucune procédure standard pour l'isolement des phages et de leur ADN ainsi que de leur séquençage et analyse bio-informatique. Il est donc très difficile de comparer les études entre elles. Dans ce contexte nous avons testé cinq méthodes différentes d'isolement des phages. La reproductibilité de la méthode choisie a ensuite été étudiée. Nous avons analysé le biais d'une étape d'amplification d'ADN avant l'étape de séquençage, étape largement utilisée dans les études de métagénomique de phages en raison de la faible quantité d'ADN extraite. Nous avons ensuite regardé l'effet de la profondeur de séquençage sur la diversité phagique retrouvée. Puis nous proposons une approche bio-informatique pour retrouver les contigs phagiques, identifier la famille du phage, son style de vie (tempéré ou virulent) et son hôte.

3.1.2 Article publié

Scientific Report, Aout 2019

A simple, reproducible and cost-effective procedure to analyse gut phageome: from phage isolation to bioinformatic approach

Camille d'Humieres^{1,2,3*}, Marie Touchon³, Sara Dion⁴, Jean Cury³, Amine Ghozlane⁵, Marc Garcia-Garcera³, Christiane Bouchier⁶, Laurence Ma⁶, Erick Denamur⁷, Eduardo P. C. Rocha³

¹IAME, UMR 1137, INSERM, Université Paris Diderot, Sorbonne Paris Cité, Laboratoire de Bactériologie, Hôpital Bichat, AP-HP, Paris, France.

²Ecole doctorale Frontières du vivant, Université Paris Diderot, 75013 Paris, France.

³Microbial Evolutionary Genomics, Institut Pasteur, CNRS, UMR3525, Paris, 75015, France.

⁴IAME, UMR 1137, INSERM, Université Paris Diderot, Sorbonne Paris Cité Paris, France.

⁵Institut Pasteur – Bioinformatics and Biostatistics Hub – C3BI, USR 3756 IP CNRS – Paris, France.

⁶Institut Pasteur, BIOMICS, 25-28 rue Dr Roux, Paris, 75015, France.

⁷IAME, UMR 1137, INSERM, Université Paris Diderot, Sorbonne Paris Cité, Laboratoire de Génétique Moléculaire, Hôpital Bichat, AP-HP, Paris, France.

OPEN

A simple, reproducible and cost-effective procedure to analyse gut phageome: from phage isolation to bioinformatic approach

Camille d'Humières^{1,2,3,4}, Marie Touchon⁴, Sara Dion¹, Jean Cury⁴, Amine Ghozlane⁵, Marc Garcia-Garcera⁴, Christiane Bouchier⁶, Laurence Ma⁶, Erick Denamur^{1,7} & Eduardo P.C. Rocha⁴

The microbiota of the human gut is a complex and rich community where bacteria and their viruses, the bacteriophages, are dominant. There are few studies on the phage community and no clear standard for isolating them, sequencing and analysing their genomes. Since this makes comparisons between studies difficult, we aimed at defining an easy, low-cost, and reproducible methodology. We analysed five different techniques to isolate phages from human adult faeces and developed an approach to analyse their genomes in order to quantify contamination and classify phage contigs in terms of taxonomy and lifestyle. We chose the polyethylene glycol concentration method to isolate phages because of its simplicity, low cost, reproducibility, and of the high number and diversity of phage sequences that we obtained. We also tested the reproducibility of this method with multiple displacement amplification (MDA) and showed that MDA severely decreases the phage genetic diversity of the samples and the reproducibility of the method. Lastly, we studied the influence of sequencing depth on the analysis of phage diversity and observed the beginning of a plateau for phage contigs at 20,000,000 reads. This work contributes to the development of methods for the isolation of phages in faeces and for their comparative analysis.

Our understanding of human gut microbiota has drastically increased in the past decade, but the majority of gut metagenomic studies have focused on the bacterial component of the microbiota in healthy subjects and in patients suffering from various pathological conditions. The human gut virome consists of the whole community of viruses in the gut and is mainly composed of bacteriophages (henceforth called phages)¹. In human gut, there are between 10^9 to 10^{12} virus-like particles per gram of faeces^{2,3}, a density within the order of magnitude of bacteria. The contribution of phages to gut microbiota ecology and their effects on human host are just beginning to be highlighted^{4,5}. As an example, an expansion of the number of enteric Caudovirales phages has been observed in patients with Crohn disease and ulcerative colitis⁶.

The human gut phageome varies according to the age of individuals. The infant phageome is highly dynamic and associated with early life changes in its composition [8]. In contrast, the phageome of adults is stable⁷. Healthy adults have a gut phageome mainly constituted of temperate DNA phages belonging to the Caudovirales order (Siphoviridae, Myoviridae, Podoviridae)^{3,8}. There is a high inter-individual variation between the phageome of healthy individuals⁹, but a recent study showed that a small set of phages are found in the majority of healthy people¹⁰. One of these is the crAssphage, a 97 kbp Podoviridae phage that is highly abundant and ubiquitous in the human gut metagenome¹¹. Our limited knowledge about gut phages is reflected in the limited host taxonomic

¹IAME, UMR 1137, INSERM, Université Paris Diderot, 75018, Paris, France. ²AP-HP, Laboratoire de Bactériologie, Hôpital Bichat, 75018, Paris, France. ³Ecole doctorale Frontières du vivant, Université Paris Diderot, 75013, Paris, France. ⁴Microbial Evolutionary Genomics, Institut Pasteur, CNRS, UMR3525, Paris, 75015, France. ⁵Hub de Bioinformatique et Biostatistique – Département Biologie Computationnelle, Institut Pasteur, USR 3756 CNRS, Paris, France. ⁶Genomics Platform, BIOMICS, Institut Pasteur, 25-28 rue Dr Roux, Paris, 75015, France. ⁷AP-HP, Laboratoire de Génétique Moléculaire, Hôpital Bichat, Paris, France. Correspondence and requests for materials should be addressed to C.d.H. (email: camille.dhumieres@aphp.fr)

Received: 5 March 2019

Accepted: 18 July 2019

Published online: 05 August 2019

	AA ₀				
Phage isolation					
Method	I	II	III	IV	V
Quantity (gr)	5	5	5	1	1
Filtration (µm)	2 0.45 0.22	2 0.45 0.22	2 0.45 0.22	2 0.45	2 0.45
Concentration					
Method	UF	UC UF	UC UC-CICs	PEG	UC
Price (Euros)/sample	12.5	17.5	23	0.4	5
Time (WD)	0.5	1	1.5	1*	0.5
DNA extraction	+	+	+	+	+
DNA quantity (ng/gr of faeces)	1058	61.5	54	107	77
MDA	–	–	+	–	–
Sequencing					
Low coverage Number of reads	+“\n”557,844	+“\n”1,052,700	+“\n”618,938	+“\n”1,183,870	+“\n”2,112,614
High coverage Number of reads	–	–	–	+87,148,950	–

Table 1. Main characteristics of the analyses performed on the AA₀ faecal sample. Abbreviations: UF-Ultracentrifugation, UC-Ultracentrifugation; CsCl-Cesium chloride, PEG- Polyethylene glycol, WD-Working day, MDA- Multiple displacement amplification, *Including O/N incubation.

span of sequence databases and results in that the vast majority of reads in human viromes cannot be annotated into a functional or taxonomic category¹².

In the context of a project where we study the bacterial microbiota of the faeces of several healthy individuals during several months¹³, we looked for a simple, efficient, reproducible and inexpensive methodology for (i) phage isolation from faeces, (ii) DNA extraction and sequencing and (iii) computational analyses of the sequencing data. The methods developed for these tasks are very diverse, making comparisons between studies difficult (Table S1). Recently, two studies^{14,15} have compared different methods to isolate phages from faeces. Many of the methods in Castro-Meija *et al.*¹⁴ have merits, but they are not easily scalable to hundreds of samples because they all include a cesium chloride gradient (CsCl) step, which is complex and time consuming. Five other phage isolation methods were presented in Kleiner *et al.*¹⁵, including a method using polyethylene glycol (PEG), which is cheaper, faster and more easily scalable, compared to CsCl. However, they failed to use it due to the formation of a viscous mass. After the phage isolation and DNA extraction step, most of the methods found in phageome papers used an amplification step (called multiple displacement amplification, MDA) of phage DNA before sequencing (Table S1). MDA is known to introduce biases associated with sequence composition^{16,17} and depending on the exact method it can produce a large excess of ssDNA phages¹⁸. This has been well documented for mock communities^{19,20}, but not for human gut samples.

The goal of this work was to produce an easy, reproducible and low-cost methodology to analyse a large number of samples of gut phageome, from viral isolation to bioinformatic analysis. To this purpose, we have tested five different methods of phage isolation. For the one we favoured by the results, we assessed its reproducibility, estimated the bias of an MDA step before sequencing, the effect of sequencing depth, and developed bioinformatic approaches. The prophages of bacterial genomes are ignored in the study, because they are retrieved when sequencing the bacterial fraction of the samples.

Results

A scalable, efficient and cheap method for phage isolation from faeces. We started our study by testing five different methods to isolate phage DNA from faecal samples (described in Table 1). These methods differ mainly by the method of concentration of phages in the sample: method I uses an ultrafiltration technique, method II an ultracentrifugation and an ultrafiltration technique, method III an ultracentrifugation step followed by a CsCl ultracentrifugation, method IV a PEG concentration step and method V a simple ultracentrifugation step. In terms of feasibility, methods II, III and V required an ultracentrifuge, unlike methods I and IV that can be set up easily in a routine lab. The yield of the method, in nanograms of DNA per gram of faeces, varied by a factor of 20 (Table 1). The reference method using CsCl (method III), produced small amounts of DNA and required an additional MDA step before sequencing, while other methods did not. We also estimated the prices of the different concentration steps - including only the reagents - and the working time for phage isolation (Table 1). Method IV was, by far, the cheapest method whereas methods I and V were the fastest. Of note, the time of the method IV included an overnight incubation at 4 °C.

To assess the efficiency of the methods, we sequenced the extracted DNA at low coverage (LC) and obtained 5,525,966 reads, *i.e.*, an average of 1,105,193 reads per method (details in Table 1). The reads were subjected to quality control and then assembled per method, producing a total of 110,709 contigs. The distribution of the sizes of contigs showed a majority of extremely small contigs (median = 389 bp). We excluded all contigs with less than

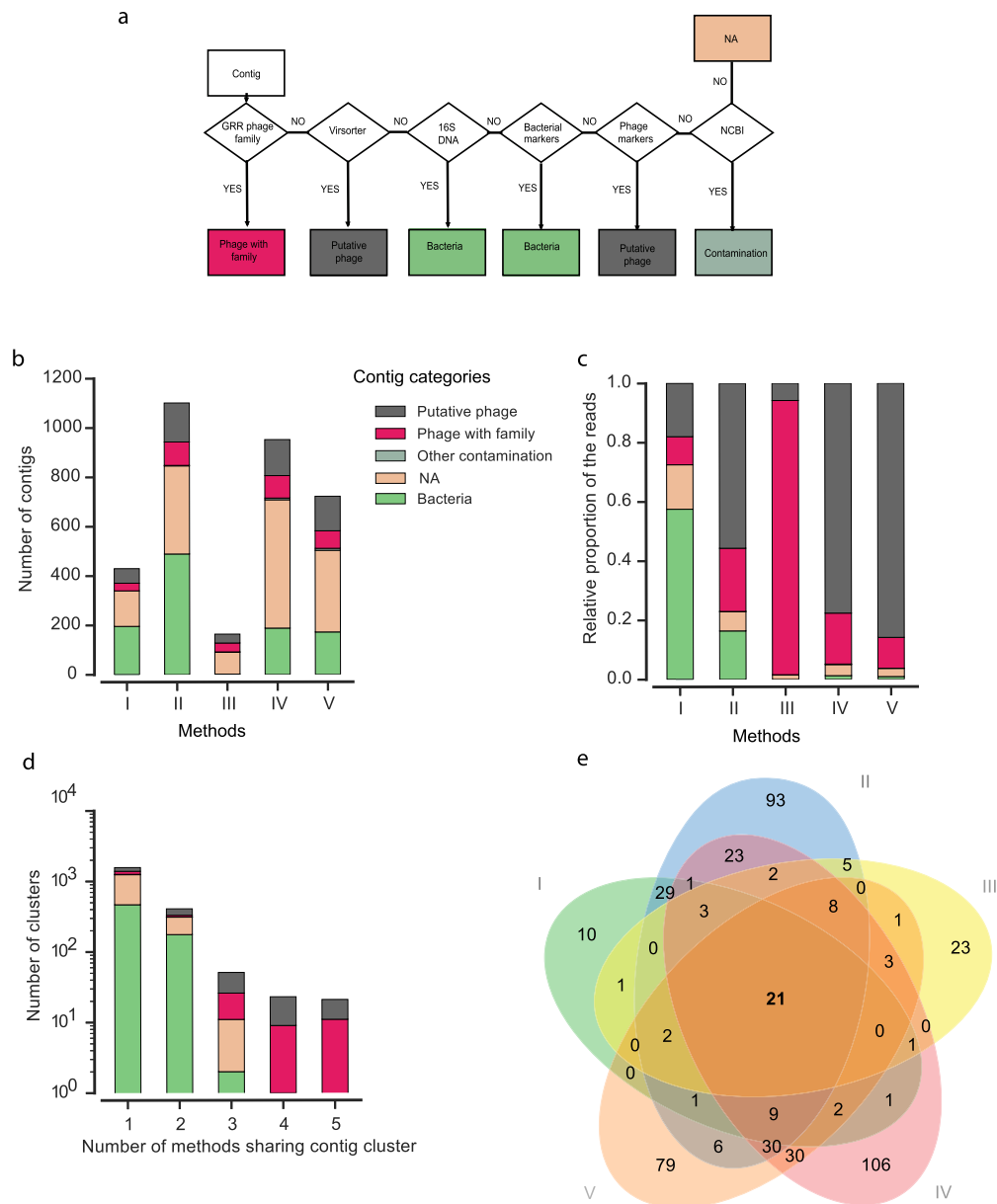


Figure 1. Algorithm to explore contigs and comparison between the five different phage isolation methods (I to V) used to analyse the faeces sample AA₀ of the volunteer AA. **(a)** Flowchart for the taxonomic attribution of the contigs with a complete method. It permitted to classify a contig according to the following groups: “phage with family”, “putative phage”, “bacteria”, “contamination”, “non-attributable (NA)” according to wGRR strategy, Virsorter software, HMM bacterial markers and 16S profile and NCBI database. **(b)** Number of contigs with a minimum of three ORFs, after a rarefaction step, classified following the procedure in **(a)**. **(c)** Frequency of reads mapping the different classes of contigs. **(d)** Number of methods sharing contig clusters. For example, 21 contig clusters (all categorized as phage) were retrieved by all the methods. **(e)** Venn diagram of the number of phage clusters identified for each method (only clusters where the referenced sequence is classified as “phage with family” and “putative phage”). NA: non-attributable.

three open reading frames (ORFs) because these analyses are prone to contamination with small amounts of DNA from bacteria and because very small contigs are hard to identify as phages. This reduced the dataset to 3,369 contigs that represented only 3% of the number of contigs, but were mapped by 84.5% of all reads. Hence, this threshold reduces dramatically the number of contigs, but not the number of reads actually used in the analysis.

To compare the samples in terms of richness (the number of contigs and their content), each sample was rarefied at 322,285 reads. The contigs were then classified as bacteria, phage with family, putative phage, other contamination or non-attributable (NA), using the flowchart described in the Methods (see Fig. 1a). The number of contigs, and their categories, were very diverse across the different methods (Fig. 1b). Method III presented the smallest number of contigs, in opposition to methods II and IV. The majority of contigs could not be assigned to

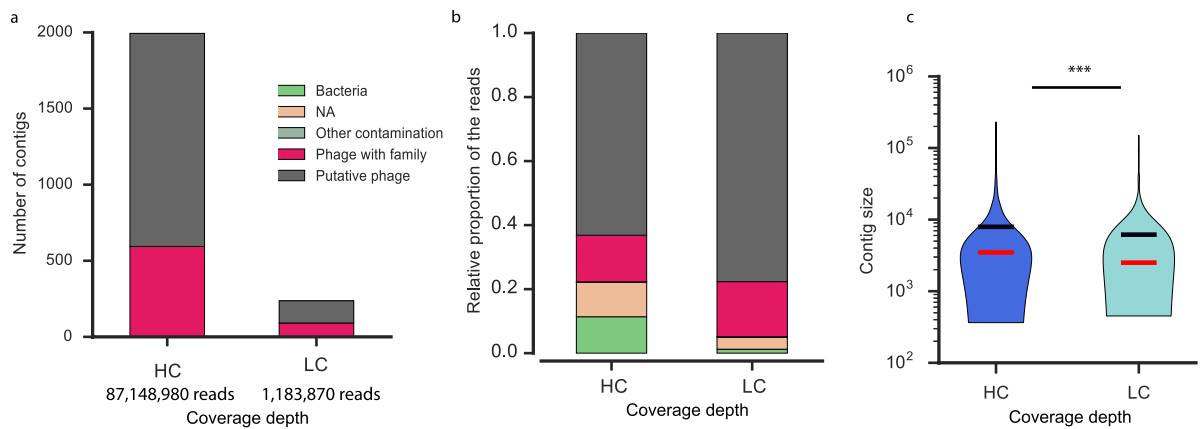


Figure 2. Comparison of the phage contigs obtained from the same sample (AA_0) with different sequencing depths: low coverage (LC) and high coverage (HC). **(a)** Number of phage contigs (classes “phage with family” and “putative phage”) detected in the two datasets. **(b)** Relative proportion of the reads for the different categories (classified with the flowchart presented in Fig. 1a) in the two datasets. **(c)** Violin plot of the distribution of the sizes of phage contigs in the two datasets. The red line corresponds to the median and the black line to the mean. ***Significant difference, Wilcoxon test, $p < 0.001$. NA: non-attributable.

phages or bacteria. But when we weighted each contig by the number of reads mapping it (Fig. 1c), we observed that the vast majority of reads maps the contigs classified as phages in all but the first method, which showed high levels of bacteria DNA. All methods produced phage contigs with median size around 5 kb (Fig. S1).

We then compared the similarities between contigs produced by each method. We clustered all contigs by sequence similarity and assessed the number of methods that contributed for each cluster of contigs (for a variant using more local alignment and producing similar results, see Fig. S2). The resulting 2,075 clusters of non-redundant contigs were mostly identified by one or two methods (Fig. 1d). However, the few contigs identified by four or all of the methods were all classed as phages (or putative phages). A Venn diagram (Fig. 1e) describing the phage (or putative) contigs indicates that methods II and IV exhibited the most diverse sets of phage contigs.

These results led us to continue our study using method IV because it is scalable and rapid (excluding the overnight incubation) while producing a large number of very diverse phage contigs, with relatively little contamination and at a low cost.

The effect of sequencing depth. The depth of sequencing is often decided based on budget constraints. Nevertheless, it is important to assess how much is lost or gained when making this decision. To tackle this point, we sequenced the same sample extracted with method IV at a high coverage (HC) (87,148,950 reads) (Table 1). Statistics comparing this dataset and the previous one (LC, 74 times fewer reads) are described in supplementary material (Table S3). Each dataset was assembled separately, the LC dataset produced 952 contigs (with at least three ORFs) and the HC dataset produced 16,608 contigs (with at least three ORFs). We searched for phages (phages with family or putative phages) on these two pools of contigs using the developed flowchart (Fig. 1a), and found 237 phage contigs in the LC dataset versus 1994 in the HC dataset (Fig. 2a). The number of phage contigs, weighted by the number of reads mapping on them, was very large in both datasets, even if the HC dataset showed more contaminants or unassigned contigs (Fig. 2b). The distribution of sizes of these contigs was significantly different: phage contigs from the HC dataset were significantly longer (mean size of 7,926 bp vs 6,163 bp, $P < 0.001$, Wilcoxon test, Fig. 2c). Hence, the HC analysis has 79 times more reads, but the average contig is only 28.6% larger.

We pooled phage contigs (phage with family and putative phage) from the two datasets and clustered them using CD-HIT. We found 189 clusters that included contigs from both datasets. We found 1803 phage contigs that were exclusively present in the HC dataset, showing that increasing the depth of sequencing significantly increases the number of novel phage contigs. On the other hand, 41 phage contigs were exclusively present in the LC dataset. To enquire on the possibility that these 41 contigs were only partly homologous to contigs in the HC dataset, leading them not to cluster together, we re-ran the CD-HIT with a local alignment (option $-G 0$) and with a less strict coverage on the shorter contig of 50% instead of 95%. This resulted in only 14 contigs exclusively present in the LC dataset (Table S4) representing 2% of the LC reads. Hence, very few contigs assembled in the LC dataset are absent from the HC dataset.

We quantified the gains associated with increased sequencing by simulating different sequencing depths. This was done by randomly selecting reads (in-house script) from the sample IV (HC sequencing). We performed steps from 1% to 90% in steps of 10 (Table S5). Each subset was assembled and analysed independently (in the same way as before). To test the reliability of this simulation method, we plotted on the same figure the number of phage, bacteria and NA contigs for simulated datasets and for the LC dataset. The number of contigs identified in the LC dataset is very close to the curve fitting the results of the simulations from the HC dataset (Fig. 3b), confirming the adequacy of our simulations. Expectedly, the number of contigs increases less than linearly with the number of reads: smaller datasets have proportionally more contigs. Interestingly, the abundance

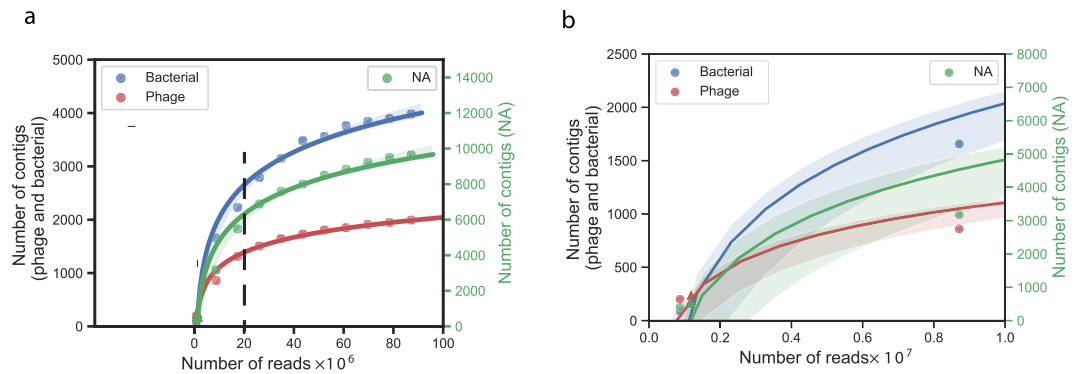


Figure 3. Analysis of the effect of sequencing depth on the number and type of contigs. **(a)** Contigs obtained by re-sampling the high coverage (HC) dataset at different depths. The black y-axis on the left indicates the number of contigs of phage and bacteria. The green y-axis on the right indicates the number of “NA” contigs. The vertical black dashed line represents the sequencing depth recommended (20,000,000 reads). **(b)** Zoom of **(a)** on 0 to 10 millions of reads, simulated data are in circles and observed data [from the low coverage (LC) dataset] in triangles. NA: non-attributable.

Samples									
Samples	AA ^a			BB			CC		
Replicates	1	2	3	1	2	3	1	2	3
Method	IV	IV	IV	IV	IV	IV	IV	IV	IV
Sephadex	YES	YES	YES	YES	YES	YES	YES	YES	YES
For each replicate, 3 different amplification methods were tested									
No MDA	✓		✓		✓		✓		✓
MDA 30'		✓		✓		✓		✓	
MDA 90'			✓		✓		✓		✓
Sequencing	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table 2. Main characteristics of the analyses performed on the faeces of the three human volunteers (AA, BB, CC) for assessing the reproducibility of the method IV and the effects of MDA. ^aThe AA and AA₀ (see in Table 1) faecal samples correspond to two different faecal samples from the same individual. Abbreviations: MDA-Multiple displacement amplification.

of phage contigs, relative to bacterial and unassigned contigs, decreases as the sequence depth increases. This suggests a lower diversity of phage DNA, relative to other elements of the community, and faster saturation of this population.

This analysis does not reveal any obvious point where to stop sequencing a sample, because increasing sequencing depth always provides more information. We observed a decreasing number of novel contigs with increasing sequencing depth. The number of novel phage contigs initiate a plateau after 20,000,000 reads (Fig. 3a), from this point the vast majority of novel contigs are not classed as phages.

MDA decreases the reproducibility of the isolation method. To assess the reproducibility of method IV and how it is affected by MDA, we extracted phage DNA from faeces of three healthy volunteers (AA, BB, CC) (Table 2). Each faecal sample was separated in nine samples: one set of three replicates was treated as above (S), another set of three replicates was amplified by MDA for a short period (SP30), and a final set of three replicates was amplified for a longer period (SP90). The sequencing of these samples resulted in 22,241,488 reads, for an average of 823,758 reads per sample. All reads from the 27 samples were assembled together to obtain a contig reference catalog called ABC-Reference-Contig-Catalog that contained 6,056 contigs with a mean size of 4,796 bp and a median of 2,173 bp. We then mapped the reads of each sample on this set of contigs. The resulting count matrix was rarefied at 716,808 reads to allow straightforward comparisons across samples.

MDA resulted in the fewer different contigs, an effect that increased with the amplification period of time ($P < 0.05$, student t-test). This conclusion remains significant when we adjusted the number of contigs to account for their average size. This shows that MDA decreases the diversity of contigs observed in the samples. The composition of the samples from the same faeces was expected to be more similar (and thus show lower beta-diversity) than that of samples from different faeces. We thus computed the beta-diversity between replicates to assess the reproducibility of the methods. The method without MDA presented the lowest beta-diversity in each healthy volunteer, and longer periods of MDA lead to increased values of beta-diversity (Fig. 4b). We computed a heatmap of the Bray-Curtis dissimilarity between all the 27 samples, complemented with hierarchical clustering to assess the similarities between samples (Fig. 4c). This analysis showed a very clear clustering of the samples by individual. Yet, some amplified samples had very high within-individual Bray Curtis dissimilarity, and

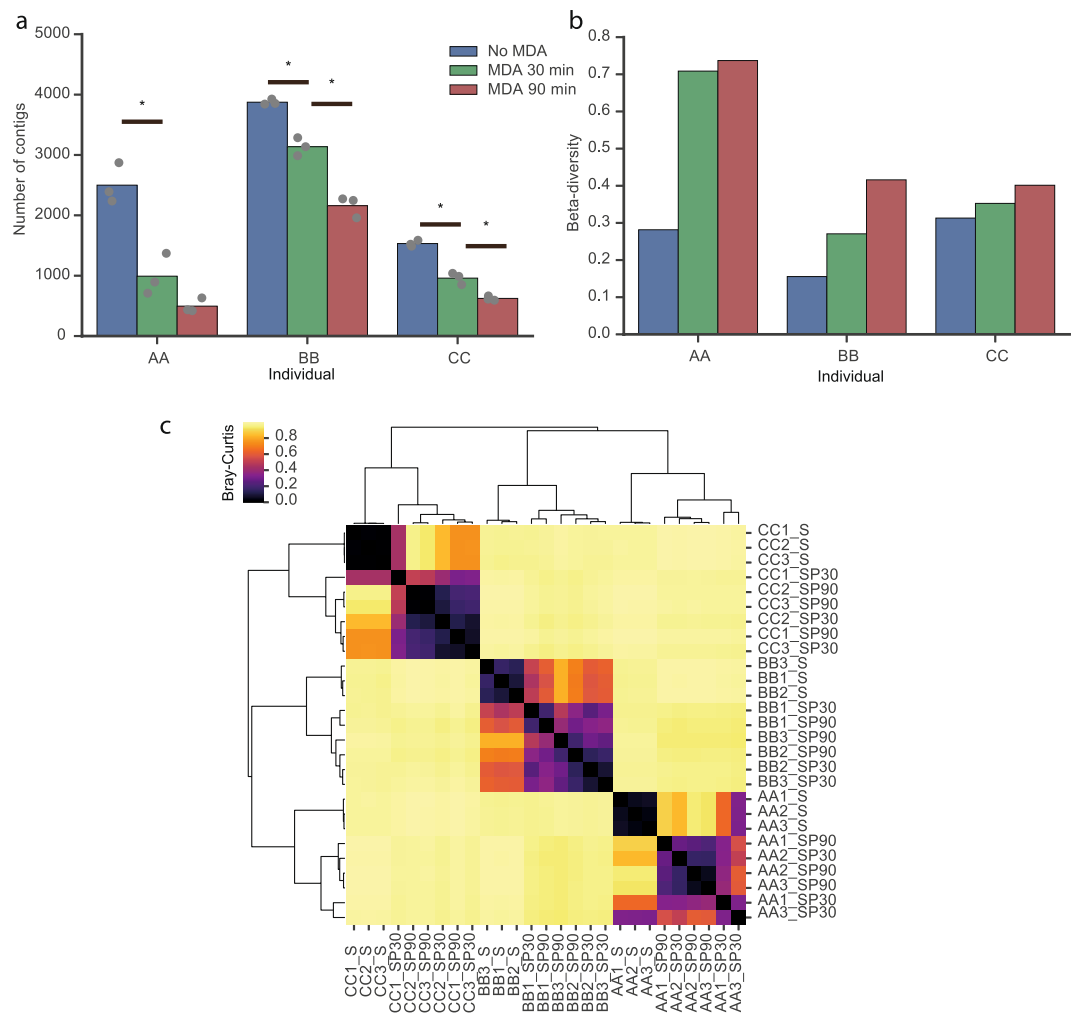


Figure 4. Analysis of the diversity and reproducibility of phage extraction method IV on the samples of three individuals (AA, BB, CC) using two protocols of multiple displacement amplification (MDA) (No MDA, 30 minutes MDA, 90 minutes MDA). **(a)** Number of different contigs (alpha-diversity) in the 27 samples. The bars indicate the average values. The differences in the values of pairwise alpha-diversity between S, SP30 and SP90 are statistically different ($*P < 0.05$, t-student test). **(b)** Beta-diversity computed between replicates. **(c)** Heatmap of the Bray-Curtis dissimilarity between the 27 samples. The trees above and on the left of the heat map were obtained using hierarchical clustering with the function `cluster.hierarchy.linkage` (method ward) from the library `SciPy` (v1.0.1) of python.

SP30 and SP90 samples were sometimes intermingled in the hierarchical classification. Overall, the tightest and most consistent clusters were always those of non-amplified samples, which showed an average intra-group Bray Curtis dissimilarity much smaller than the others (Fig. S3). These samples always clustered together in each individual, contrary to those of the other two methods. In summary, MDA severely decreased the genetic diversity and the reproducibility of the method, and method IV without MDA showed high reproducibility and ability to discriminate between individuals.

The wGRR strategy informs the phageome. We developed a method to attribute a family, a lifestyle or a host to contigs. We searched for sequence similarity between the contigs and (pro)phages of three databases (phage Refseq database, prophages from HMP bacteria and prophages from RefSeq bacteria, see Methods). These results were used to compute a score of similarity called weighted gene repertoire relatedness (wGRR) for each pair of contig/phage. This corresponds to the assignment of the properties of the best hit (phage, prophage) to the global characteristics of the contig (family/host/lifestyle attribution), which facilitates processes that aim at classing the entire contig. To quantify the quality of such assignments, we defined a wGRR cut-off per database. We computed the wGRR of each pair of phages of the Refseq database and found that a wGRR cut off of 5.2 gives 95% correct predictions of phage family (Fig. S4a). However, this cut-off is valid only for comparisons between complete phages. We produced sets of fragments of phage genomes of different sizes to simulate phage contigs of different size. We then recalculated the wGRR cut-off between the complete phage sequences and these different sets (Fig. S4b,c). This produced a smooth increase of the wGRR threshold value with decreasing contig size that

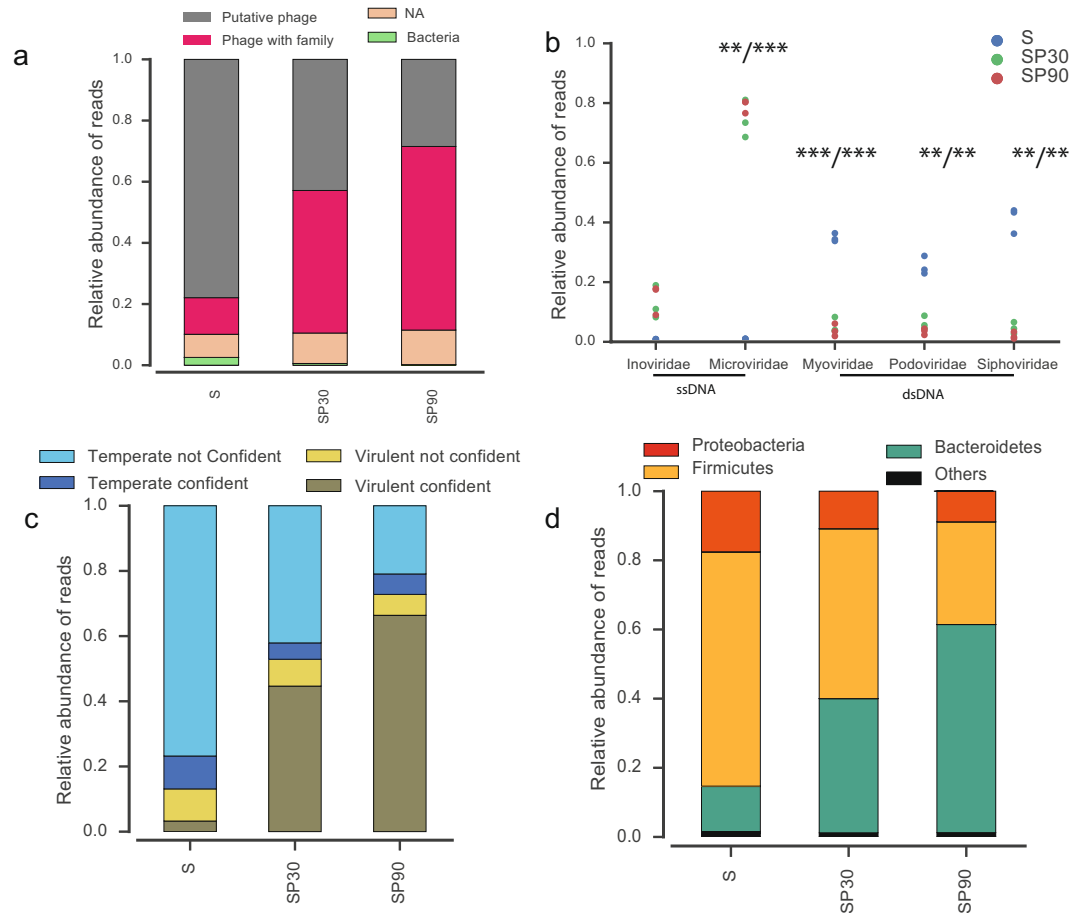


Figure 5. Analysis of the contigs of the healthy volunteers' (AA, BB, CC) samples. **(a)** Bar plots of the relative abundance of reads for each contigs category in each method: method IV for phage isolation following with a 30 min multiple displacement amplification MDA (SP30), 90 min MDA (SP90) or no MDA (S). **(b)** Plot of the relative abundance of reads for each phage family. For Microviridae, Myoviridae and Siphoviridae the relative abundance of reads are significantly different between MDA protocols (SP30, SP90) and no MDA protocols (S), as assessed by paired t-tests ($* < 0.05$, $** < 0.01$, $*** < 0.001$). **(c)** Bar plots of the relative abundance of the reads mapping the contigs of a certain lifestyle. **(d)** Bar plots of the relative abundance of the reads mapping the contigs of a certain host phyla. NA: non-attributable.

could be fit with a negative exponential function (Fig. S4d). We used a similar approach to define wGRR thresholds for lifestyle and host attributions (Table S6).

We then used our method to characterize the phage contigs in terms of family, host and lifestyle attribution and assess the differences between samples in terms of their level of MDA (using the ABC-Reference-Contig-Catalog, details in Table S7). The vast majority of reads were classed as phage contigs. MDA had little effect on the amount of non-assigned or bacteria contigs (Fig. 5a). The ability to identify their family increased with the period of MDA (Fig. 5b). The method without MDA (S) showed a large majority of double stranded DNA (dsDNA) Caudovirales (Siphoviridae, Myoviridae, Podoviridae) and very few single stranded DNA (ssDNA) phages (Inoviridae, Microviridae). This is expected on two accounts. First, dsDNA phages are thought to be the most abundant³. Second, ssDNA phages are not sequenced using the Nextera XT DNA Library Prep Kit that we have used. The use of MDA (SP30 and SP90) drastically increased the proportion of ssDNA phages (Inoviridae, Microviridae) and almost excluded dsDNA phages (Fig. 5b). The predicted lifestyle of the contigs was also very different between methods, with a dramatic increase in contigs classed as virulent phages in samples with MDA; a consequence of the fact that most (73%) ssDNA phages of phage Refseq database are classed as virulent in PHACTS²¹ (Fig. 5c). MDA was also accompanied by an increase of Bacteroidetes phages (Fig. 5d), in agreement with works identifying many Microviridae in this clade².

Biological insights. Our analyses revealed the following categories in the ABC-Reference-Contig-Catalog: 57.6% of the contigs could not be assigned, 18.5% were classed as putative phages, 15% as bacteria and 8.3% to a known family of phages. Importantly, the order of decisions in the decision algorithm used in Fig. 1a has little impact on the overall conclusions since very few contigs are ever classed as phage and bacterial or phage and contamination (all details are in Fig. S5). Intriguingly, 218 out of 506 contigs classed as “Phage with family” were not detected by Virsorter software (Fig. S5). These contigs are small and this is probably the reason why they

were missed by Virsorter (Fig. S6a) even if they have homologs of phage genes. To check that these 218 contigs provide reliable information, we re-analysed Fig. 5 (panels b–d) after removing the contigs missed by Virsorter. The results are similar (Fig. S6b–d), showing that the wGRR approach provides a means of classifying the phage contigs and of identifying novel ones.

We found that only 0.41% of the contigs corresponded to other types of contamination: 2 contigs matched eukaryota DNA, 1 contig matched archaea DNA and 22 contigs had sequence similarities to environmental DNA from uncultured bacteria found by metagenomic methods. Most contigs did not match any database but when we looked at the relative abundance of reads mapped on ABC-Reference-Contig-Catalog, we observed that in method S the vast majority of reads mapped on putative phage or phage contigs (Fig. 5a). We studied more precisely the composition of the gut phageome of the three volunteers when the procedure did not include MDA (Fig. S7). Here, we observed that Caudovirales were largely predominant, but within this group each volunteer had a different predominant family (Podoviridae for AA, Myoviridae for BB and Siphoviridae for CC).

The analysis of the phage lifestyle showed that nearly half of the phage contigs could not be classified according to the lifestyle (46.6%) (Table S7B). The vast majority of the remaining contigs were classed as temperate phages in the non-amplified dataset, although most of the times as “non-confident”. Lastly, most predicted hosts of the gut phages correspond to bacterial phyla frequent in the gut (Firmicutes, Bacteroidetes, and Proteobacteria)^{22,23} (Table S7C).

The crAssphage is a 97 kbp bacteriophage that is frequently found at high abundance in human faeces¹¹. It infects *Bacteroides intestinalis*²⁴. We could not identify the originally identified crAssphage when we mapped (using bowtie2) our reads on the phage DNA sequence. But when we analysed homologies between proteins encoded in the contigs and the proteins of the phage, we found three contigs with many distant homologs to proteins of the crAssphage (with respectively a wGRR of 19, 27 and 26, Fig. S8). All these contigs were found in all the replicates of the three healthy volunteers with or without MDA, even if the amplification decreased its frequency, as expected since it's a dsDNA phage²⁵. Overall, this suggests that our samples include phages distantly related to the crAssphage.

Discussion

There is still a need to standardize protocols to isolate, sequence, and analyse phages from faeces. Here, we tested different phage isolation methods on a single faeces sample to identify a method that could be used in projects with many samples. We decided to test our method (PEG) in real samples, using multiple replicates and different sequencing depths. This allows assessing how it fares in complex samples, at the cost of not knowing the accuracy of the method because its exact composition is unknown. The use of mock communities in the future could allow to quantify the association between the coverage of the contigs and their abundance in the sample, even if the poor current knowledge of most of the human gut phages makes such analyses hard to translate into the context of natural complex communities. Accordingly, the majority of contigs in our samples have no homology to known phages (or bacteria).

All the methods had a centrifugation step that was followed by the filtration of the supernatant; this was performed to eliminate debris. Bacteria were subsequently eliminated by a filtration step and each specific step of phage concentration were performed. We observed a high variability in terms of contigs provided by the five methods, which highlights the difficulty of comparing phageomes obtained with different procedures. The small size of contigs is quite typical of metagenomics datasets. The extraction techniques may lead to damage and fragmentation of the DNA and contribute to the existence of so many small contigs. If so, the effect is similar for the five extraction techniques because we find small contigs in all the samples. The historical reference method, Cesium chloride phage isolation, gives the lowest bacterial contamination and the largest contigs. However, it is very fastidious, has the lowest phage diversity, and shows poor reproducibility, in agreement with a previous study¹⁵. This may be due to the need to amplify DNA given the low DNA yield of the method when dealing with small samples. In any case, the cost and time required for this method make it poorly applicable to large-scale studies. Other isolation phage methods tested in this study provided interesting results. For technical reasons, we could only analyse one sample per method. Hence, we cannot claim that the method we have chosen is necessarily better than the other methods. However, the PEG isolation method presented good reproducibility and the best cost/efficiency ratio, which are very important for large-scale studies. Besides, it produced a large number and a high diversity of phage contigs.

Some steps performed in the method IV (PEG) may have contributed to its good performance: (i) the initial mechanic agitation could help to better dissolve phage faeces in PBS (ii) filtration step with 2 µm and 0.45 µm filters, and not 0.22 µm filter as in method I to III. It has been shown that the 0.22 µm filter reduced by half the number of viral particles observed under the microscope³ (iii) the PEG concentration method seems to be an efficient method to concentrate phages.

For many years, due to the low amount of phage DNA extracted and the high amount of DNA required in the library preparation kits, research teams had to use MDA before sequencing. This study shows, along with several others^{16,17}, that MDA amplifies preferentially ssDNA. There is some controversy on the relative frequency of dsDNA Caudovirales or ssDNA microviridae in the healthy human gut^{7–9,26}. Some of these works used a MDA step before sequencing, which preferentially amplifies ssDNA^{6–9}, whereas others - like ours - used library preparation kits that don't account for ssDNA phages¹⁹. Microscopy studies³ of gut phageome showed a majority of dsDNA phage (from the Caudovirales family). Recently, the use of a novel commercial kit allowed preparing dsDNA and ssDNA libraries compatible with Illumina sequencing in six aquatic samples, and revealed the frequent presence of ssDNA viruses. Nevertheless, these were vastly outnumbered by dsDNA viruses in all cases¹⁹. It remains to be tested in gut phageome. Independently of the real frequency of ssDNA phages in the gut, MDA

leads to less diverse and reproducible phage contigs, posing a problem for the analysis and comparison of multiple samples. Hence, MDA should be avoided, at least when studying the dsDNA component of this community.

Phages evolve rapidly by exchanging functional modules by recombination²⁷, and genome databases do not adequately represent the diversity of the phage world. As a consequence, phage genomes have many genes lacking clear homology to known genomes. To analyse the gut phageome, a global genomic similarity approach can provide interesting additional information relative to approaches based on local similarity. This spurred our approach of assembling reads into contigs and then search similarity to known phages using a global approach (the wGRR index). This index has been previously used to analyse several types of mobile genetic elements, including phages^{28,29}. Here, it allowed the inference of the family, lifestyle and host phyla of many phage contigs, representing a large fraction of the reads. This method is complementary to typical individual protein-based sequence similarity analyses since the wGRR index collects information on the presence of homology and degree of identity of the set of genes of the contig. This global index is thus particularly appropriate to make classifications of the entire contigs. Its limit is that if one wishes to change the reference database, then the fit of a new wGRR cut-off equation will be needed. For example, since the beginning of our work, several novel programs were published to identify prophages in bacterial genomes with higher sensitivity or specificity than PhageFinder. We made a rapid analysis of the effect on the wGRR method of changing the source of prophages by using Phaster³⁰ instead of PhageFinder. It showed that the introduction of a more accurate method further improves the analysis of wGRR, since the threshold upon which classifications can be made decreases (Fig. S9). The wGRR method is general and can be used with any prophage finder method.

The majority of the reads present in samples mapped phage or putative phage contigs, supporting the quality of our isolation method. Nevertheless, the majority of the contigs could not be classed. Most of them are small and poorly covered, they account for a small amount of reads (Fig. S1), which may explain the lack of homologous sequences in the databases. Finally, some contigs were classed as of bacterial origin. It is not possible to determine the relative frequency of them corresponding to bacterial contamination or to virions carrying bacterial DNA arising from generalized transduction.

The phageome of each individual is unique⁷. This study shows that Bray-Curtis dissimilarity between two gut phageomes content is much smaller for samples from the same individual than for samples from different individuals. The Bray-Curtis dissimilarity between two replicates is low (0.08). This provides a measure of the error of the method, and suggests that the method is reproducible.

Conclusion

The phages of the gut have a lot to teach us, so it is essential to know how to isolate them and to extract and analyse their genomes. Our study contributes to the standardization of the methods used to study faeces phageome. It should provide reliable comparisons between studies from different laboratories. Future studies could aim at combining on the same sample the phageome and microbiome analysis to better characterise microbial communities.

Methods

Sample collection. A faecal sample (AA₀) from a healthy adult volunteer (AA) was used to compare the five phage isolation methods designated I to V (Table 1). To compare the reproducibility of method IV and to analyse the bias of MDA procedure, three unrelated healthy adult volunteers (AA, BB and CC) were recruited to provide one faecal sample (Table 2). Samples were stored at 4 °C (for a maximal period of 24 h), homogenized with a Stomacher paddle blender (Interscience, France), and stored at −80 °C.

Phage isolation and DNA extraction. All isolation methods are summarized in Table 1. In each of methods I, II and III, 5 gr from AA₀ were weighed and homogenized in 40 ml of phosphate buffered saline (PBS) (Sigma-Aldrich). Samples were centrifuged three times at 3,000 g for 20 min by taking the supernatant at each time. Then, the supernatant was filtered successively with a 2 μm (hydrophilic glass fibre, Millipore), 0.45 μm (Polyethersulfone membrane, Millipore) and 0.22 μm (Polyethersulfone membrane, Millipore) membranes.

Various procedures were then performed to enrich for phages.

- In method I, a double ultrafiltration was performed on the filtrate with an Amicon Ultra-15 (Merck Millipore) during 20 min at 5,000 g in a swinging out rotor as described in⁷.
- In method II, the filtrate was ultracentrifuged at 100,000 g during 2.5 hours at 4 °C, the phage pellet was suspended in 10 ml of Tris-HCL, NaCl (TN) buffer and then ultrafiltered with Amicon Ultra-15 (Merck Millipore) during 20 min at 5,000 g in a fixed rotor as described in⁸.
- In method III, the filtrate was ultracentrifuged at 100,000 g during 2.5 hours at 4 °C and the phage pellet was suspended in 6 ml of TN buffer. The sample was then loaded onto a CsCl gradient composed of 1.7 g/ml, 1.5 g/ml, 1.35 g/ml steps and centrifuged for 2.5 hours in a SW Ti-41 rotor at 4 °C. The interface between the 1.35 and 1.5 g/ml region was removed based on the protocol published by Thurber *et al.*³¹ and CsCl was removed by dialysis.

In each of methods IV and V, 1 gr from AA₀ was weighted and homogenized in 40 mL of PBS.

- In method IV, the sample was then agitated with a mechanic laboratory agitator during 1 hour at 4 °C at 900 rpm (VWR incubating mini shaker), centrifuged at 17,000 g for 5 min and the supernatant was filtered at 2 μm and 0.45 μm. Phages were then concentrated using a PEG method. One molar solid NaCl and 10% (v/v) PEG 8000 (Sigma) were dissolved into the filtered culture fluid and incubated overnight at 4 °C as recommended for a constant and stable precipitation³². The solutions with the phages were pelleted by

centrifugation at 5,250 g for 1 h at 4 °C and re-suspended in 500 µL of SM buffer (NaCl 100 mM, MgSO₄·7H₂O 8 mM, Tris-Cl 50 mM).

- In method V, the sample was centrifuged twice (one at 2,500 g for 5 min and one at 5,000 g for 15 min), and the supernatant was filtered at 2 µm and 0.45 µm. The filtered fluid was ultracentrifuged at 100,000 g during 2.5 hours at 4 °C in a SWTi-41 rotor as described in¹⁵. The phage pellet was then suspended in 1 mL of SM buffer.

All samples were treated with 10 U/ml of DNase (Sigma) for 30 min at 37 °C followed by 10 min at 65 °C to stop the reaction. DNA was then extracted using the commercial kit “Phage DNA extraction” (Norgen biotek Corp, Ontario, Canada). DNA extracted from method III was amplified with phi-29 DNA polymerase (REPLI-g Mini kit, Qiagen, Germany) (MDA step) as recommended by the manufacturer because there wasn't enough DNA to prepare the sequencing library.

To test the reproducibility of method IV and assess the potential bias of MDA, we produced three different faecal samples (from three different individuals AA, BB and CC) and isolated phages on three different aliquots from each faecal sample (Table 2, 27 samples in total). An additional DNA purification step with a Sephadex column (Sigma-Aldrich, France) was done for the 27 samples in order to improve the purity of DNA. To assess the bias of MDA we made a replicate without MDA before sequencing (S), another with a 30 min MDA (SP30) and the last one with a 90 min MDA as recommended by the manufacturer (SP90) (Table 2).

DNA sequencing. Phage DNA was sequenced with the Illumina MiSeq Nano V2 PE_250 bases method using the Kit Nextera XT with an input of 1 ng DNA, called low coverage (LC) sequencing. Phage DNA from method IV was also sequenced with a deeper coverage with Illumina HiSeq2500 PE_250 bases, called high coverage (HC) sequencing. All samples were sequenced using paired-end reads (numbers of reads mentioned in this manuscript always refer to paired-end reads).

Genome sequences. The sequences and corresponding annotations of 1,943 complete bacteriophage genomes were retrieved from GenBank Refseq (<ftp://ftp.ncbi.nih.gov/genomes>; last accessed November 2016) called phage Refseq database in this paper. The order, the family, the type of nucleic acids and the bacterial host were also extracted from the GenBank files of each phage (when available). This dataset contained 1,756 Caudovirales (90%) of which 3 main families were identified, i.e.: 53% were Siphoviridae, 27% Myoviridae, and 19% Podoviridae. In order to create prophage databases we retrieved bacterial genomes from different locations: (i) 5562 complete bacterial genomes from Refseq GenBank (<ftp://ftp.ncbi.nih.gov/genomes/>, last accessed in November 2016) and (ii) 457 draft genomes of the Human Microbiome Project (HMP) (<http://hmpdacc.org/HMRGD>, last accessed in December 2015).

wGRR. The weighted gene repertoire relatedness (wGRR) is a similarity score between two genetic elements. We used blastp v.2.6.0 (default parameters) to compute sequence identity between proteins encoded in pairs of contigs (A and B). We kept all bi-directional best hits with an e-value lower than 10⁻⁵. The wGRR score is computed as:

$$wGRR = \frac{\sum_i id(A_i, B_i)}{\min(A, B)} \times 100$$

where A_i and B_i is the pair i of homologous proteins present in A and B , $id(A_i, B_i)$ is the percent sequence identity of their alignment, and $\min(A, B)$ is the number of proteins of the contig encoding the fewest proteins (A or B). The wGRR varies between zero and one hundred; it represents the fraction of homologs in the smallest of the two contigs weighted by their sequence similarity²⁸.

Prophage identification. We identified 9927 prophages in the chromosomes of complete bacterial genomes from GenBank Refseq using Phage Finder v4.6³³ as in³⁴. Many of these prophages were very similar. We clustered them with a wGRR cut off of 95% and we took a representative of each cluster (6426 phages). We used Virsorter v.1.0.3³⁵ to identify prophages in draft genomes because this software searches for phages in draft genomes and metagenomic datasets (contrary to Phage Finder). We excluded the putative prophages classed under categories 3 and 6 of VirSorter, which often correspond to other mobile genetic elements (e.g. pathogenicity islands). We thus identified 1243 putative prophages in the draft genomes of the Human Microbiome Project (HMP) (<http://hmpdacc.org/HMRGD>, last accessed in December 2015).

Prediction of phage lifestyles. The lifestyle of phages from phage Refseq database was predicted using PHACTS v.0.3²¹. Predictions were considered as confident only if the average probability score of the predicted lifestyle was two standard deviations (SD) away from the average probability score of the other lifestyle, as recommended by the authors (who claim a precision rate of 99% with this parameter). Using these criteria, we classified 54% of phage Refseq database into 420 virulent and 629 temperate phages.

Trimming, assembly, and mapping of reads. The sequence reads were trimmed (i) to remove the Illumina adapters using Adapter Trimming with cutadapt (v1.12)³⁶ with default parameter values (ii) and reads with low quality were removed using UrQt (v1.0.18)³⁷ with parameters: `-min_QC_length 80 -min_read_size 100 -min_QC_phred 20`. In order to optimize the assembly step, we reduced the redundancy of the reads with the Khmer software (v2.0)³⁸, using a k-mer size of 20 and a median k-mer coverage level of 20, which makes a digital normalization and a trimming of sequences based on the k-mer abundance. The resulting reads were assembled using SPAdes 3.10.1³⁹ with `-meta` option. Gene prediction was done using Prodigal (v2.6.2)⁴⁰ with `-p meta`

option. We excluded genes lacking start and stop codons. In order to focus our analysis on contigs sufficiently large to study genetic contexts, we excluded contigs with less than 3 open reading frames (ORFs).

We used CD-HIT (v4.7, option -g 1, -c 0.95, -aS 0.9 or option -g 1 -G 0 -c 0.95 -aS 0.5 for local alignment)⁴¹ to define families of contigs. For each resulting family, we took a representative contig (the largest) to create a non-redundant contig catalog. We then mapped the samples' reads on this "contigs reference catalog" using bowtie2 (v2.2.9-local-sensitive-local options)⁴² and exploit SAM files with samtools (v1.3.1 with the following commands: views, sort, index, idxstat)⁴³. As result, we obtained a matrix representing the number of reads of a sample (columns) mapping each contig reference catalog (rows) in the dataset.

Distance and diversity analysis. The count matrix was normalized with the following "total-count" method. We first compute the sum of the number of reads for all contigs (k lines) for a given sample (column S_i) where $a_{i,j}$ is the observed count of contig $j \in \{1, \dots, k\}$ in the sample $i \in \{1, \dots, N\}$:

$$S_i = \sum_{j=1}^k a_{i,j}$$

The mean (M) of the column's sum (S_i) where N is the number of sample was calculated as:

$$M = \frac{\sum_{i=1}^N S_i}{N}$$

Finally, a size factor (SF_i) was determined for each column (i):

$$SF_i = \frac{S_i}{M}$$

Each value of the count matrix was then divided by the size factor of the respective column in order to normalize each column.

We computed the pairwise dissimilarity between the samples using the Bray-Curtis compositional dissimilarity measure between columns of the normalized count matrix. This was done with the skbio python package (v4.2) (<http://scikit-bio.org/>). For diversity analysis, we rarefied the data, *i.e.*, we generated a count matrix where each column had the same sum of reads by randomly sampling the original data. This procedure can result in a waste of significant data, and alternatives have been designed to deal with this problem⁴⁴. Yet, in this work samples were of very similar sizes (by design), so rarefaction can be done without much loss of information. The count matrix was rarefied with the "rarefy" function of the vegan package in R⁴⁵.

The richness of a sample was defined as the number of different contigs (with at least three ORFs). Alpha-diversity was defined as the average of the richness of different samples for a given condition.

Beta-diversity within a set of samples was calculated as:

$$\text{Beta-diversity} = 1 - \frac{\text{Gamma-diversity}}{(\text{Alpha-diversity})}$$

where gamma-diversity is the total number of contigs present in the samples (with a minimum of one read that mapped on the contig).

Identification of phage contigs. Putative phage contigs were detected with Virsorter v1.0.3 (option db2 and -virome 1). Given the small size of contigs, we used putative prophages from all categories given by the software. Additionally, we searched for genes that could constitute markers of the presence of phages. HMMER (v3.0, option-cut_ga)⁴⁶ was performed on proteins with HMM phage profile from PFAM (30.0) and TIGRFAM (15.0) with a keyword like "capsid, portal, protease, replication, tails" (details of the profiles are given in supplementary data, Table S2).

Identification of bacteria contigs and non-bacterial contamination contigs. To identify bacterial DNA, we searched for the presence of specific bacterial markers in the contigs. We used 128 out of 139 HMM profiles (Table S2) of bacterial conserved marker genes⁴⁷ using HMMER (v3.0, option-cut_ga). We removed 11 profiles from the original dataset because they were matched by some proteins in the phage Refseq database (indicating that they are not specific to bacteria) (Table S2). We also searched the contigs for similarity with a 16S HMM profile using HMMER (v3.0, option -E 0.1).

To identify contigs representing non-bacterial contamination (human, archaea and eukaryotic virus DNA), we used Blastn (v2.6.0, options -evalue 0.00001 -max_target_seqs. 10 -task megablast, select only match with a coverage higher than 90%) to search for homologs of the contig genes in the NCBI-nt catalogue (last accessed September 2017). We classed the contigs with hits using the NCBI taxonomy, attributed using the python package taxabd (<https://github.com/HadrienG/taxadb>).

Classification of phage contigs. We computed the wGRR between all contigs with more than three ORFs and between them and phages from the three phage databases (see above). The contigs were sorted with the following data: phage family (*i.e.*, Myoviridae, Podoviridae, Siphoviridae, Inoviridae, Microviridae), lifestyle (virulent or temperate), and bacterial host (phyla) associated with the phage with highest wGRR value. However, since these values were sometimes very low, we defined a statistically meaningful wGRR threshold for each type of information. First, we computed all wGRR values between all pairs of phages (or prophages) in each database.

We calculated the wGRR cut-off between two elements of known classification so that if one of the elements had a given classification, the other had at least a probability of 95% of having the same classification. This method provides wGRR thresholds to class phages given the classification of the best hit (*i.e.*, given the known phage with highest wGRR).

To apply the previous method to metagenomics, we had to devise a way of dealing with phage sequence fragmentation in small contigs. We recalculated the wGRR cut-off in the same way, except that we made the analyses on fragments of known phage genomes. We used fragments of different sizes (5, 10, 15, 20, 30, 40, 50, and 60 genes). This allowed adjusting the wGRR cut-off to the size of the smallest element (the contig in the real analysis). For each database and for each type of information (family, lifestyle, host) we fitted a negative exponential curve in the data using curve fit function from Scipy package (v1.0.1). This equation has the form:

$$Y = \alpha \times e^{-\beta X} + \gamma$$

where Y is the wGRR cut-off and X is the number of proteins encoded in the fragment of the phage, α , β and γ depends of the database used and the information searched.

We then computed the best wGRR between the contig and known phages from the phage Refseq database. The contig was classed using the information of the best hit among the phages if the wGRR is higher than the cut-off determined with the equation above.

We used the wGRR values to classify the phage contigs in terms of taxonomy. The classification of the lifestyle depended on the number of proteins encoded on the contig. We used the PHACTS software prediction for contigs with more than 19 proteins (4 categories: virulent/temperate confident or virulent/temperate non-confident). For shorter contigs, PHACTS provided unreliable classifications, and we used the wGRR approach. For host phyla attribution, we used the wGRR calculated in comparisons with three different phage/prophage databases (see section genome sequences and prophage identification). We only required one hit above the threshold in one database to make a phylum attribution, but we left without classification the contigs for which there was a contradiction between the databases.

Decision algorithm. Flowchart in Fig. 1a presents the algorithm for taxonomic attribution of the contigs. This permitted to classify a contig according to the following groups: “phage with family”, “putative phage”, “bacteria”, “contamination”, “non-attributable (NA)” according to wGRR strategy, Virsorter software, HMM bacterial markers and 16S profile and NCBI database as described above.

Ethics approval and consent to participate. This work was approved by the Comité d’Evaluation de l’Ethique des projets de Recherche Biomédicale (CEERB) Paris Nord” (Institutional Review Board -IRB 00006477- of HUPNVS, Paris 7 University, AP-HP). We confirm that all methods were performed in accordance with relevant guidelines and regulations, and that informed consent was obtained from all participants.

Data Availability

DNA sequences are available at the EBI ENA under accession number PRJEB30796. Scripts used for the analyses and the figures can be provided here https://gitlab.pasteur.fr/cdhumier/phage_isolation_to_bioinformatic_approach.

References

- Breitbart, M. *et al.* Metagenomic Analyses of an Uncultured Viral Community from Human Feces. *J. Bacteriol.* **185**, 6220–6223 (2003).
- Kim, M.-S., Park, E.-J., Roh, S. W. & Bae, J.-W. Diversity and Abundance of Single-Stranded DNA Viruses in Human Feces. *Appl. Environ. Microbiol.* **77**, 8062–8070 (2011).
- Hoyles, L. *et al.* Characterization of virus-like particles associated with the human faecal and caecal microbiota. *Res. Microbiol.* **165**, 803–812 (2014).
- Reyes, A., Wu, M., McNulty, N. P., Rohwer, F. L. & Gordon, J. I. Gnotobiotic mouse model of phage-bacterial host dynamics in the human gut. *Proc. Natl. Acad. Sci. USA* **110**, 20236–20241 (2013).
- Lengeling, A., Mahajan, A. & Gally, D. L. Bacteriophages as pathogens and immune modulators? *mBio* **4**, e00868–00813 (2013).
- Norman, J. M. *et al.* Disease-Specific Alterations in the Enteric Virome in Inflammatory Bowel Disease. *Cell* **160**, 447–460 (2015).
- Minot, S. *et al.* Rapid evolution of the human gut virome. *Proc. Natl. Acad. Sci.* **110**, 12450–12455 (2013).
- Minot, S. *et al.* The human gut virome: Inter-individual variation and dynamic response to diet. *Genome Res.* **21**, 1616–1625 (2011).
- Reyes, A. *et al.* Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* **466**, 334–338 (2010).
- Manrique, P. *et al.* Healthy human gut phageome. *Proc. Natl. Acad. Sci. USA* **113**, 10400–10405 (2016).
- Dutilh, B. E. *et al.* A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat. Commun.* **5**, 4498 (2014).
- Aggarwala, V., Liang, G. & Bushman, F. D. Viral communities of the human gut: metagenomic analysis of composition and dynamics. *Mob. DNA* **8**, 12 (2017).
- Burdet, C. *et al.* Ceftriaxone and Cefotaxime Have Similar Effects on the Intestinal Microbiota in Human Volunteers Treated by Standard-Dose Regimens. *Antimicrob. Agents Chemother.* **63**, (2019).
- Castro-Mejía, J. L. *et al.* Optimizing protocols for extraction of bacteriophages prior to metagenomic analyses of phage communities in the human gut. *Microbiome* **3**, 64 (2015).
- Kleiner, M., Hooper, L. V. & Duerkop, B. A. Evaluation of methods to purify virus-like particles for metagenomic sequencing of intestinal viromes. *BMC Genomics* **16** (2015).
- Yilmaz, S., Allgaier, M. & Hugenholtz, P. Multiple displacement amplification compromises quantitative analysis of metagenomes. *Nat. Methods* **7**, 943–944 (2010).
- Džunková, M. *et al.* Direct sequencing from the minimal number of DNA molecules needed to fill a 454 picotiterplate. *PLoS One* **9**, e97379 (2014).
- Kim, K.-H. & Bae, J.-W. Amplification methods bias metagenomic libraries of uncultured single-stranded and double-stranded DNA viruses. *Appl. Environ. Microbiol.* **77**, 7663–7668 (2011).

19. Roux, S. *et al.* Towards quantitative viromics for both double-stranded and single-stranded DNA viruses. *PeerJ* **4**, e2777 (2016).
20. Marine, R. *et al.* Caught in the middle with multiple displacement amplification: the myth of pooling for avoiding multiple displacement amplification bias in a metagenome. *Microbiome* **2**, 3 (2014).
21. McNair, K., Bailey, B. A. & Edwards, R. A. PHACTS, a computational approach to classifying the lifestyle of phages. *Bioinforma. Oxf. Engl.* **28**, 614–618 (2012).
22. Li, J. *et al.* An integrated catalog of reference genes in the human gut microbiome. *Nat. Biotechnol.* **32**, 834–841 (2014).
23. Hugon, P. *et al.* A comprehensive repertoire of prokaryotic species identified in human beings. *Lancet Infect. Dis.* **15**, 1211–1219 (2015).
24. Shkoporov, A. N. *et al.* Φ CrAss001 represents the most abundant bacteriophage family in the human gut and infects *Bacteroides intestinalis*. *Nat. Commun.* **9**, 4781 (2018).
25. Yutin, N. *et al.* Discovery of an expansive bacteriophage family that includes the most abundant viruses from the human gut. *Nat. Microbiol.* **3**, 38–46 (2018).
26. Waller, A. S. *et al.* Classification and quantification of bacteriophage taxa in human gut metagenomes. *ISME J.* **8**, 1391–1402 (2014).
27. Botstein, D. A theory of modular evolution for bacteriophages. *Ann. N. Y. Acad. Sci.* **354**, 484–490 (1980).
28. Cury, J., Touchon, M. & Rocha, E. P. C. Integrative and conjugative elements and their hosts: composition, distribution and organization. *Nucleic Acids Res.* **45**, 8943–8956 (2017).
29. Bobay, L.-M., Rocha, E. P. C. & Touchon, M. The Adaptation of Temperate Bacteriophages to Their Host Genomes. *Mol. Biol. Evol.* **30**, 737–751 (2013).
30. Arndt, D. *et al.* PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* **44**, W16–W21 (2016).
31. Thurber, R. V., Haynes, M., Breitbart, M., Wegley, L. & Rohwer, F. Laboratory procedures to generate viral metagenomes. *Nat. Protoc.* **4**, 470–483 (2009).
32. Jones, T. H. & Johns, M. W. Improved Detection of F-Specific RNA Coliphages in Fecal Material by Extraction and Polyethylene Glycol Precipitation. *Appl. Environ. Microbiol.* **75**, 6142–6146 (2009).
33. Fouts, D. E. Phage_Finder: Automated identification and classification of prophage regions in complete bacterial genome sequences. *Nucleic Acids Res.* **34**, 5839–5851 (2006).
34. Touchon, M., Bernheim, A. & Rocha, E. P. Genetic and life-history traits associated with the distribution of prophages in bacteria. *ISME J.* **10**, 2744–2754 (2016).
35. Roux, S., Enault, F., Hurwitz, B. L. & Sullivan, M. B. VirSorter: mining viral signal from microbial genomic data. *PeerJ* **3**, e985 (2015).
36. Marcel Martin, M. M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBNet.journal* **17**, 10–12 (2011).
37. Modolo, L. & Lerat, E. UrQt: an efficient software for the Unsupervised Quality trimming of NGS data. *BMC Bioinformatics* **16**, 137 (2015).
38. Crusoe, M. R. *et al.* The khmer software package: enabling efficient nucleotide sequence analysis. *F1000Research* **4**, 900 (2015).
39. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol. J. Comput. Mol. Cell Biol.* **19**, 455–477 (2012).
40. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
41. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
42. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
43. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.* **25**, 2078–2079 (2009).
44. McMurdie, P. J. & Holmes, S. Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Comput. Biol.* **10**, e1003531 (2014).
45. Oksanen, J. *et al.* *vegan: Community Ecology Package* (2018).
46. Eddy, S. R. Accelerated Profile HMM Searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).
47. Rinke, C. *et al.* Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431–437 (2013).

Acknowledgements

We thank Stevann Volant for his help in count matrix normalization, Remi Denise and Amandine Perrin for their help in script optimization. We thank Etienne Ruppé for insightful comments on a previous version of this manuscript. High-throughput sequencing was performed on the Institut Pasteur Genomics Platform, member of “France Génomique” consortium (ANR10-INBS-09-08). We thank two anonymous reviewers providing helpful comments on earlier drafts of the manuscript. This work was supported by a grant from the “Agence Nationale de la Recherche” (ANR-16-CE15-0022). It was also partially supported by a grant from the “Fondation pour la Recherche Médicale” to ED (Equipe FRM 2016, grant number DEQ. 20161136698).

Author Contributions

E.R., E.D. and C.D. conceived and planned the experiments. C.D. and S.D. carried out the experiment of phage isolation. C.B. and L.M. made the DNA sequencing. C.D. helped by M.T., J.C., M.G.G. and A.G. performed the analysis. C.D. wrote the manuscript, supervised by E.R. and E.D. All authors have read and approved the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-47656-w>.

Competing Interests: The authors declare no competing interests.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019

Supplementary information for:

A simple, reproducible and cost-effective procedure to analyse gut phageome: from phage isolation to bioinformatic approach

Camille d'Humieres^{1,2,3,4*}, Marie Touchon⁴, Sara Dion¹, Jean Cury⁴, Amine Ghozlane⁵, Marc Garcia-Garcera⁴, Christiane Bouchier⁶, Laurence Ma⁶, Erick Denamur^{1,7}, Eduardo P. C. Rocha⁴

¹IAME, UMR 1137, INSERM, Université Paris Diderot, 75018, Paris, France.

²AP-HP, Laboratoire de Bactériologie, Hôpital Bichat, 75018 Paris, France.

³Ecole doctorale Frontières du vivant, Université Paris Diderot, 75013 Paris, France.

⁴Microbial Evolutionary Genomics, Institut Pasteur, CNRS, UMR3525, Paris, 75015, France.

⁵Bioinformatics and Biostatistics Hub, Institut Pasteur, C3BI, USR 3756 IP CNRS, Paris, France.

⁶BIOMICS, Institut Pasteur, 25-28 rue Dr Roux, Paris, 75015, France.

⁷AP-HP, Laboratoire de Génétique Moléculaire, Hôpital Bichat, AP-HP, Paris, France.

THIS SUPPLEMENTARY FILE CONTAINS:

SUPPLEMENTARY FIGURES.....	2
SUPPLEMENTARY TABLE.....	11
LEGENDS OF EXCEL SUPPLEMENTARY TABLES	17
LIST OF SUPPLEMENTARY REFERENCES.....	18

Supplementary figures

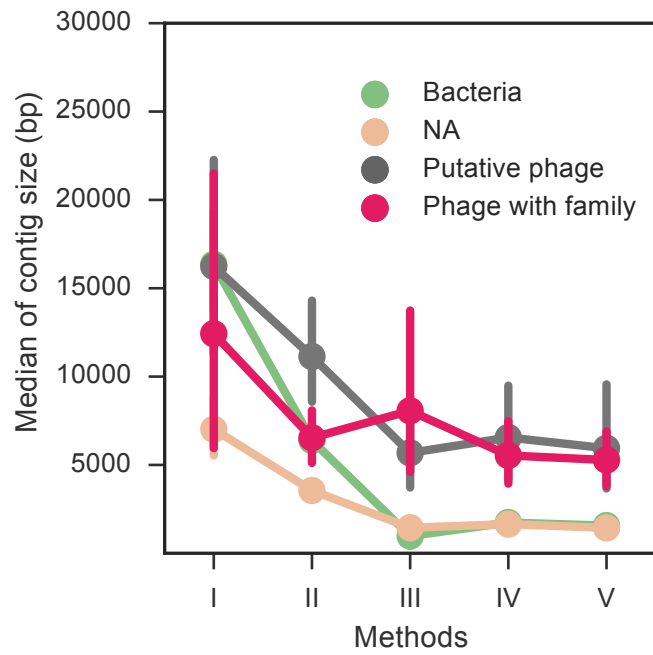


Figure S1: Median of the contigs size for each method and for each category (bacteria, “NA” and phage). NA: non-attributable

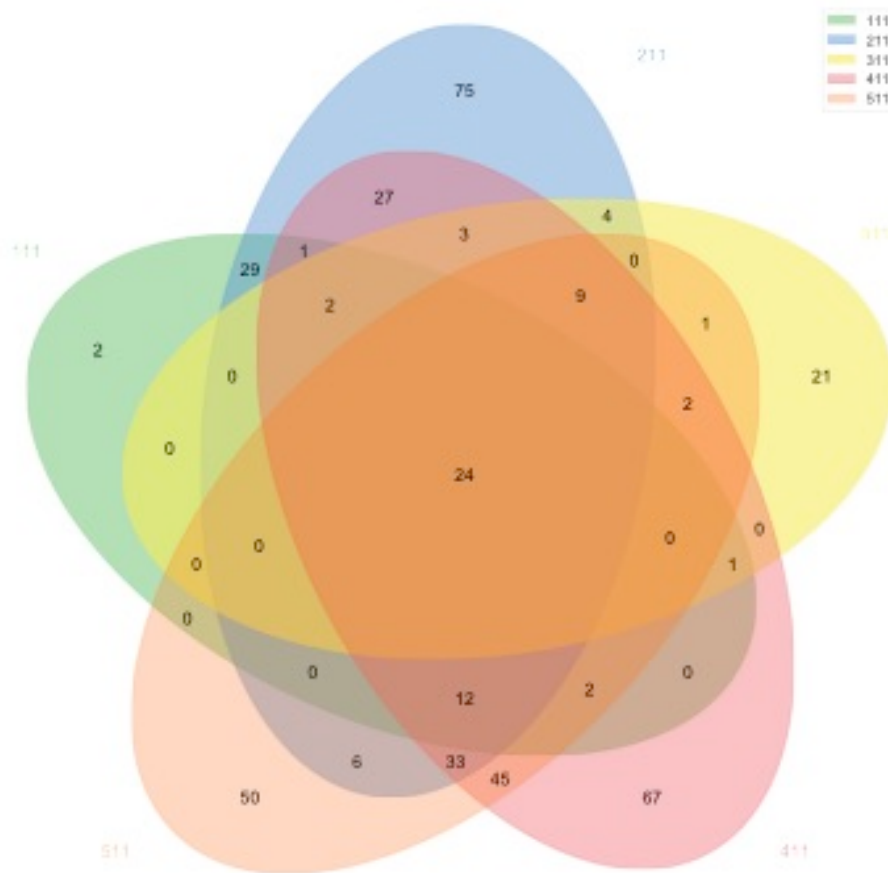


Figure S2: Venn diagram of phage clusters between each method after a local clusterisation step

Description: Venn diagram of the number of phage clusters identified for each method (only clusters where the referenced sequence is classified as "phage with family" and "putative phage"). Clustering step was performed with CD-HIT local criteria.

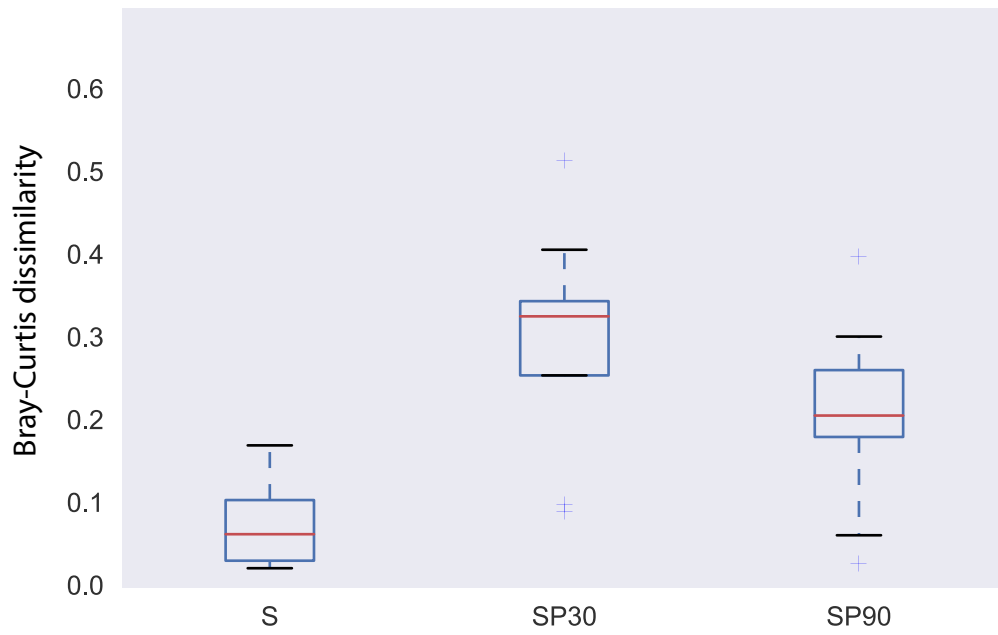


Figure S3: Box plot of the Bray Curtis dissimilarity between replicate for each methods

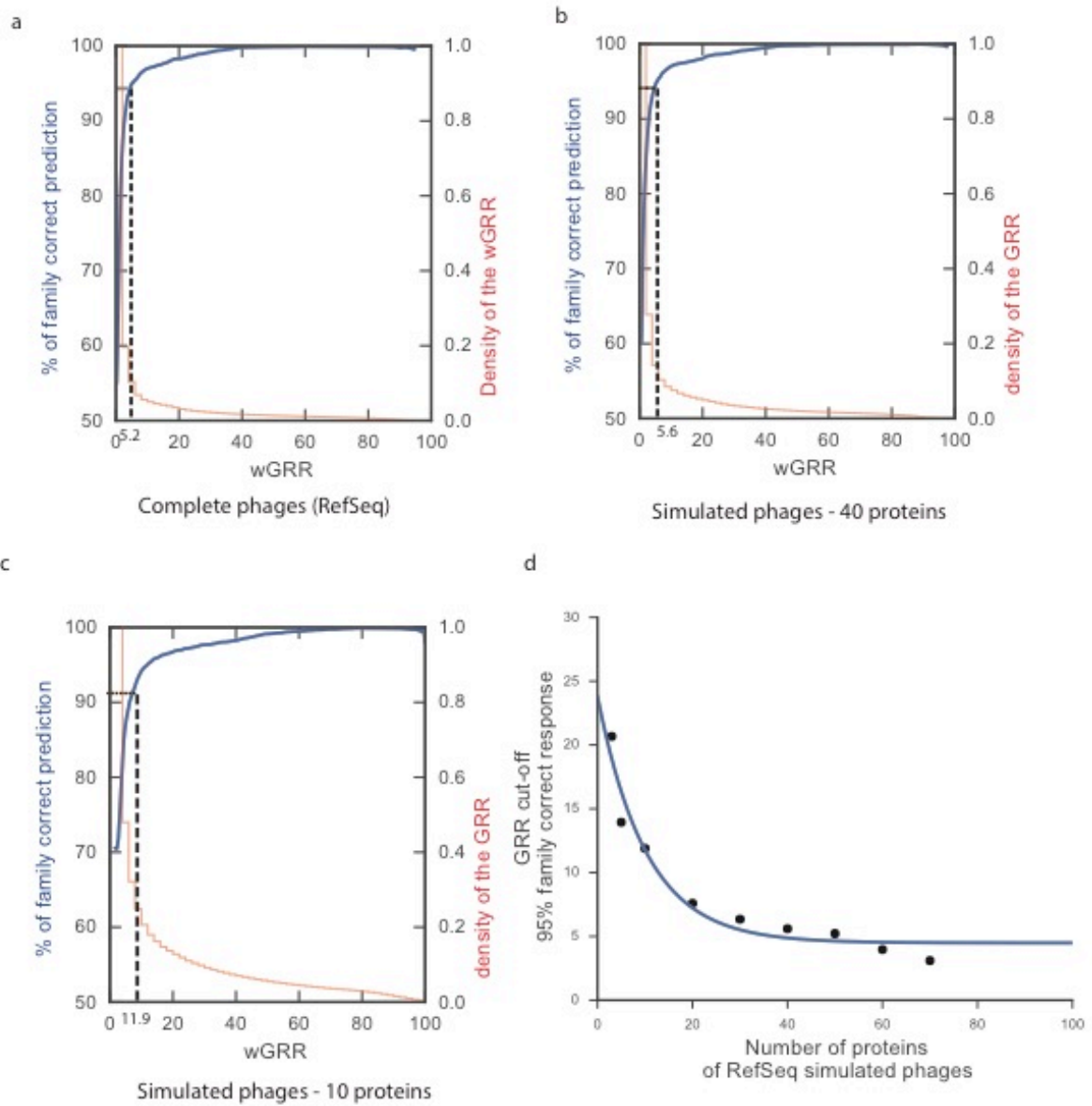


Figure S4: Determination of wGRR cut off for phage family attribution with phage Refseq Database.

Description: a: Evolution of the percentage of phage family correct response in terms of wGRR value for the phage of phage Refseq database. The blue line represents the proportion of phage family correct response. The red line is the decreasing cumulative density of wGRR. The dashed line represents the wGRR cut-off (5.2) for 95% correct family phage predictions. b: Evolution of the percentage of phage family correct response in terms of wGRR value between phage of phage Refseq database and a sub-sampling of this

database with only part of phage of 40 proteins. The dashed line represents the wGRR cut-off (5.6) for 95% correct family phage predictions. c: Evolution of the percentage of phage family correct response in terms of wGRR value between phage of phage Refseq database and a sub-sampling of this database with only part of phage of 10 proteins. The dashed line represents the wGRR cut-off (11.6) for 95% correct family phage predictions. d: wGRR cut-off for a phage family correct response in terms of size of simulated phages. This curve fit is 0.95 with this equation $Y = 19.33 \times e^{-0.098X} + 4.49$ where Y is the wGRR cut-off and X the number of proteins of the simulated phages (or of the contigs).

	GRR phage family	Virsorter putative phage	16S DNA bacterial	HMM Markers bacterial	HMM markers phage	NCBI
	506	288	0	6	274	17B* – 14V*
		663	0	11	329	19B* – 11V*
		16S DNA Bacterial	6	0	0	0
			HMM markers bacterial	443	26	49B
				HMM markers phage	444	34B* – 2V* – 1E*
					NCBI	506

* B: Bacterial contig, V: Virus contig, E: Eukaryota contig

Figure S5: Details of the repartition of the 6055 contigs in the ABC-Reference-Contig-Catalog with the decision algorithm in Figure 1a.

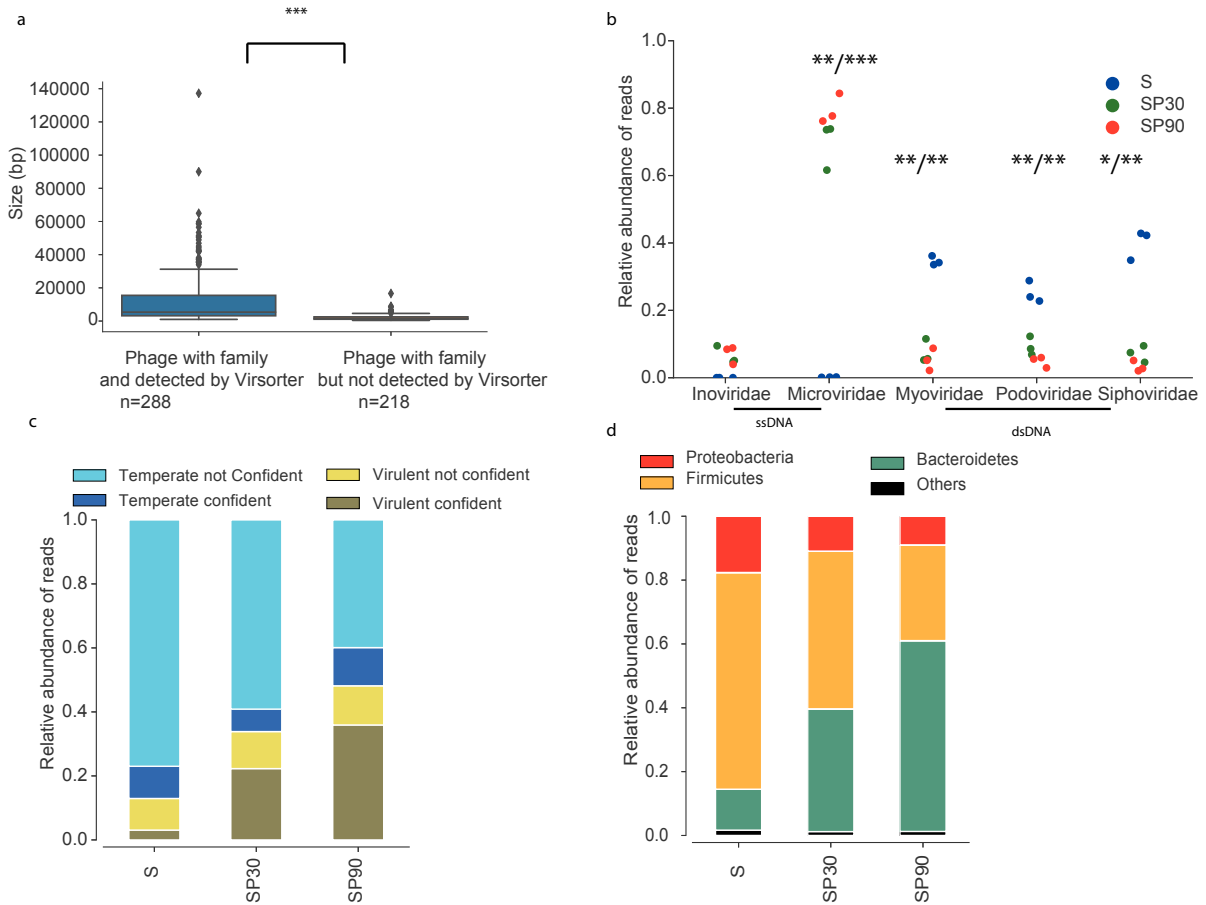


Figure S6: Re-analysis of the contigs of the healthy volunteers' (AA, BB, CC) samples, removing the 218 contigs classified as "Phage with family" but not detected by Virsorter. (a): Boxplot of the size (bp) of the 508 contigs classified as "Phage with family" and divided into two groups (detected or not by Virsorter as phages, *** $p < 0.001$, Wilcoxon test) (b): Plot of the relative abundance of reads for each phage family, conciliating only the 208 contigs classified as "Phage with family" and detected by Virsorter. For microviridae, myoviridae and siphoviridae the relative abundance of reads are significantly different between MDA protocols (SP30, SP90) and no MDA protocols (S), as assessed by paired t-tests (* < 0.05 , ** < 0.01 , *** < 0.001). (c) Bar plots of the relative abundance of the reads mapping the contigs of a certain lifestyle, removing the 218 contigs classified as "Phage with family" but

not detected by Virsorter. (d) Bar plots of the relative abundance of the reads mapping the contigs of a certain host phyla, removing the 218 contigs classified as “Phage with family” but not detected by Virsorter. NA: non-attributable.

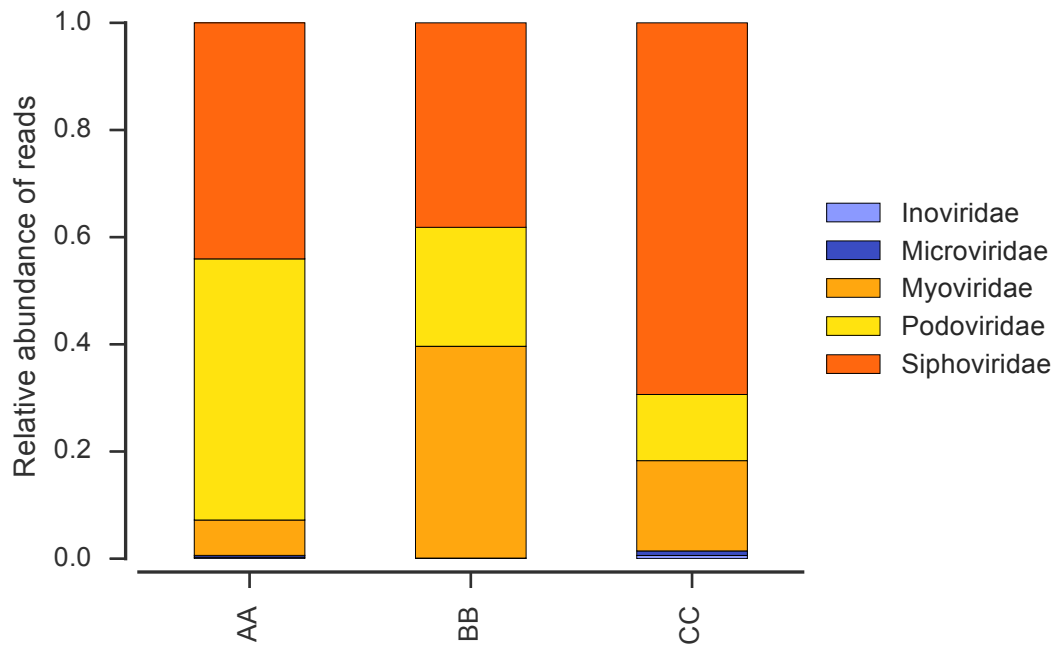


Figure S7: Bar plots of the relative abundance of reads belonging to “phage family” contig categories in the method S without MDA in the three healthy volunteers (AA, BB, CC)

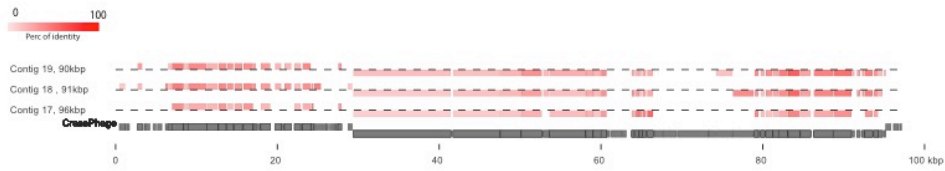


Figure S8: Details of the three contigs of the crAssphage like.

Description: Three contigs of the ABC-Reference-Contig-Catalog are related to the crAssphage, contigs numbers 17, 18 and 19 with respectively a wGRR of 19, 27 and 26. Each red little box represents a homology between a protein of the contig and a protein of the crassphage. The intensity of the red colour is proportional to the percentage of identity between the 2 proteins.

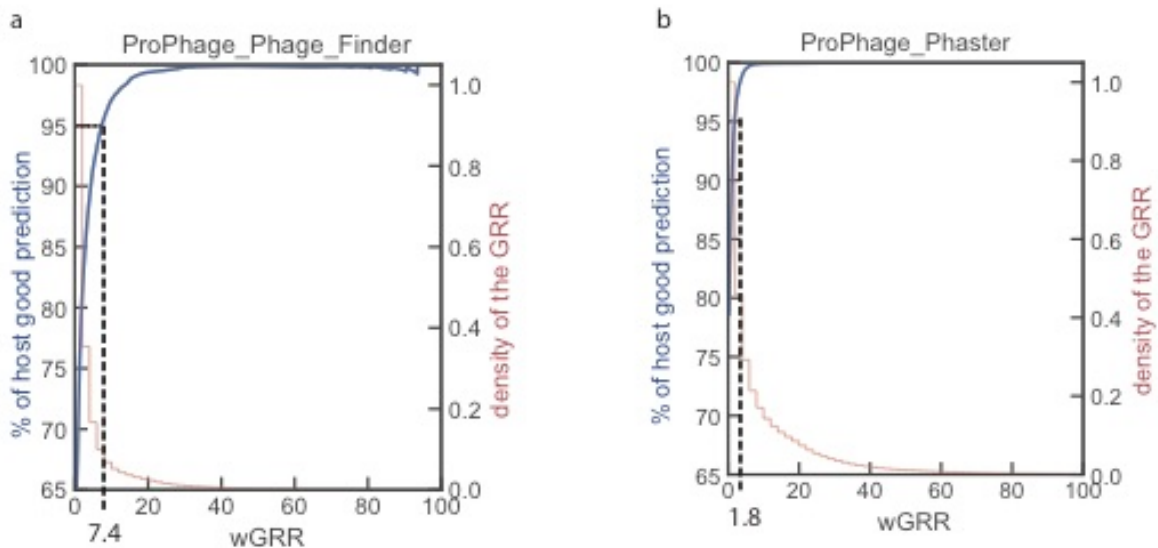


Figure S9: Comparison of the evolution of the percentage of host prediction correct response in terms of wGRR value for the prophage of bacteria Refseq database found with

two different software of prophages prediction: an old one: Phage finder and a new one: Phaster.

Description: (a) We identified 9927 prophages in the chromosomes of complete bacterial genomes from GenBank Refseq using Phage Finder v4.6. Many of these prophages were very similar. We clustered them with a wGRR cut off of 95% and we took a representative of each cluster (6426 phages). In the graph the blue line represents the proportion of host correct response. The red line is the decreasing cumulative density of wGRR. The dashed line represents the wGRR cut-off (7.4) for 95% correct host predictions. (b) We identified 11239 prophages in the chromosomes of complete bacterial genomes from GenBank Refseq using Phaster (online version, only intact prophages were selected). In order to reduce the dataset we clustered them with a wGRR cut off of 95% and we took a representative of each cluster (5786 phages). In the graph the blue line represents the proportion of host correct response. The red line is the decreasing cumulative density of wGRR. The dashed line represents the wGRR cut-off (1.8) for 95% correct host predictions

SUPPLEMENTARY TABLE

Method	Use of MDA	No MDA
Filtration	1-5	
Filtration + CsCl ^a gradient	6-10	11-13 14,15
Filtration + Ultrafiltration	16	17

^a CsCl: Cesium chloride

Table S1: Literature overview of the methods for phage isolation in faeces and the utilisation of the multiple displacement amplification (MDA) process.

Name	Number of reads	Number of contigs	Number of contigs (at least 3 ORFs)	Mean of the contig size (bp)	Median of the contig size (bp)
AA ₀ IV LC	1 183 870	42 584	952	2 789	1 451
AA ₀ IV HC	87 148 950	68 8041	15 656	4 956	1 678

Table S3: Comparing statistic between sample (AA₀) extracted with method IV at a low coverage (LC) and at a high coverage (HC)

Name	Number of reads
IV_HC	87 148 950
IV_90p	78 429 610
IV_80p	69 724 386
IV_70p	61 005 126
IV_60p	52 287 638
IV_50p	43 574 052
IV_40p	34 846 898
IV_30p	26 143 282
IV_20p	17 422 864
IV_10p	8 713 526
IV_1p	870 448
IV_LC	1 183 870

Table S5: Number of reads of the 10 simulated datasets and of the two observed datasets

IV_HC (high coverage) and IV_LC (low coverage)

Database	Information	α	β	γ	R^2
RefSeq Phage	Family	19.33	-0.098	4.49	0.95
	Host	13.64	-0.18	7.34	0.96
	Lifestyle	10.75	-0.23	0.44	0.98
Prophage HMP	Host	28.78	-0.32	13.98	0.95
Prophage RefSeq Bacteria	Host	30.30	-0.22	5.26	0.99

Table S6: Summary of the different values (α , β and γ) of the equation type

$Y = \alpha \times e^{-\beta X} + \gamma$ where Y is the wGRR cut-off and X is the number of proteins of the simulated phage (or of the contig) for each databases and for each information.

The last column " R^2 " represents the fit of the regression line

a

Categories	Number of contigs
No attribution	3489
Putative PHAGE	1122
Bacterial Contamination	914
PHAGE with family	506
Unknown contamination	22
Eukaryota Contamination	2
Archaea Contamination	1

b

Categories	Number of contigs
No attribution	705
Temperate-Confident	603
Temperate-Not-Confident	160
Lytic-Confident	96
Lytic-Not-Confident	54

c

Categories	Number of contigs
Firmicutes	818
No attribution	544
Bacteroidetes	144
Proteobacteria	20
Actinobacteria	27
Fusobacteria	9
Synergistetes	3
Acidobacteria	1
Deinococcus-Thermus	2
Crenarchaeota	1

Table S7: Exploration of the 6,056 contigs found in healthy volunteers (AA, BB, CC) samples

Description: a: Table of description of the different categories of contigs

b: Table of description of the different categories of lifestyle of contigs,

c: Table of description of the different bacteria host (phyla) of contigs.

LEGENDS OF EXCEL SUPPLEMENTARY TABLES

Table S2

File format: XLS file

Title: Tab number 1: HMM profile, from PFAM (30.0) and TIGRFAM (15.), phage specific with a keyword like “capsid, core, portal, protease, replication, tails” and HMM profile bacteria specific. Tab number 2: Details of the HMM profile of the bacterial genes.

Table S4

File format: XLS file

Title: Exploration of different CD-HIT clustering settings on the two dataset: 411 low coverage (LC) dataset and 411 high coverage (HC) dataset.

LIST OF SUPPLEMENTARY REFERENCES

1. Lim, E. S. *et al.* Early life dynamics of the human gut virome and bacterial microbiome in infants. *Nat. Med.* **21**, 1228–1234 (2015).
2. Zuo, T. *et al.* Bacteriophage transfer during faecal microbiota transplantation in Clostridium difficile infection is associated with treatment outcome. *Gut* **67**, 634–643 (2018).
3. Conceição-Neto, N. *et al.* Modular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis. *Sci. Rep.* **5**, 16532 (2015).
4. Reyes, A. *et al.* Gut DNA viromes of Malawian twins discordant for severe acute malnutrition. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 11941–11946 (2015).
5. Kim, K.-H. & Bae, J.-W. Amplification methods bias metagenomic libraries of uncultured single-stranded and double-stranded DNA viruses. *Appl. Environ. Microbiol.* **77**, 7663–7668 (2011).
6. Reyes, A. *et al.* Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* **466**, 334–338 (2010).
7. Minot, S. *et al.* The human gut virome: Inter-individual variation and dynamic response to diet. *Genome Res.* **21**, 1616–1625 (2011).
8. Norman, J. M. *et al.* Disease-Specific Alterations in the Enteric Virome in Inflammatory Bowel Disease. *Cell* **160**, 447–460 (2015).
9. Modi, S. R., Lee, H. H., Spina, C. S. & Collins, J. J. Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome. *Nature* **499**, 219–222 (2013).

10. Abeles, S. R., Ly, M., Santiago-Rodriguez, T. M. & Pride, D. T. Effects of Long Term Antibiotic Therapy on Human Oral and Fecal Viromes. *PLoS One* **10**, e0134941 (2015).
11. Breitbart, M. *et al.* Viral diversity and dynamics in an infant gut. *Res. Microbiol.* **159**, 367–373 (2008).
12. Manrique, P. *et al.* Healthy human gut phageome. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 10400–10405 (2016).
13. Allen, H. K. *et al.* Antibiotics in feed induce prophages in swine fecal microbiomes. *mBio* **2**, (2011).
14. Górska, A. *et al.* Dynamics of the human gut phageome during antibiotic treatment. *Comput. Biol. Chem.* **74**, 420–427 (2018).
15. Johnson, T. A. *et al.* The In-Feed Antibiotic Carbadox Induces Phage Gene Transcription in the Swine Gut Microbiome. *mBio* **8**, (2017).
16. Minot, S. *et al.* Rapid evolution of the human gut virome. *Proc. Natl. Acad. Sci.* **110**, 12450–12455 (2013).
17. Fernández-Orth, D. *et al.* Faecal phageome of healthy individuals: presence of antibiotic resistance genes and variations caused by ciprofloxacin treatment. *J. Antimicrob. Chemother.* **74**, 854–864 (2019).

3.1.3 Principaux résultats de l'article et perspectives

Les principaux résultats de l'étude sont les suivants :

- Cinq méthodes d'isolement des phages ont été testées sur une même selle. Ces méthodes différaient de par la taille des filtres utilisés (0,45µm suivi de 0,22 µm versus 0,45µm seulement) et de la technique de concentration des phages (ultrafiltration, ultracentrifugation suivi d'une ultrafiltration, ultracentrifugation au chlorure de césium, polyéthylène glycol et ultracentrifugation simple). La technique de concentration des phages par le PEG s'est avérée la plus intéressante pour extraire les phages des selles pour les raisons suivantes : (i) méthode de loin la moins coûteuse, (ii) méthode ne nécessitant pas de machine onéreuse comme une ultracentrifugeuse, (iii) méthode produisant une quantité d'ADN de phage satisfaisante ne nécessitant pas d'étape d'amplification, contrairement au chlorure de césium, (iv) les séquences obtenues par cette méthode ont montré une grande quantité et diversité de contigs de phage et (v) la proportion de contamination par des reads de séquences bactériennes est très faible (< 2%).
- Sur l'ADN extrait par la méthode au polyéthylène glycol, deux profondeurs de séquençage ont été réalisées afin d'évaluer l'impact de ce paramètre sur la diversité des contigs de phage retrouvés. En augmentant 74 fois la profondeur de séquençage, 8,5 fois plus de contigs phagique ont été retrouvés. Différentes profondeurs de séquençage ont été simulés *in silico*, il est observé une diminution du nombre de nouveaux contigs phagiques avec l'augmentation de la profondeur. Il semble y avoir un plateau à 20 000 000 de reads, à partir duquel la grande majorité des nouveaux contigs ne sont pas classés comme des phages.
- La méthode d'extraction au polyéthylène glycol montre une bonne reproductibilité avec, en moyenne, une distance de Bray-Curtis entre deux

réplicats de 0,08. Les conséquences d'une étape d'amplification non spécifique de l'ADN phagique avant le séquençage ont été étudiées dans ce travail. Il est montré que plus le temps d'amplification est long, plus la diversité de contig de phage est moindre (avec une grande proportion de contig de phages à ADN simple brin) et plus on perd en reproductibilité.

- Une approche bio-informatique basée sur le wGRR : « weighted gene repertoire relatedness », score sur les similarités de séquences entre les contigs et les prophages ou phages de trois bases de données, a permis d'attribuer à certain contig une famille de phage précise, un hôte (au niveau du phyla) et un style de vie (tempéré, lytique). Cette approche prend en compte la taille du contig, paramètre important dans le cas de données de métagénomique où les contigs sont majoritairement petits. La plupart (57 %) des contigs ne correspond à aucune des bases de données mais en abondance relative les contigs étiquetés phage et phage putatif sont les plus abondants. Un homologue du crAssphage a été retrouvé chez les 3 volontaires sains testés.

Ce travail méthodologie nous a permis d'isoler et d'analyser les phages du microbiote intestinal. La question de l'impact des antibiotiques sur le phageome intestinal est encore très peu documentée et a fait l'objet de la suite de ce travail de thèse.

3.2 Etude clinique d'impact de la céfotaxime et du ceftriaxone sur le microbiote intestinal chez des volontaires sains : résultats bactériologiques

3.2.1 Présentation du contexte de l'étude clinique

L'augmentation de la résistance bactérienne aux antibiotiques et sa diffusion aussi bien à l'hôpital que dans la communauté constitue une menace pour l'homme. Une des raisons de cette évolution est le mésusage des antibiotiques aussi bien chez l'homme que chez l'animal. En effet, les antibiotiques favorisent l'émergence de bactéries résistantes dans le microbiote intestinal. La bonne prescription antibiotique caractérisée par (i) un spectre le plus étroit sur le germe ciblé (ii) un moindre impact sur le microbiote intestinal, est une priorité. Dans ce contexte, plus d'études sur l'impact des différents antibiotiques sur le microbiote intestinal sont nécessaires.

Ceftriaxone et cefotaxime sont deux céphalosporines de troisième génération administrées par voie injectable avec le même spectre d'activité mais des modes d'élimination différents : la ceftriaxone est éliminée à 60% par les reins et à 40% par le foie (excrétion biliaire) ¹⁷⁸ alors que le cefotaxime est éliminé à 90% par les reins et 10% par le foie ¹⁷⁹. Dans deux études ^{180,181} randomisées contrôlées, il est suggéré que la ceftriaxone, dûe à son mode d'élimination biliaire, a un impact plus prononcé que le cefotaxime sur la sélection de bacilles à Gram négatif résistants aux CIIG sur le microbiote intestinal. Dans ces études, les 2 molécules sont utilisées à la même dose (2 grammes par jour). Or, en pratique clinique, les deux molécules ne sont pas administrées à la même posologie en raison de leur demi-vie très différente : le cefotaxime a une demi-vie de 1h et se donne 3 fois par jour (3 grammes par 24h) alors que la ceftriaxone a une demi-vie de 7h et se donne une fois par jour (1gramme par 24h). Aucune étude avec ces schémas d'administration n'est disponible.

L'objectif de cette étude est de comparer l'impact de l'élimination biliaire de la ceftriaxone (40%) et du cefotaxime (10%) sur l'émergence de la résistance dans le microbiote intestinal en respectant les rythmes d'administration usuels.

C'est une étude prospective, ouverte, randomisée chez des volontaires sains. Il y a eu 22 volontaires inclus (11 dans chaque bras : ceftriaxone versus cefotaxime). Le

traitement était d'une durée de 3 jours par voie intraveineuse, soit par ceftriaxone (1 gramme/jour) ou céfotaxime (1 gramme/8 heures). Les selles ont été recueillies avant (J-14, J-7, J-1) et après (J1, J2, J3, J4, J7, J10, J15, J30, J90 et J180) la première administration pour des analyses bactériologiques (phénotypique et 16S) et les dosages pharmacocinétiques à l'équilibre. Des échantillons de plasma ont été prélevés au jour 3 également pour des analyses pharmacocinétiques.

3.2.2 Article publié

Antimicrobial Agents and Chemotherapy, Mai 2019-10-18

Cefotaxime and ceftriaxone Have Similar Effects on the Intestinal Microbiota in Human Volunteers Treated by Standard-Dose Regimens

Charles Burdet (1, 2), Nathalie Grall (1, 3), Morgane Linard (2), Antoine Bridier-Namias (1), Michèle Benhayoun (4), Khadija Bourabha (1), Mélanie Magnan (1), Olivier Clermont (1), **Camille d'Humières** (1, 3), Olivier Tenaillon (1), Erick Denamur (1,5), Laurent Massias (1,6), Sarah Tubiana (4), Loubna Alavoine (4), Antoine Andremont (1), France Mentré (1, 2), Xavier Duval (1, 4) for the CEREMI group

(1) INSERM, IAME, UMR 1137, Paris, France; Université Paris Diderot, Sorbonne Paris Cité, Paris, France

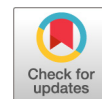
(2) AP-HP, Department of Epidemiology, Biostatistic and Clinical Research, Bichat Hospital, Paris, France

(3) AP-HP, Department of Bacteriology, Bichat Hospital, Paris, France

(4) INSERM Clinical Investigation Center 1425, Paris, France

(5) AP-HP, Laboratory of Genetics, Bichat Hospital, Paris, France

(6) AP-HP, Laboratory of Toxicology, Bichat Hospital, Paris, France



Ceftriaxone and Cefotaxime Have Similar Effects on the Intestinal Microbiota in Human Volunteers Treated by Standard-Dose Regimens

Charles Burdet,^{a,b,c} Nathalie Grall,^{a,b,d} Morgane Linard,^c  Antoine Bridier-Nahmias,^{a,b} Michèle Benhayoun,^e Khadija Bourabha,^{a,b} Mélanie Magnan,^{a,b} Olivier Clermont,^{a,b} Camille d'Humières,^{a,b,d} Olivier Tenailon,^{a,b} Erick Denamur,^{a,b,f} Laurent Massias,^{a,b,g} Sarah Tubiana,^e Loubna Alavoine,^e Antoine Andremont,^{a,b} France Mentré,^{a,b,c} Xavier Duval,^{a,b,e} for the CEREMI Group

^aINSERM, IAME, UMR 1137, Paris, France

^bUniversité Paris Diderot, Sorbonne Paris Cité, Paris, France

^cAP-HP, Department of Epidemiology, Biostatistic and Clinical Research, Bichat Hospital, Paris, France

^dAP-HP, Department of Bacteriology, Bichat Hospital, Paris, France

^eINSERM Clinical Investigation Center 1425, Paris, France

^fAP-HP, Laboratory of Genetics, Bichat Hospital, Paris, France

^gAP-HP, Laboratory of Toxicology, Bichat Hospital, Paris, France

ABSTRACT Ceftriaxone has a higher biliary elimination than cefotaxime (40% versus 10%), which may result in a more pronounced impact on the intestinal microbiota. We performed a monocenter, randomized open-label clinical trial in 22 healthy volunteers treated by intravenous ceftriaxone (1 g/24 h) or cefotaxime (1 g/8 h) for 3 days. We collected fecal samples for phenotypic analyses, 16S rRNA gene profiling, and measurement of the antibiotic concentration and compared the groups for the evolution of microbial counts and indices of bacterial diversity over time. Plasma samples were drawn at day 3 for pharmacokinetic analysis. The emergence of 3rd-generation-cephalosporin-resistant Gram-negative enteric bacilli (*Enterobacteriales*), *Enterococcus* spp., or noncommensal microorganisms was not significantly different between the groups. Both antibiotics reduced the counts of total Gram-negative enteric bacilli and decreased the bacterial diversity, but the differences between the groups were not significant. All but one volunteer from each group exhibited undetectable levels of antibiotic in feces. Plasma pharmacokinetic endpoints were not correlated to alteration of the bacterial diversity of the gut. Both antibiotics markedly impacted the intestinal microbiota, but no significant differences were detected when standard clinical doses were administered for 3 days. This might be related to the similar daily amounts of antibiotics excreted through the bile using a clinical regimen. (This study has been registered at ClinicalTrials.gov under identifier NCT02659033.)

KEYWORDS cefotaxime, ceftriaxone, intestinal microbiota, metagenomics, pharmacokinetics

The rise in bacterial resistance to antibiotics, particularly that of Gram-negative bacteria, constitutes a serious threat to our medical system (1). Initially restrained to nosocomial infections, antibiotic resistance has spread worldwide (2), affecting treatment of community-acquired infections. One of the drivers of this evolution is the impact that antibiotics, in particular, beta-lactams, exert on the intestinal commensal microbiota of both humans and animals (3).

In their studies of healthy volunteers receiving 5-day courses of ciprofloxacin, Dethlefsen and colleagues observed that antibiotic administration had a rapid effect in

Citation Burdet C, Grall N, Linard M, Bridier-Nahmias A, Benhayoun M, Bourabha K, Magnan M, Clermont O, d'Humières C, Tenailon O, Denamur E, Massias L, Tubiana S, Alavoine L, Andremont A, Mentré F, Duval X, for the CEREMI Group. 2019. Ceftriaxone and cefotaxime have similar effects on the intestinal microbiota in human volunteers treated by standard-dose regimens. *Antimicrob Agents Chemother* 63:e02244-18. <https://doi.org/10.1128/AAC.02244-18>.

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

Address correspondence to Charles Burdet, charles.burdet@inserm.fr.

Received 23 October 2018

Returned for modification 23 November 2018

Accepted 20 March 2019

Accepted manuscript posted online 1 April 2019

Published 23 May 2019

reducing bacterial diversity and in modifying the gut microbiome composition (4, 5). Similar observations were reported with moxifloxacin (6). These results are a strong incentive to better document the impact of antibiotics on the gut microbiome (7) and to identify the drivers of its disruption. Indeed, the gut microbiome has been shown to contribute to health maintenance of its host (8).

A limited number of studies have suggested that taking into account antibiotic pharmacokinetic characteristics might help to reduce the impacts of antibiotics on the microbiome. The two 3rd-generation cephalosporins ceftriaxone and cefotaxime share their antibacterial spectra, indications, and intravenous administration route, but they have different pharmacokinetic characteristics: ceftriaxone is administered once daily and has a 40% biliary elimination, while cefotaxime needs thrice daily administration but has a 10% biliary elimination. In two randomized controlled trials, female patients requiring gynecological surgery received a single high-dose (2 g of ceftriaxone or 2 g of cefotaxime) antibiotic prophylactic treatment (9, 10). Both studies suggested a more pronounced impact of ceftriaxone on the gut microbiota in terms of the selection of Gram-negative enteric bacillus populations resistant to 3rd-generation cephalosporins. Similar trends were reported in observational studies (11, 12).

However, no study dealing with usual antibiotic regimens administered in clinical practice is available. Indeed, cefotaxime is usually administered at higher daily doses than ceftriaxone, which may ultimately lead to quite similar effects on the gut microbiota. Moreover, next-generation sequencing methods were not used to investigate this question but may provide a precise description of the microbiome.

Here, we report the results of a prospective randomized clinical trial performed in healthy volunteers receiving a clinical course of ceftriaxone or cefotaxime for 3 days for the comparison of their impact on the gut microbiota, determined using phenotypic analysis and 16S rRNA gene profiling.

RESULTS

Subjects. Thirty-three subjects were screened, and 22 subjects were randomized (11 in each treatment group); all participated in the trial until the end of follow-up. Their baseline characteristics are presented in Table 1. No serious adverse event was reported, but treatment was stopped early in one subject in the cefotaxime group who presented a vasovagal malaise while the intravenous line was being placed for the 8th infusion.

Ceftriaxone and cefotaxime pharmacokinetics. For both antibiotics, plasma pharmacokinetics were best described by a 2-compartment model with first-order elimination from the central compartment. All population parameters could be estimated with a reasonable precision (see Table S1 in the supplemental material), and the goodness of fit was satisfactory (Fig. S1 and S2). Derived pharmacokinetic endpoints estimated for the 11 subjects included in each treatment group are presented in Table 2.

All subjects exhibited undetectable concentrations of antibiotics in feces between day -1 and day 30, except for one subject in each treatment group. Subject 16 had detectable concentrations of ceftriaxone in feces between day 2 and day 7. The concentrations ranged from 5.0 $\mu\text{g/g}$ to 93.7 $\mu\text{g/g}$, with the maximal value being measured at day 4. Subject 3 had detectable concentrations of cefotaxime in feces at day 4 (1.6 $\mu\text{g/g}$). These 2 subjects had derived pharmacokinetic endpoints close to the median values observed for their respective treatment groups.

Phenotypic analysis of the fecal samples. Overall, 283 fecal samples were available for phenotypic analyses. The baseline counts of the microorganisms studied are presented in Table 1.

(i) Third-generation-cephalosporin-resistant Gram-negative enteric bacilli. No major change in the counts of resistant Gram-negative enteric bacilli was observed in either of the two treatment groups over time (Fig. 1 and Fig. S3A and B), and no significant difference was observed between groups (Table 3). The molecular characteristics of the detected resistant strains are presented in Table S2.

TABLE 1 Baseline characteristics of the 22 included subjects^a

Characteristic	Value for subjects treated with:	
	Ceftriaxone (n = 11)	Cefotaxime (n = 11)
Clinical characteristics		
No. (%) of men	2 (18)	4 (36)
Median (min, max) age (yr)	30 (24, 55)	24 (18, 61)
Median (min, max) total body wt (kg)	72.0 (57.9, 78.4)	63.4 (49.9, 87.0)
Median (min, max) body mass index (kg/m ²)	24.6 (20.0, 28.7)	22.0 (18.6, 28.9)
Median (min, max) count of microorganisms in feces at baseline (log ₁₀ CFU/g)		
Resistant enterobacteria (AES agar)	<2.0 (<2.0, 5.4)	<2.0 (<2.0, 2.8)
Resistant enterobacteria (ChromID ESB agar)	<2.0 (<2.0, 5.7)	<2.0 (<2.0, <2.0)
Total enterobacteria	7.7 (5.8, 8.7)	7.8 (3.3, 9.1)
<i>Enterococcus</i> spp.	6.2 (3.4, 9.2)	6.7 (3.3, 7.8)
<i>S. aureus</i>	<2.0 (<2.0, 4.2)	<2.0 (<2.0, 3.0)
<i>P. aeruginosa</i>	<2.0 (<2.0, 5.1)	<2.0 (<2.0, 2.7)
<i>C. difficile</i>	<2.0 (<2.0, <2.0)	<2.0 (<2.0, <2.0)
Yeasts	<2.0 (<2.0, 2.7)	<2.0 (<2.0, 2.4)
Median (min, max) index of α -diversity at day -1		
Shannon index	3.9 (3.2, 4.3)	4.0 (2.8, 4.3)
No. of OTUs	451 (308, 591)	451 (267, 521)
Median (min, max) index of β -diversity between day -1 and day -15		
Bray-Curtis dissimilarity	0.35 (0.23, 0.51)	0.29 (0.17, 0.52)
Unweighted UniFrac distance	0.41 (0.32, 0.46)	0.38 (0.29, 0.49)
Median (min, max) relative abundance (%) of the main bacterial phyla at day -1		
<i>Actinobacteria</i>	4.7 (1.3, 13.5)	4.3 (2.1, 10.0)
<i>Bacteroidetes</i>	34.1 (23.2, 52.2)	40.2 (24.2, 60.4)
<i>Firmicutes</i>	55.7 (43.9, 69.1)	54.1 (32.7, 68.2)
<i>Proteobacteria</i>	0.9 (0.3, 1.6)	0.6 (0.2, 7.4)
Other phyla	1.2 (0.3, 2.6)	1.4 (0.7, 3.2)

^aBaseline counts of microorganisms were computed as the arithmetic mean of the log₁₀ counts observed for the available pretreatment samples at days -15, -7, and -1. min, minimum; max, maximum.

At baseline, colonization with resistant Gram-negative enteric bacilli was detected in 5 subjects (45.5%) and 1 subject (9.7%) in the ceftriaxone and cefotaxime groups, respectively. Among the uncolonized subjects, 1 (16.7%; 95% confidence interval [CI], 0.4%, 64.1%) from the ceftriaxone group and 2 (20%; 95% CI, 2.5%, 55.6%) from the cefotaxime group acquired resistant Gram-negative enteric bacilli between the beginning of treatment and day 15 ($P > 0.99$) (Table 4).

(ii) Other microorganisms studied. No significant difference in the variation from baseline of any of the other microorganisms studied was observed between the groups (Table 3; Fig. 1; Fig. S3C to H). Both antibiotics profoundly reduced the counts of total Gram-negative enteric bacilli, with the highest median reduction from baseline being 4.4 log₁₀ CFU/g (2.4, 8.7 log₁₀ CFU/g [minimum, maximum]) in the ceftriaxone group and 3.8 (1.9, 8.1 log₁₀ CFU/g [minimum, maximum]) in the cefotaxime group (Fig. 1 and Fig. S3C). At day 15, the counts of Gram-negative enteric bacilli returned to their baseline value. Carriage of toxigenic *Clostridium difficile* was detected in 3 patients, 1 (at day 10) in the ceftriaxone group and 2 (1 at day 4 and 1 at day 15) in the cefotaxime group ($P > 0.99$). Results regarding the emergence of other noncommensal microorganisms of the intestinal tract are presented in Table 4.

TABLE 2 Derived pharmacokinetic endpoints of ceftriaxone and cefotaxime for the 11 subjects in each treatment group derived from the estimated individual pharmacokinetic parameters^a

Treatment	fAUC _{0-24,ss} (mg · h/liter)	fC _{max} (mg/liter)	fC _{min} (mg/liter)	fT > 1 (%)
Ceftriaxone (n = 11)	83.8 (59.0, 105.3)	13.5 (10.8, 14.9)	0.9 (0.4, 1.5)	94.2 (68.6, 100)
Cefotaxime (n = 11)	121.5 (83.6, 187.0)	40.6 (29.4, 55.9)	0.1 (0.1, 0.2)	53.0 (44.5, 66.7)

^aThe data represent the median (minimum, maximum).

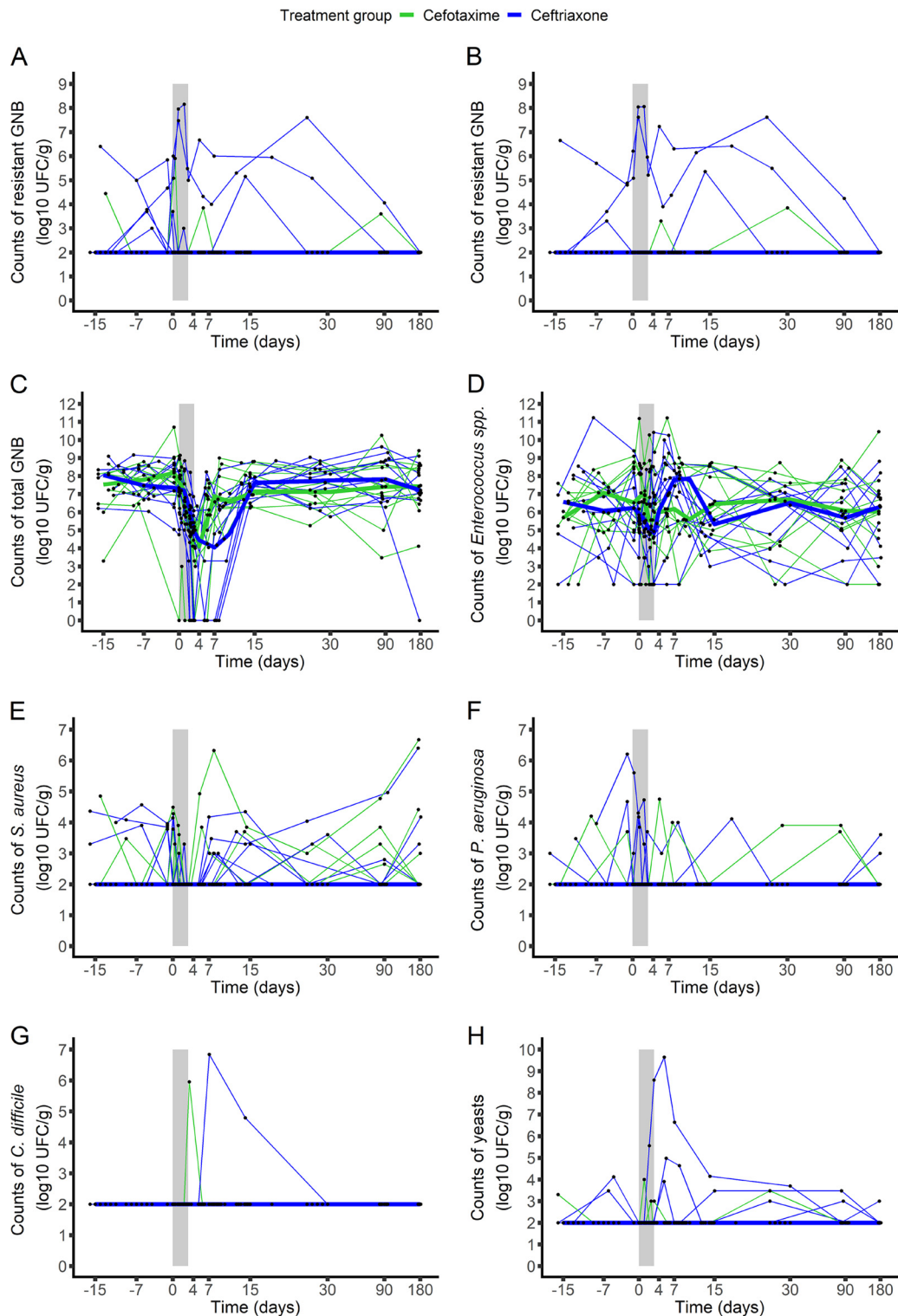


FIG 1 Evolution of the counts of the studied microorganisms in the fecal samples from the 22 included subjects treated with ceftriaxone ($n = 11$, blue) or cefotaxime ($n = 11$, green). The studied microorganisms included 3rd-generation-cephalosporin-resistant Gram-negative enteric bacilli (measured on AES plates [A] or ChromID ESBL plates [B]), total Gram-negative enteric bacilli (C), *Enterococcus* spp. (D), *S. aureus* (E), *P. aeruginosa* (F), *C. difficile* (G), and yeasts (H). The light gray zone represents the treatment period. Thin lines represent individual values, and thick lines represent the median change from the baseline of the log₁₀ counts at each time. UFC, number of colony-forming units; GNB, Gram-negative enteric bacilli.

TABLE 3 AUC of variation from baseline of the counts of the studied microorganisms between baseline and day 7 or day 15 and of bacterial diversity indices between day -1 and day 30 in fecal samples from the 22 subjects^a

Parameter	Value for subjects treated with:		P value ^b
	Ceftriaxone (n = 11)	Cefotaxime (n = 11)	
AUC _{D0D7} of raw bacterial counts (no. of log ₁₀ CFU · day/g)			
Resistant enterobacteria (AES agar)	0.0 (0.0, 23.4)	0.0 (0.0, 2.7)	0.40
Resistant enterobacteria (ChromID ESBL agar)	0.0 (0.0, 24.3)	0.0 (0.0, 1.5)	0.28
AUC _{D0D7} of count change from baseline (no. of log ₁₀ CFU · day/g)			
Total enterobacteria	-14.8 (-30.0, -7.0)	-11.2 (-26.0, -6.2)	0.26
<i>Enterococcus</i> spp.	-0.3 (-8.6, 2.6)	-0.9 (-3.5, 4.1)	0.56
<i>S. aureus</i>	0.0 (-9.7, 1.3)	0.0 (-2.7, 1.7)	0.23
<i>P. aeruginosa</i>	0.0 (-8.4, 5.5)	0.0 (-4.0, 2.9)	0.87
<i>C. difficile</i>	0.0 (0.0, 0.0)	0.0 (0.0, 7.0)	0.36
Yeasts	0.0 (-3.9, 18.4)	0.0 (-2.5, 2.4)	0.85
AUC _{D0D15} of count change from baseline (no. of log ₁₀ CFU · day/g)			
Resistant enterobacteria (AES agar)	0.0 (-50.7, 40.3)	0.0 (-12.2, 4.9)	>0.99
Resistant enterobacteria (ChromID ESBL agar)	0.0 (-6.5, 44.6)	0.0 (0.0, 6.0)	0.79
Total enterobacteria	-30.7 (-79.2, -12.4)	-21.8 (-37.5, 16.4)	0.088
<i>Enterococcus</i> spp.	7.6 (-38.1, 39.5)	6.2 (-31.5, 19.3)	0.75
<i>S. aureus</i>	1.2 (-18.5, 6.6)	0.0 (-7.3, 21.2)	0.82
<i>P. aeruginosa</i>	0.0 (-25.5, 13.6)	0.0 (-8.4, 7.0)	0.69
<i>C. difficile</i>	0.0 (0.0, 33.7)	0.0 (0.0, 7.5)	>0.99
Yeasts	0.0 (-10.5, 53.7)	0.0 (-6.5, 2.5)	0.46
AUC _{D-1D30} of change from baseline of diversity indices (unit · day)			
Shannon diversity index	-12.0 (-26.6, 2.4)	-10.5 (-21.6, 15.1)	0.85
No. of OTUs	-1,391 (-5,708, 114)	-1,816 (-2,677, 907)	>0.99
Bray-Curtis dissimilarity	12.2 (10.3, 19.1)	12.2 (10.7, 17.6)	0.95
Unweighted UniFrac distance	12.1 (10.0, 18.7)	12.5 (10.9, 16.9)	0.75

^aBaseline counts were computed as the arithmetic mean of the log₁₀ counts observed for pretreatment samples. AUC_{D0D15} and AUC_{D-1D30} were normalized by the actual time observed between baseline and day 15 and baseline and day 30, respectively. Data are presented as the median (minimum, maximum).

^bP values refer to the results of the nonparametric Wilcoxon test.

16S rRNA gene profiling of the fecal samples. 16S rRNA gene profiling was performed on 148 available samples. The baseline characteristics taxonomic composition and diversity indices are presented in Table 1.

(i) Taxonomic composition. Both antibiotics reduced the relative abundance of the *Firmicutes* (median maximal change [minimum, maximum], -9.3% [-47.6%, 2.2%] in the ceftriaxone group and -12.3% [-22.4%, -4.0%] in the cefotaxime group), of the *Actinobacteria* (median maximal change [minimum, maximum], -3.6% [-11.0%, 0.0%] in the ceftriaxone group and -2.3% [-8.9%, -0.1%] in the cefotaxime group), and of *Bacteroidetes* (median maximal change [minimum, maximum], -2.2% [-31.4%, 10.8%] in the ceftriaxone group and -4.0% [-23.8%, 15.2%] in the cefotaxime group). The relative abundance of the *Proteobacteria* remained roughly unchanged (median max-

TABLE 4 Emergence of 3rd-generation-cephalosporin-resistant Gram-negative enteric bacilli or noncommensal microorganisms of the intestinal tract between day 1 and day 15 in the 22 included subjects^a

Microorganism	No. (%) of subjects treated with:				P value ^b
	Ceftriaxone		Cefotaxime		
	Subjects at risk	Subjects colonized	Subjects at risk	Subjects colonized	
Resistant Gram-negative enteric bacilli	6	1 (16.7)	10	2 (20)	>0.99
<i>S. aureus</i>	7	6 (85.7)	8	3 (37.5)	0.12
<i>P. aeruginosa</i>	8	1 (12.5)	9	1 (11.1)	>0.99
<i>C. difficile</i>	11	1 (9.1)	11	1 (9.1)	>0.99
Yeasts	9	3 (33.3)	10	3 (30)	>0.99

^aOnly subjects who were not colonized before the beginning of treatment were included in the analysis. For resistant Gram-negative enteric bacilli, results from either AES or ChromID ESBL agar plates were considered for definition of the emergence of resistance.

^bP values refer to the result of a nonparametric Fisher's exact test.

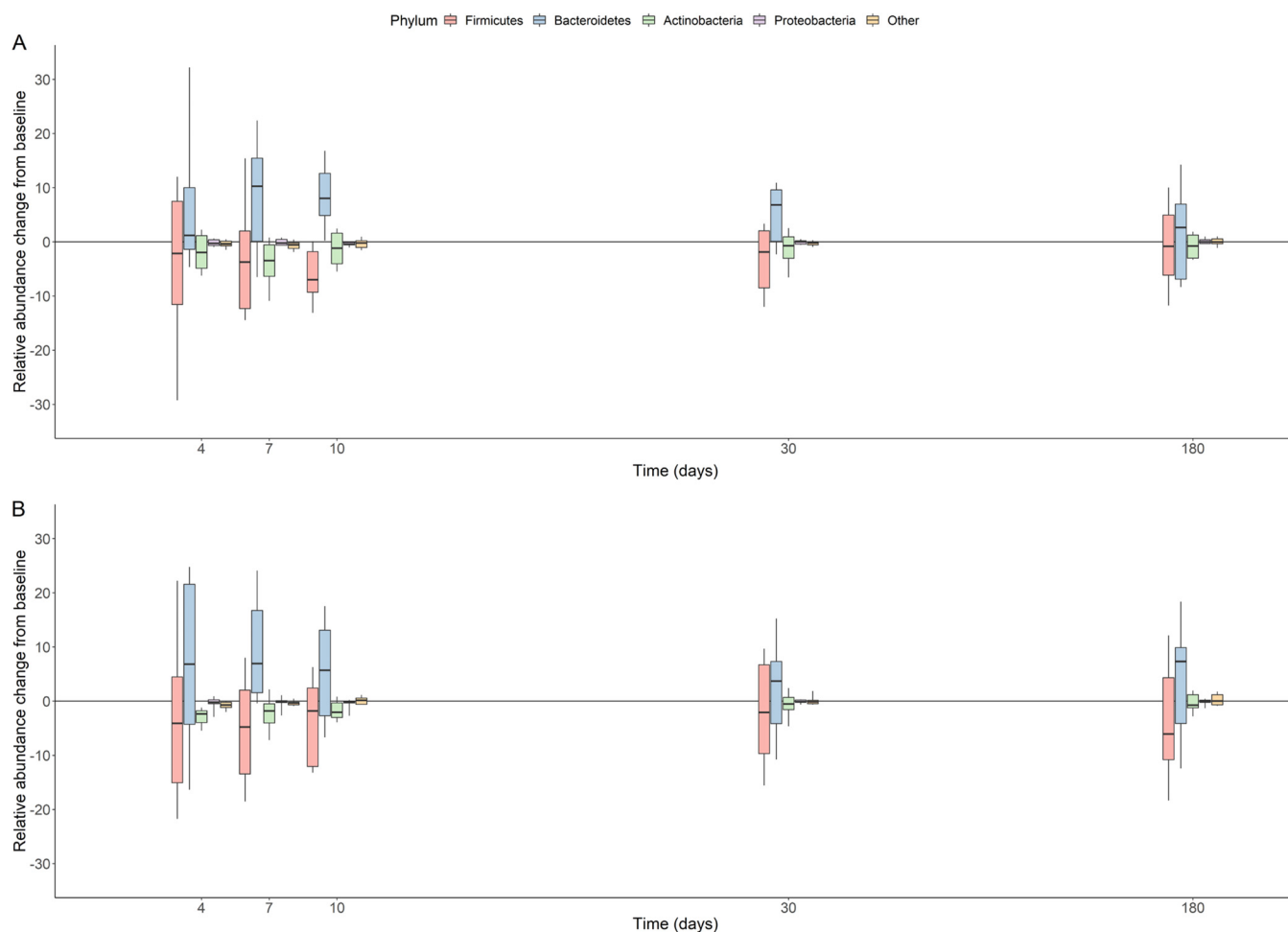


FIG 2 Box plots of the change from baseline of relative abundances (in percent) of the main bacterial phyla at day 4, day 7, day 10, day 30, and day 180 in the 22 included subjects treated with ceftriaxone ($n = 11$) (A) or cefotaxime ($n = 11$) (B). The boxes present the 25th and 75th percentiles, and the horizontal black bars report the median value, while the whiskers report the 10th and 90th percentiles.

imal change [minimum, maximum], -0.5% [-1.6% , 0.0%] in the ceftriaxone group and 0.0% [-5.7% , 0.2%] in the cefotaxime group) (Fig. 2).

(ii) Bacterial diversity. Although no significant difference was observed between ceftriaxone and cefotaxime (Table 3), both antibiotics exhibited a profound impact on bacterial diversity. The highest reduction of the Shannon index and the number of operational taxonomic units (OTUs) (minimum, maximum) from baseline was 0.8 Shannon unit (0.3, 2.9 Shannon units) and 100 OTUs (40, 268 OTUs) in the ceftriaxone group and 1.1 Shannon units (0.2, 1.8 Shannon units) and 144 OTUs (97, 234 OTUs) in the cefotaxime group (Fig. 3 and Fig. S4).

At day -1 , interindividual UniFrac distances appeared to be quite homogeneous, except for subject 3 (from the cefotaxime group), whose distance from the other subjects was notably higher (Fig. S5). His distance to the others increased at day 4, whereas other interindividual distances remained roughly unchanged. At day 7, subject 3 and subject 16 (from the ceftriaxone group) were very distant from the other subjects. These 2 subjects were those who exhibited detectable concentrations of antibiotics in feces. They also exhibited the highest changes of diversity over time (Fig. 3). Similar patterns were observed for the Bray-Curtis dissimilarity (Fig. S6).

At day 30, the α -diversity indices returned to their baseline values, while the β -diversity indices between day -1 and day 30 were similar to the values observed between day -15 and day -1 in both groups (Fig. 3 and Fig. S7), attesting to the temporal intraindividual variability of the microbiome composition.

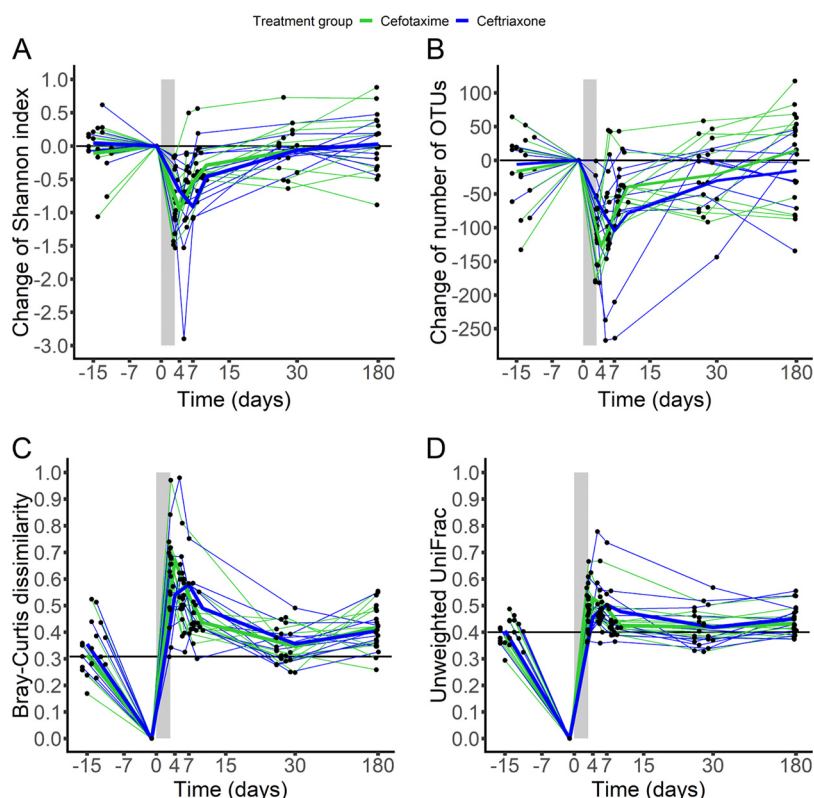


FIG 3 Evolution of the change from baseline of the Shannon index (A), change from baseline of the number of OTUs (B), Bray-Curtis dissimilarity from baseline (C), and unweighted UniFrac distance from baseline (D) in fecal samples from the 22 included subjects treated with ceftriaxone ($n = 11$, blue) or cefotaxime ($n = 11$, green). The light gray zone represents the treatment period. Thin lines represent individual values, and thick lines represent the mean of the index values at each time. In panels C and D, horizontal black lines represent the median value of the β -diversity index for each subject between samples collected at day -15 and samples collected at day -1.

Pharmacokinetic endpoints were not significantly correlated with any index of bacterial diversity (Table S3).

DISCUSSION

Data from preclinical and clinical trials reported differences in ceftriaxone and cefotaxime pharmacokinetic characteristics, in particular, differences in their biliary elimination. The proportion of administered ceftriaxone excreted by bile after each dose has been estimated to be approximately 40% (13, 14), which is 4 times higher than that of cefotaxime (15, 16). In their animal study, van Ogtrop and colleagues treated mice with either cefoperazone, ceftriaxone, cefepime, or ceftazidime, whose intestinal elimination ranges from 0.2% to 37% of the administered dose (17). Although all 4 cephalosporins exhibited similar effects on Gram-negative bacteria, some differences were observed: when higher doses were administered, the impact of the antibiotics, in particular, impacts regarding colonization resistance, increased with the fractional intestinal elimination. However, differences in the antibacterial spectrum and the dosing regimens of the studied antibiotics in these animals might be responsible for some of these differences. In the CEREMI trial, we compared the impacts of two 3rd-generation cephalosporins, whose antibacterial spectrum has previously been reported to be similar, at their clinical doses to analyze specifically the impact of pharmacokinetic differences (18).

Interestingly, the fecal concentrations of both antibiotics were very low in the subjects included in the study. It was reported previously that the ceftriaxone excreted through the bile is microbiologically inactivated (19, 20). This might be due to the

presence of endogenous beta-lactamases of microbial origin. Several authors reported that beta-lactam antibiotics can be inactivated by the enzymatic activity of the gut microbiota (21, 22). In particular, using a semiquantitative test for measuring the beta-lactamase activity in the fecal content, Leonard et al. observed that 2 of 6 volunteers treated by ceftriaxone exhibited undetectable levels of antibiotic when fecal beta-lactamase activity was detectable (23). This has been further illustrated by the administration of an oral beta-lactamase, which was shown to reduce the fecal concentrations of ceftriaxone (24). This beta-lactamase activity of commensal bacteria might protect against the selection of resistant Gram-negative enteric bacillus populations (25, 26).

Despite these low antibiotic concentrations, the intestinal microbiota was deeply affected by the 3-day course of both antibiotics, although no significant difference was found between the groups. It is probable that antibiotics exert their effect on the upper part of the intestine, whereas we could measure the concentrations of antibiotics only in the fecal contents. In addition, a majority of intestinal *Bacteroides* bacteria might not be affected by administered antibiotics and replace suppressed Gram-negative enteric bacillus populations. This would thereby leave no room for the selection of resistant Gram-negative enteric bacillus populations. To our knowledge, however, no data supporting this hypothesis are available.

The absence of a difference regarding the selection of phenotypic resistance among Gram-negative enteric bacilli might be explained by the design of the trial, which was performed in healthy volunteers. This community setting probably results in a low selective pressure, thereby limiting the risk of colonization with resistant bacterial strains. The exposure to the health care system of patients requiring intravenous antibiotic therapy might be an additional factor contributing to the differences observed in existing ecological studies (11, 12).

More startling was the similarity of both antibiotics' impact on the microbiome. This observation might be further explained by considering the dose of each antibiotic administered. The total daily dose of cefotaxime is indeed 3 times higher than that of ceftriaxone. Using known data on biliary elimination of the 2 antibiotics, we can infer that the daily amount of antibiotic reaching the gut is roughly similar for both antibiotics (400 mg for ceftriaxone at 1 g/day and 300 mg for cefotaxime at 1 g/8 h). It is noteworthy that despite being administered at a lower daily dose, the pharmacodynamic profile of ceftriaxone, as measured by the fraction of the time during which the free plasma concentration of cefotaxime is above the susceptibility breakpoint of *Enterobacteriaceae*, is more favorable than that of cefotaxime, with a similar impact on the intestinal microbiota. This suggests that ceftriaxone might be better suited to reduce the overall consumption of antibiotics and thereby the environmental spread of antibiotic residues.

This work has some limitations. First, we included only healthy volunteers. Our results thus might not reflect the global effect of both antibiotics on the intestinal microbiota when treating infected patients. Here, we aimed at analyzing the intrinsic impact of ceftriaxone and cefotaxime, avoiding confounding factors, such as repeated antibiotic treatments or frequent contact with the health care system. Another limitation is the absence of data regarding beta-lactamase activity in feces. This would have provided further details on the results regarding the fecal concentrations of antibiotics. This activity might also be used as a proxy for the exposure of the intestinal microbiota to beta-lactam antibiotics. Of note, the prevalence of Gram-negative bacilli was somewhat lower than expected (27). It might have been underestimated as we chose to use a selective agar (Drigalski) in order to quantify total enterobacteria from fecal microbiota in a reproducible way, which is difficult without using selective media. Furthermore, analyses were performed on the same medium for all patients, which allowed us to avoid any differential bias.

In spite of these limitations, this trial provides further insights into the impacts of ceftriaxone and cefotaxime on the intestinal microbiota and allowed us to compare those impacts. Under our experimental conditions, our results suggest that both

antibiotics exert a pronounced impact on the intestinal microbiota, extending what is already known on their antibacterial spectrum for cultivable bacteria. This was observed while the total daily dose of cefotaxime administered was 3 times higher than that of ceftriaxone, as requested in clinical practice. Deeper analysis of the resistance within the microbiota, such as the use of shotgun metagenomics for studying the fecal content of resistance-conferring genes or the use of the recently released targeted sequence capture ResCap (28), would be of great value in order to draw a complete picture of the possible difference in selective pressure between these 2 antibiotics.

MATERIALS AND METHODS

Study design. We conducted a prospective, open-label, randomized clinical trial (registered at ClinicalTrials.gov under identifier NCT02659033) from March 2016 to August 2017 in adult healthy volunteers in the Clinical Investigation Center at Bichat-Claude Bernard Hospital (Paris, France). All participants received oral and written information and provided signed consent before inclusion. The trial obtained approval from the Independent Ethics Committee Île-de-France 1 on 21 December 2015 (2015-oct-14028) and from the National Agency for Security of Medicinal Products on 24 July 2015 (150527A-41) and was conducted with respect to good clinical practice and the Declaration of Helsinki as last amended. It was sponsored by Assistance Publique-Hôpitaux de Paris.

Subject selection criteria. Male and female healthy volunteers between 18 and 65 years of age were eligible if their body mass index was between 18.5 and 30 kg/m², if they had normal digestive transit, and if they were considered healthy by medical history, vital signs, physical examination, electrocardiogram, and blood laboratory results at a screening visit 16 to 26 days prior to randomization. Females of childbearing age were eligible if they were using contraceptives and had a negative pregnancy test at screening.

Subjects with a history of hospitalization within the previous 6 months or antibiotic exposure within the previous 3 months were not eligible, nor were those who had any history of chronic active disease, including HIV, hepatitis C virus, or hepatitis B virus infection, or a history of allergy to beta-lactams or who were under legal protection. Members of the subjects' households could not have any active chronic disease or have received any antibiotic treatment in the preceding 15 days. The subjects were secondarily excluded if they did not provide a fecal sample before randomization or if more than 1 fecal sample was missing between day 1 and day 7 after antibiotic treatment initiation.

Treatments. Volunteers were randomized (at a 1:1 ratio) to receive from day 1 to day 3 either ceftriaxone (1 g/24 h) or cefotaxime (1 g/8 h) as a 30-min intravenous infusion using an automatic high-precision infusion pump. Day 1 was defined as the first day of antibiotic treatment.

Plasma sampling and analyses. For each volunteer, six blood samples were collected at day 3 for determination of the total concentration of antibiotics in plasma. In the ceftriaxone treatment group, blood samples were collected just before and at 0.5, 1, 2, 4, and 8 h after the beginning of the 3rd infusion. In the cefotaxime treatment group, blood samples were collected just before and at 0.5, 1, 2, 4, and 6 h after the 7th infusion. The exact times of the beginning and end of infusion, as well as the exact sampling times, were recorded. Blood samples were taken from the arm opposite that used for antibiotic administration.

Blood samples were immediately centrifuged (4,000 rpm for 5 min), and plasma was stored at -80°C until analysis. The total plasma concentrations of ceftriaxone and cefotaxime were determined by high-performance liquid chromatography coupled with UV spectrophotometric detection (at 280 nm for ceftriaxone and 254 nm for cefotaxime). The limit of quantification was 0.5 mg/liter for both antibiotics.

Fecal sampling and analyses. A total of 13 fecal samples was obtained from each participant (at days -15 ± 2 , -6 ± 2 , -1 , 1, 2, 3, 4, 7, 10 ± 1 , 15 ± 1 , 30 ± 3 , 90 ± 7 , and 180 ± 7). All bacteriological analyses were performed following the PROBE guidelines (29).

(i) Determination of bacterial counts in feces. Fecal samples were stored at 4°C after emission and transmitted to the bacteriology laboratory after they were anonymized. One hundred milligrams of feces was suspended in 1 ml of brain heart infusion broth containing 30% glycerol and stored at -80°C .

Total Gram-negative enteric bacilli (*Enterobacteriales*) were counted by plating serial dilutions of broth on Drigalski agar (bioMérieux, Marcy l'Etoile, France). Third-generation-cephalosporin-resistant Gram-negative enteric bacilli were detected and counted on ChromID ESBL agar (bioMérieux, Marcy l'Etoile, France) and biplate ESBL agar (AES Chemunex, Ivry-sur-Seine, France) (37°C for 48 h). All distinct colonies were studied. *Enterococcus* spp. were detected and counted using Enterococcosel plates (Becton, Dickinson and Company, Franklin Lakes, NJ, USA). Intestinal colonization by *Candida* spp., *Pseudomonas aeruginosa*, *Clostridium difficile*, and *Staphylococcus aureus* was detected, and the organisms were counted on ChromID *Candida* (bioMérieux), cetrimide agar (Oxoid, Basingstoke, UK), ChromID *Clostridium difficile* (bioMérieux), and BBL mannitol salt agar (Becton, Dickinson and Company, Franklin Lakes, NJ, USA), respectively. The limit of quantification was $2.0 \log_{10}$ CFU per gram of feces for all microorganisms.

Third-generation-cephalosporin-resistant Gram-negative enteric bacillus strains were tested for antibiotic susceptibility by the disk diffusion method described by EUCAST. The presence of an extended-spectrum beta-lactamase (ESBL) was detected using the double-disk synergy test. The overproduction of intrinsic or plasmid AmpC cephalosporinase was detected by studying susceptibility on cloxacillin agar (30). Resistant strains were identified by mass spectrometry (MALDI Biotyper; Bruker Daltonics, Bremen, Germany). Third-generation-cephalosporin-resistant *Escherichia coli* strains were typed or sequenced by

PCR (Illumina Technology) for determination of the molecular support of resistance as well as the phylogroup, sequence type, and O:H type (see Text S1 in the supplemental material).

(ii) 16S rRNA gene profiling of the intestinal microbiota. Samples obtained at days -15 ± 2 , -1 , 4 , 7 , 10 ± 1 , and 30 ± 3 were also analyzed by 16S rRNA gene profiling. Operational taxonomic units (OTUs) were aggregated at the phylum level, and the relative abundance of each bacterial phylum was determined. Various indices of diversity were computed. α -diversity (intrasample) metrics included the Shannon diversity index (31) and the number of observed OTUs. β -Diversity (intersample) metrics included the Bray-Curtis dissimilarity (32) and the unweighted UniFrac distance (33). Detailed methods are presented in Text S1.

(iii) Determination of fecal antibiotic concentration. Fecal samples were stored at -80°C until the assay was performed. The fecal concentrations of active ceftriaxone and cefotaxime (and their metabolites) were measured in samples collected between day -1 and day 10 by microbiological assay (with *E. coli* ATCC 25922 for both antibiotics) after incubation at 37°C for 24 h, and the limit of quantification was $1.25 \mu\text{g/g}$ for both antibiotics.

Pharmacokinetic and bacteriological endpoints used for treatment comparison. The primary endpoint was the area under the curve (AUC) between the baseline (defined as day 0) and day 7 ($\text{AUC}_{\text{D}0\text{D}7}$) of the \log_{10} counts of 3rd-generation-cephalosporin-resistant Gram-negative enteric bacilli.

Bacteriological secondary endpoints included the following: the AUC between baseline and day 15 ($\text{AUC}_{\text{D}0\text{D}15}$) of the variation from baseline of the \log_{10} counts of 3rd-generation-cephalosporin-resistant Gram-negative enteric bacilli; the $\text{AUC}_{\text{D}0\text{D}7}$ and $\text{AUC}_{\text{D}0\text{D}15}$ of the variation from baseline of the \log_{10} counts of total Gram-negative enteric bacilli, *Enterococcus* spp., and noncommensal microorganisms in the intestinal microbiota; and the proportion of uncolonized subjects with the emergence of 3rd-generation-cephalosporin-resistant Gram-negative enteric bacilli or noncommensal microorganisms in the intestinal microbiota between day 1 and day 15.

The pharmacokinetic endpoints were the area under the curve of the free plasma concentration over 24 h at steady state ($f\text{AUC}_{0-24,ss}$), the maximal and trough (minimum) free plasma concentrations (fC_{max} and fC_{min} , respectively) at steady state, and the fraction of the time during which the free plasma concentration was above 1 mg/liter at steady state ($fT > 1$) (34).

As an exploratory analysis, we computed for each subject the AUC from day -1 to day 30 ($\text{AUC}_{\text{D}-1\text{D}30}$) of the variation of the 2 indices of α -diversity. For β -diversity indices, for each sample obtained at day x , we determined the index value for each subject between day x and day -1 and computed the $\text{AUC}_{\text{D}-1\text{D}30}$ of the diversity index.

Statistical methods. The sample size of the trial was computed using data from the study by Michéa-Hamzhepour et al. (9). Assuming a common standard deviation of $2.2 \log_{10}$ CFU/g of stool, the inclusion of 12 subjects in each group was required to support a $3\text{-log}_{10}\text{-CFU/g}$ difference between treatment groups with a 90% power and a type I error of 0.05. Due to recruitment difficulties, inclusions were stopped on 18 April 2017. On this date, 22 volunteers (11 in each group) had been recruited. With this number, the power of the trial was estimated to be 86% using the above-described hypotheses.

We computed the variation from baseline of the bacterial and fungal counts in feces and computed the AUCs using the actual date and time of stool emission and the trapezoidal method. For each subject, the baseline value was computed as the arithmetic mean of the \log_{10} counts observed on available pretreatment samples. The $\text{AUC}_{\text{D}0\text{D}15}$ and $\text{AUC}_{\text{D}-1\text{D}30}$ were normalized using the observed delay between day 0 the actual time of collection of the day 15 sample and day -1 and the actual time of collection of the day 30 sample, respectively.

The plasma concentrations of antibiotics were analyzed using the population approach (Text S1). Pharmacokinetic endpoints were derived for each antibiotic at steady state, assuming 90% protein binding for ceftriaxone (19) and 40% for cefotaxime (35). Derived pharmacokinetic endpoints ($f\text{AUC}_{0-24,ss}$, fC_{max} , fC_{min} , and $fT > 1$) were computed for each subject using the predicted individual pharmacokinetic profiles. We used the 1-mg/liter threshold, as this value corresponds to the susceptibility breakpoint of *Enterobacteriaceae* species for ceftriaxone and cefotaxime (http://www.eucast.org/clinical_breakpoints/).

Endpoints were compared between groups using bilateral nonparametric Wilcoxon or Fisher's exact tests, as appropriate, and a type I error of 0.05. We computed 95% confidence intervals of proportions using the binomial distribution. We analyzed the link between each pharmacokinetic endpoint and the $\text{AUC}_{\text{D}-1\text{D}30}$ of the diversity indices studied using the Spearman rank correlation coefficient.

Data are presented as the median (minimum, maximum) or the number (percent) of subjects. Statistical analyses were performed using SAS (version 9.4) software (SAS Institute, USA).

Accession number(s). Sequence data have been submitted to the EBI database under BioProject accession number [PRJEB28341](https://www.ebi.ac.uk/bioproject/28341).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/AAC.02244-18>.

SUPPLEMENTAL FILE 1, PDF file, 2.3 MB.

ACKNOWLEDGMENTS

The members of the CEREMI Group include the following: for the Scientific Council, Antoine Andreumont (INSERM, Université Paris Diderot, Bichat Hospital), Charles Burdet (INSERM, Université Paris Diderot, Bichat Hospital), Xavier Duval (INSERM, Université

Paris Diderot, Bichat Hospital), Bruno Fantin (INSERM, Université Paris Diderot, Beaujon Hospital), Nathalie Grall (INSERM, Université Paris Diderot, Bichat Hospital), Estelle Marcault (Bichat Hospital), Laurent Massias (INSERM, Université Paris Diderot, Bichat Hospital), France Mentré (INSERM, Université Paris Diderot, Bichat Hospital), and Sarah Tubiana (INSERM, Bichat Hospital); for inclusion of participants and follow-up, Loubna Alavoine (INSERM, Bichat Hospital), Michèle Benhayoun (INSERM, Bichat Hospital), Fatima Djerdjour (INSERM, Bichat Hospital), Jean-Luc Ecobichon (INSERM, Bichat Hospital), Emila Ilic-Habensus (INSERM, Bichat Hospital), Albane Laparra (INSERM, Bichat Hospital), Milica Mandic (INSERM, Bichat Hospital), Marie Ella Nisus (INSERM, Bichat Hospital), Sandra Raine (INSERM, Bichat Hospital), Pascal Ralaimazava (INSERM, Bichat Hospital), and Valérie Vignali (INSERM, Bichat Hospital); for phenotypic analyses and microbiologic assays, Antoine Andremont (INSERM, Université Paris Diderot, Bichat Hospital), Camille d'Humières (INSERM, Université Paris Diderot, Bichat Hospital), Nathalie Grall (INSERM, Université Paris Diderot, Bichat Hospital), and Julie Riberty (Bichat Hospital); for sequencing analyses, Khadija Bourabha (INSERM, Université Paris Diderot), Antoine Bridier Nahmias (INSERM, Université Paris Diderot), Olivier Clermont (INSERM, Université Paris Diderot), Erick Denamur (INSERM, Université Paris Diderot), Mélanie Magnan (INSERM, Université Paris Diderot), and Olivier Tenaillon (INSERM, Université Paris Diderot); for plasma concentration measurements, Laurent Massias (INSERM, Univ Paris Diderot, Bichat Hospital); for data management and monitoring, Estelle Marcault (Bichat Hospital), Marion Schneider (Bichat Hospital), and Isabelle Vivaldo (Bichat Hospital); and for statistical analyses, Charles Burdet (INSERM, Université Paris Diderot, Bichat Hospital), Morgane Linard (Bichat Hospital), and France Mentré (INSERM, Université Paris Diderot, Bichat Hospital).

We all report no conflict of interest.

C.B., N.G., L.M., S.T., A.A., F.M., and X.D. conceived the study. C.B., E.D., A.A., F.M., and X.D. obtained funding. C.B., M.B., L.A., and X.D. included the subjects. N.G., C.D., and A.A. performed the phenotypic analyses. A.B.-N., K.B., M.M., O.T., and E.D. performed the 16S rRNA gene profiling. K.B., O.C., and E.D. performed the molecular analysis of the resistant strains. L.M. performed the pharmacologic analyses. C.B., M.L., and F.M. performed the statistical analyses. C.B., N.G., F.M., and X.D. wrote the first draft of the manuscript. All authors approved the final version of the manuscript.

This clinical trial was sponsored by Assistance Publique-Hôpitaux de Paris (Paris, France) and funded by the Contrat de Recherche Clinique 2013 (Assistance Publique-Hôpitaux de Paris, Département de la Recherche Clinique et du Développement, CRC13-179). This work was partially supported by a grant from the Fondation pour la Recherche Médicale to E.D. (Equipe FRM 2016, grant number DEQ20161136698).

REFERENCES

- Harper K, Armelagos G. 2010. The changing disease-scape in the third epidemiological transition. *Int J Environ Res Public Health* 7:675–697. <https://doi.org/10.3390/ijerph7020675>.
- Woerther PL, Burdet C, Chachaty E, Andremont A. 2013. Trends in human fecal carriage of extended-spectrum beta-lactamases in the community: toward the globalization of CTX-M. *Clin Microbiol Rev* 26:744–758. <https://doi.org/10.1128/CMR.00023-13>.
- Andremont A. 2003. Commensal flora may play key role in spreading antibiotic resistance. *ASM News* 69:601–607.
- Dethlefsen L, Huse S, Sogin ML, Relman DA. 2008. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol* 6:e280. <https://doi.org/10.1371/journal.pbio.0060280>.
- Dethlefsen L, Relman DA. 2011. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proc Natl Acad Sci U S A* 108(Suppl 1):4554–4561. <https://doi.org/10.1073/pnas.1000087107>.
- de Gunzburg J, Ghozlane A, Ducher A, Le Chatelier E, Duval X, Ruppé E, Armand-Lefevre L, Sablier-Gallis F, Burdet C, Alavoine L, Chachaty E, Augustin V, Varastet M, Levenez F, Kennedy S, Pons N, Mentré F, Andremont A. 2018. Protection of the human gut microbiome from antibiotics. *J Infect Dis* 217:628–636. <https://doi.org/10.1093/infdis/jix604>.
- Ruppe E, Burdet C, Grall N, de Lastours V, Lescure FX, Andremont A, Armand LL. 2018. Impact of antibiotics on the intestinal microbiota needs to be re-defined to optimize antibiotic usage. *Clin Microbiol Infect* 24:3–5. <https://doi.org/10.1016/j.cmi.2017.09.017>.
- Clemente JC, Ursell LK, Parfrey LW, Knight R. 2012. The impact of the gut microbiota on human health: an integrative view. *Cell* 148:1258–1270. <https://doi.org/10.1016/j.cell.2012.01.035>.
- Michéa-Hamzehpour M, Auckenthaler R, Kunz J, Pechère JC. 1988. Effect of a single dose of cefotaxime or ceftriaxone on human faecal flora. A double-blind study. *Drugs (Suppl 2)*:6–11. <https://doi.org/10.2165/00003495-198800352-00004>.
- Brautigam HH, Knothe H, Rangoonwala R. 1988. Impact of cefotaxime and ceftriaxone on the bowel and vaginal flora after single-dose prophylaxis in vaginal hysterectomy. *Drugs (Suppl 2)*:163–168. <https://doi.org/10.2165/00003495-198800352-00035>.
- Gbaguidi-Haore H, Dumartin C, L'Heriteau F, Pefau M, Hocquet D, Rogues AM, Bertrand X. 2013. Antibiotics involved in the occurrence of antibiotic-resistant bacteria: a nationwide multilevel study suggests dif-

- ferences within antibiotic classes. *J Antimicrob Chemother* 68:461–470. <https://doi.org/10.1093/jac/dks406>.
12. Grohs P, Kerneis S, Sabatier B, Lavollay M, Carbonnelle E, Rostane H, Souty C, Meyer G, Gutmann L, Mainardi JL. 2014. Fighting the spread of AmpC-hyperproducing *Enterobacteriaceae*: beneficial effect of replacing ceftriaxone with cefotaxime. *J Antimicrob Chemother* 69:786–789. <https://doi.org/10.1093/jac/dkt403>.
 13. Patel IH, Chen S, Parsonnet M, Hackman MR, Brooks MA, Konikoff J, Kaplan SA. 1981. Pharmacokinetics of ceftriaxone in humans. *Antimicrob Agents Chemother* 20:634–641. <https://doi.org/10.1128/AAC.20.5.634>.
 14. Patel IH, Kaplan SA. 1984. Pharmacokinetic profile of ceftriaxone in man. *Am J Med* 77:17–25.
 15. Kemmerich B, Lode H, Belmega G, Jendroschek T, Borner K, Koeppe P. 1983. Comparative pharmacokinetics of cefoperazone, cefotaxime, and moxalactam. *Antimicrob Agents Chemother* 23:429–434. <https://doi.org/10.1128/AAC.23.3.429>.
 16. Jehl F, Peter JD, Picard A, Dupeyron JP, Marescaux J, Sibilly A, Monteil H. 1987. Biliary excretion of cefotaxime and desacetylcefotaxime. *Rev Med Interne* 8:487–492. (In French.) ([https://doi.org/10.1016/S0248-8663\(87\)80198-4](https://doi.org/10.1016/S0248-8663(87)80198-4)).
 17. van Ogtrop ML, Guidet HF, Mattie H, van Furth R. 1991. Modulation of the intestinal flora of mice by parenteral treatment with broad-spectrum cephalosporins. *Antimicrob Agents Chemother* 35:976–982. <https://doi.org/10.1128/AAC.35.5.976>.
 18. Neu HC. 1982. The in vitro activity, human pharmacology, and clinical effectiveness of new beta-lactam antibiotics. *Annu Rev Pharmacol Toxicol* 22:599–642. <https://doi.org/10.1146/annurev.pa.22.040182.003123>.
 19. Stoeckel K, McNamara PJ, Brandt R, Plozza-Nottebrock H, Ziegler WH. 1981. Effects of concentration-dependent plasma protein binding on ceftriaxone kinetics. *Clin Pharmacol Ther* 29:650–657. <https://doi.org/10.1038/clpt.1981.90>.
 20. Bakken JS, Cavalieri SJ, Gangeness D. 1990. Influence of plasma exchange pheresis on plasma elimination of ceftriaxone. *Antimicrob Agents Chemother* 34:1276–1277. <https://doi.org/10.1128/AAC.34.6.1276>.
 21. Stiefel U, Tima MA, Nerandzic MM. 2015. Metallo-beta-lactamase-producing *Bacteroides* species can shield other members of the gut microbiota from antibiotics. *Antimicrob Agents Chemother* 59:650–653. <https://doi.org/10.1128/AAC.03719-14>.
 22. Stentz R, Horn N, Cross K, Salt L, Brearley C, Livermore DM, Carding SR. 2015. Cephalosporinases associated with outer membrane vesicles released by *Bacteroides* spp. protect gut pathogens and commensals against beta-lactam antibiotics. *J Antimicrob Chemother* 70:701–709. <https://doi.org/10.1093/jac/dku466>.
 23. Leonard F, Andreumont A, Leclercq B, Labia R, Tancrede C. 1989. Use of beta-lactamase-producing anaerobes to prevent ceftriaxone from degrading intestinal resistance to colonization. *J Infect Dis* 160:274–280. <https://doi.org/10.1093/infdis/160.2.274>.
 24. Kokai-Kun JF, Roberts T, Coughlin O, Sicard E, Rufange M, Fedorak R, Carter C, Adams MH, Longstreth J, Whalen H, Sliman J. 2017. The oral beta-lactamase SYN-004 (Ribaxamase) degrades ceftriaxone excreted into the intestine in phase 2a clinical studies. *Antimicrob Agents Chemother* 61:e02197-16. <https://doi.org/10.1128/AAC.02197-16>.
 25. Baquero F, Tedim AP, Coque TM. 2013. Antibiotic resistance shaping multi-level population biology of bacteria. *Front Microbiol* 4:15. <https://doi.org/10.3389/fmicb.2013.00015>.
 26. Ruppé E, Ghozlane A, Tap J, Pons N, Alvarez A-S, Maziers N, Cuesta T, Hernando-Amado S, Clares I, Martínez JL, Coque TM, Baquero F, Lanza VF, Máiz L, Goulenok T, de Lastours V, Amor N, Fantin B, Wieder I, Andreumont A, van Schaik W, Rogers M, Zhang X, Willems RJL, de Brevérn AG, Batto J-M, Blottière HM, Léonard P, Lédard V, Letur A, Levenez F, Weisz K, Haimet F, Doré J, Kennedy SP, Ehrlich SD. 2019. Prediction of the intestinal resistome by a three-dimensional structure-based method. *Nat Microbiol* 4:112–123. <https://doi.org/10.1038/s41564-018-0292-6>.
 27. Tenaillon O, Skurnik D, Picard B, Denamur E. 2010. The population genetics of commensal *Escherichia coli*. *Nat Rev Microbiol* 8:207–217. <https://doi.org/10.1038/nrmicro2298>.
 28. Lanza VF, Baquero F, Martínez JL, Ramos-Ruiz R, González-Zorn B, Andreumont A, Sánchez-Valenzuela A, Ehrlich SD, Kennedy S, Ruppé E, van Schaik W, Willems RJ, de la Cruz F, Coque TM. 2018. In-depth resistome analysis by targeted metagenomics. *Microbiome* 6:11. <https://doi.org/10.1186/s40168-017-0387-y>.
 29. Hansson L, Hedner T, Dahlof B. 1992. Prospective Randomized Open Blinded End-Point (PROBE) study. A novel design for intervention trials. *Blood Press* 1:113–119. <https://doi.org/10.3109/08037059209077502>.
 30. Armand-Lefèvre L, Angebault C, Barbier F, Hamelet E, Defrance G, Ruppé E, Bronchard R, Lepeule R, Lucet J-C, El Mniai A, Wolff M, Montravers P, Plésiat P, Andreumont A. 2013. Emergence of imipenem-resistant gram-negative bacilli in intestinal flora of intensive care patients. *Antimicrob Agents Chemother* 57:1488–1495. <https://doi.org/10.1128/AAC.01823-12>.
 31. Shannon C. 1948. A mathematical theory of communication. *Bell Syst Tech J* 27:623–656. <https://doi.org/10.1002/j.1538-7305.1948.tb00917.x>.
 32. Bray JR, Curtis JT. 1957. An ordination of the upland forest communities of southern Wisconsin. *Ecol Monogr* 27:325–349. <https://doi.org/10.2307/1942268>.
 33. Lozupone C, Knight R. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* 71:8228–8235. <https://doi.org/10.1128/AEM.71.12.8228-8235.2005>.
 34. Craig WA. 1998. Pharmacokinetic/pharmacodynamic parameters: rationale for antibacterial dosing of mice and men. *Clin Infect Dis* 26:1–10. <https://doi.org/10.1086/516284>.
 35. Esmieu F, Guibert J, Rosenkilde HC, Ho I, Le Go A. 1980. Pharmacokinetics of cefotaxime in normal human volunteers. *J Antimicrob Chemother* 6(Suppl A):83–92. https://doi.org/10.1093/jac/6.suppl_A.83.

Supplementary Material

Supplementary text S1. Supplementary methods

Analysis of 3rd generation cephalosporins resistant Gram-negative enteric bacilli

In a first step, all *E. coli* strains resistant to 3rd generation cephalosporins were phylogrouped and O-typed by PCR-based approaches as in (1, 2). According to the result, the belonging to the ST131/506 was also performed as in (3, 4). In a second step, one to four strains per subject, as well as an *Enterobacter cloacae* strain, were sequenced on an Illumina MiSeq[®] (2 X 300 bp). Assembly was performed with SPAdes, version 3.10.1 using standard parameters (5). Multilocus sequence type and O-and H-type determination and resistome were determined on the Center for Genomic Epidemiology platform (www.genomicepidemiology.org).

16S rRNA gene profiling of the intestinal microbiota

Microbial DNA was extracted using QIAamp DNA stool Kit (Qiagen, Germany). The V4 region of the 16S rRNA gene was then amplified using primers selected from (6). Their sequences were TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG GTG YCA GCM GCC GCG GTA A for primer 515f and GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GGG ACT ACN VGG GTW TCT for primer 806r. The PCR amplicons (292 bp) were sequenced on an Illumina MiSeq[®] platform according to the manufacturer's specifications (Illumina, USA). Demultiplexing and quality filtering were performed using mothur (7). No mismatch was allowed in the primers sequences in both forward and reverse primers, and quality-filtering was performed by truncating bases at the 3' end with Phred quality score <25. Paired-end read assembly was then performed with 100% overlap identity.

Following these pre-processing steps, chimera sequences were detected and eliminated using Vsearch (8). Then, clustering of similar sequences (97% identity threshold for an affiliation at the genus level) was performed through an open-reference OTU picking process and complete-linkage method, finally creating groups of sequences or "Operational Taxonomic Units" (OTUs). A final OTU cleaning step corresponding to the elimination of singletons and doubletons was performed. Diversity indices were computed for each sample after rarefaction of the data (59,742 sequences allowed an exhaustive description of the bacterial diversity).

Pharmacokinetic analysis of plasma concentrations of antibiotics

For the analysis of total plasma concentration of ceftriaxone and cefotaxime, 2 separate population pharmacokinetic analyses were performed using the Stochastic Approximation Expectation Minimization (SAEM) algorithm in Monolix v4.2 (Lixoft, Orsay, France, available at <http://www.lixoft.com>), which handles concentrations below the limit of quantification to improve parameters estimation (9). Both one- and two-compartment(s) models with first order elimination were tested, and parameters to be estimated were V, the central volume of distribution and k, the first-order

elimination rate. For two-compartment models, we also estimated k_{12} and k_{21} , the transfer rates between central and peripheral compartments.

We used an exponential random effects model for each pharmacokinetic parameter. We assumed the random effects to have a normal distribution with a mean of 0 and a variance of ω^2 . The residual error model was supposed to be additive, proportional or combined, with σ_{inter} being the standard deviation of the additive component and σ_{slope} the standard deviation of the proportional component.

Evaluation of the final model was conducted using graphical methods. Basic goodness-of-fit plots were used, as well as the individual weighted residuals (IWRES), the normalized prediction distribution errors (NPDE) over time and the visual predictive check (VPC). NPDE and VPC were generated using 500 Monte Carlo simulations. The best model was chosen using the Bayesian information criteria (BIC), derived for each model from the computation of likelihood by importance sampling (10).

Individual pharmacokinetic parameters were estimated as the maximum of the *a posteriori* distribution, also called empirical Bayes estimates, and used to predict individual pharmacokinetic profiles of included subjects.

Supplementary Table S1. Estimated population parameters and their relative standard errors (r.s.e) for the pharmacokinetic models of ceftriaxone and cefotaxime in the plasma. The models were separately fitted on the data from the 11 subjects treated by ceftriaxone and on the data from the 11 subjects treated by cefotaxime. V is the volume of distribution; k is the extraintestinal elimination rate from the central compartment; k_{12} and k_{21} are the transfer rates between the central compartment and the peripheral compartment; σ_{slope} is the proportional component of the residual error.

Parameter	Ceftriaxone		Cefotaxime	
	Fixed effects (r.s.e. %)	Standard deviation of the exponential random effects (r.s.e. %)	Fixed effects (r.s.e. %)	Standard deviation of the exponential random effects (r.s.e. %)
V (L)	7.4 (4)	0.11 (24)	8.7 (13)	0.19 (32)
k (/h)	0.16 (6)	0.18 (22)	1.69 (10)	0.07 (38)
k_{12} (/h)	0.35 (13)	- (fixed)	0.87 (30)	- (fixed)
k_{21} (/h)	0.54 (9)	- (fixed)	1.35 (13)	- (fixed)
σ_{slope}		0.05 (11)		0.19 (12)

Supplementary Table S2. Main phenotypic and genotypic characteristics of the 3rd-generation cephalosporin-resistant Gram-negative enteric bacilli in the 22 included subjects. Phylogroup, sequence type 131/506 or 10 and O-type were determined in all *E. coli* strains by PCR-based assays as in (1-4). Sequenced strains are presented in bold fonts. ND, not determined.

Strain ID	Treatment group	Subject	Sampling time (day)	Species	Antibioresistance phenotype*	Phylogroup	ST Warwick**	ST Pasteur Institute***	Serotype	Molecular support of resistance
E1	ceftriaxone	1	-15	<i>E. coli</i>	ESBL	B2	131	506	O16:H5	CTX-M-55
E3	ceftriaxone	1	-7	<i>E. coli</i>	ESBL	B2	131	506	O16	ND
E5	ceftriaxone	1	-1	<i>E. coli</i>	ESBL	B2	131	506	O16	ND
E7	ceftriaxone	1	1	<i>E. coli</i>	ESBL	B2	131	506	O16	ND
E9	ceftriaxone	1	2	<i>E. coli</i>	ESBL	B2	131	506	O16	ND
E8	ceftriaxone	1	4	<i>E. coli</i>	ESBL	B2	131	506	O16	ND
E10	ceftriaxone	1	7	<i>E. coli</i>	ESBL	B2	131	506	O16	ND
E12	ceftriaxone	1	10	<i>E. coli</i>	ESBL	B2	131	506	O16:H5	CTX-M-55
E13	ceftriaxone	1	15	<i>E. coli</i>	ESBL	B2	131	506	O16	ND
E14	ceftriaxone	1	30	<i>E. coli</i>	ESBL	B2	131	506	O16:H5	CTX-M-55
E15	ceftriaxone	1	90	<i>E. coli</i>	ESBL	B2	131	567	O16:H5	CTX-M-55
E11	cefotaxime	2	7	<i>E. coli</i>	pCASE	F	117	539	O8:H4	CMY-2
E2	cefotaxime	3	-15	<i>E. coli</i>	ESBL	B2	131	506	O16:H5	CTX-M-55
E4	ceftriaxone	6	-7	<i>Hafnia alvei</i>	dCASE	ND	ND	ND	ND	ND
E17	ceftriaxone	6	-1	<i>H. alvei</i>	dCASE	ND	ND	ND	ND	ND
E16	ceftriaxone	9	15	<i>E. coli</i>	ESBL	A	46	398	O9:H4	CTX-M-15
E19	cefotaxime	17	7	<i>Enterobacter cloacae</i>	ESBL	ND	ND	ND	ND	CTX-M-15
E18	cefotaxime	17	30	<i>E. coli</i>	ESBL	B1	58	24	O9:H25	CTX-M-1
E22	ceftriaxone	18	-7	<i>E. coli</i>	pCASE	A	10	913	O8:H17	CMY-4
E23	ceftriaxone	18	1	<i>E. coli</i>	pCASE	A	10	ND	O8	ND
E21	ceftriaxone	18	3	<i>E. coli</i>	pCASE	A	10	913	O8:H17	CMY-4
E20	ceftriaxone	22	-7	<i>Citrobacter freundii</i>	dCASE	ND	ND	ND	ND	ND
E25	ceftriaxone	28	-7	<i>E. coli</i>	ESBL	A	10	2	O71:H48	CTX-M-1
E26	ceftriaxone	28	-1	<i>E. coli</i>	ESBL	A	10	ND	O73	ND
E27	ceftriaxone	28	1	<i>E. coli</i>	ESBL	A	10	ND	O73	ND
E33	ceftriaxone	28	3	<i>E. coli</i>	ESBL	A	10	ND	O73	ND
E34	ceftriaxone	28	4	<i>E. coli</i>	ESBL	A	10	ND	O73	ND
E29	ceftriaxone	28	7	<i>E. coli</i>	ESBL	A	10	ND	O73	ND
E31	ceftriaxone	28	10	<i>E. coli</i>	ESBL	A	10	ND	O73	ND
E32	ceftriaxone	28	15	<i>E. coli</i>	ESBL	A	10	914	O73:H31	CTX-M-1
E30	ceftriaxone	28	30	<i>E. coli</i>	ESBL	A	10	914	O73:H31	CTX-M-1

* ESBL, extended spectrum beta-lactamase; pCASE, plasmid-encoded cephalosporinase; dCASE, derepressed cephalosporinase

** <https://enterobase.warwick.ac.uk/species/index/ecoli>

*** <https://www.pasteur.fr/fr/sante-publique/CNR/les-cnr/escherichia-coli-shigella-salmonella>

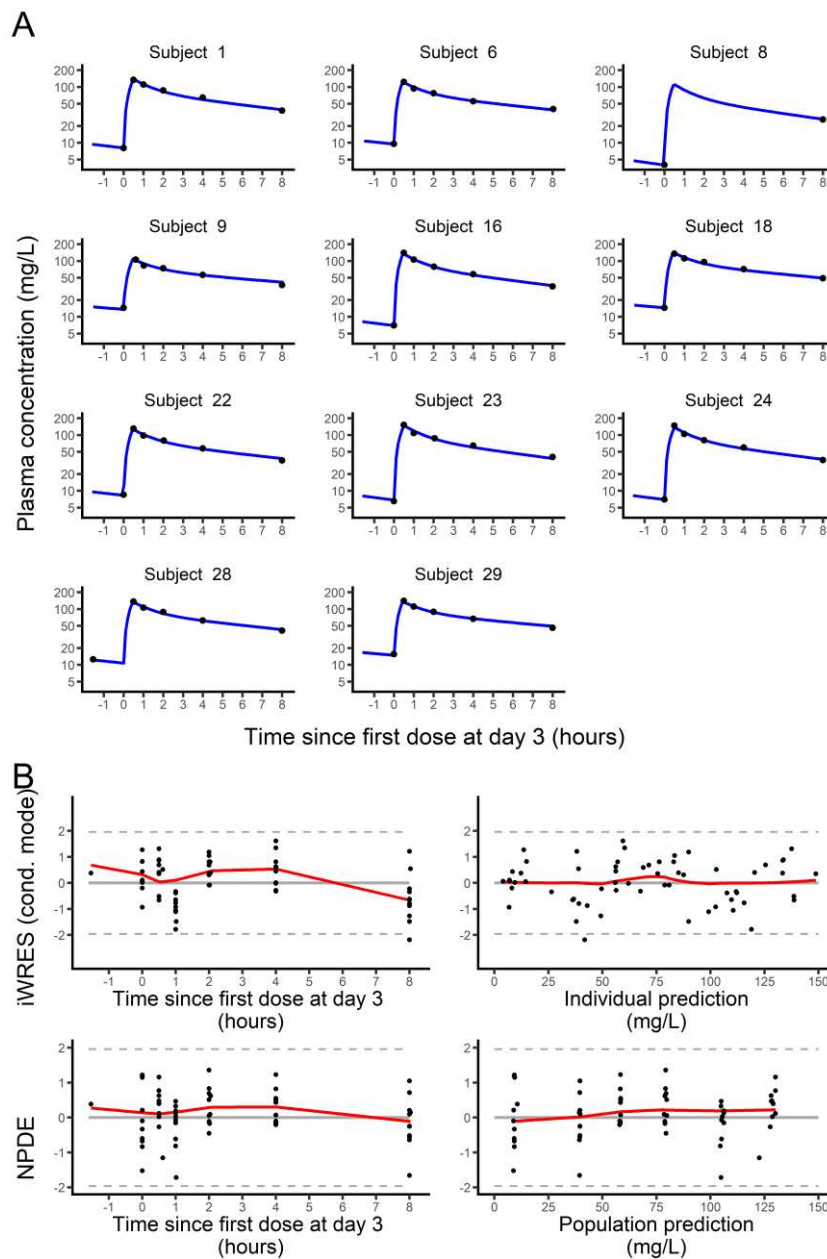
Supplementary references

1. Clermont O, Christenson JK, Denamur E, Gordon DM. 2013. The Clermont *Escherichia coli* phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. *Environ Microbiol Rep* 5:58-65.
2. Clermont O, Johnson JR, Menard M, Denamur E. 2007. Determination of *Escherichia coli* O types by allele-specific polymerase chain reaction: application to the O types involved in human septicemia. *Diagn Microbiol Infect Dis* 57:129-36.
3. Clermont O, Dhanji H, Upton M, Gibreel T, Fox A, Boyd D, Mulvey MR, Nordmann P, Ruppe E, Sarthou JL, Frank T, Vimont S, Arlet G, Branger C, Woodford N, Denamur E. 2009. Rapid detection of the O25b-ST131 clone of *Escherichia coli* encompassing the CTX-M-15-producing strains. *J Antimicrob Chemother* 64:274-7.
4. Johnson JR, Clermont O, Johnston B, Clabots C, Tchesnokova V, Sokurenko E, Junka AF, Maczynska B, Denamur E. 2014. Rapid and specific detection, molecular epidemiology, and experimental virulence of the O16 subgroup within *Escherichia coli* sequence type 131. *J Clin Microbiol* 52:1358-65.
5. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455-77.
6. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glockner FO. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* 41:e1.
7. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75:7537-41.
8. Rognes T, Flouri T, Nichols B, Quince C, Mahe F. 2016. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4:e2584.
9. Panhard X, Samson A. 2009. Extension of the SAEM algorithm for nonlinear mixed models with 2 levels of random effects. *Biostatistics* 10:121-35.
10. Lavielle M, Mentre F. 2007. Estimation of population pharmacokinetic parameters of saquinavir in HIV patients with the MONOLIX software. *J Pharmacokinet Pharmacodyn* 34:229-49.

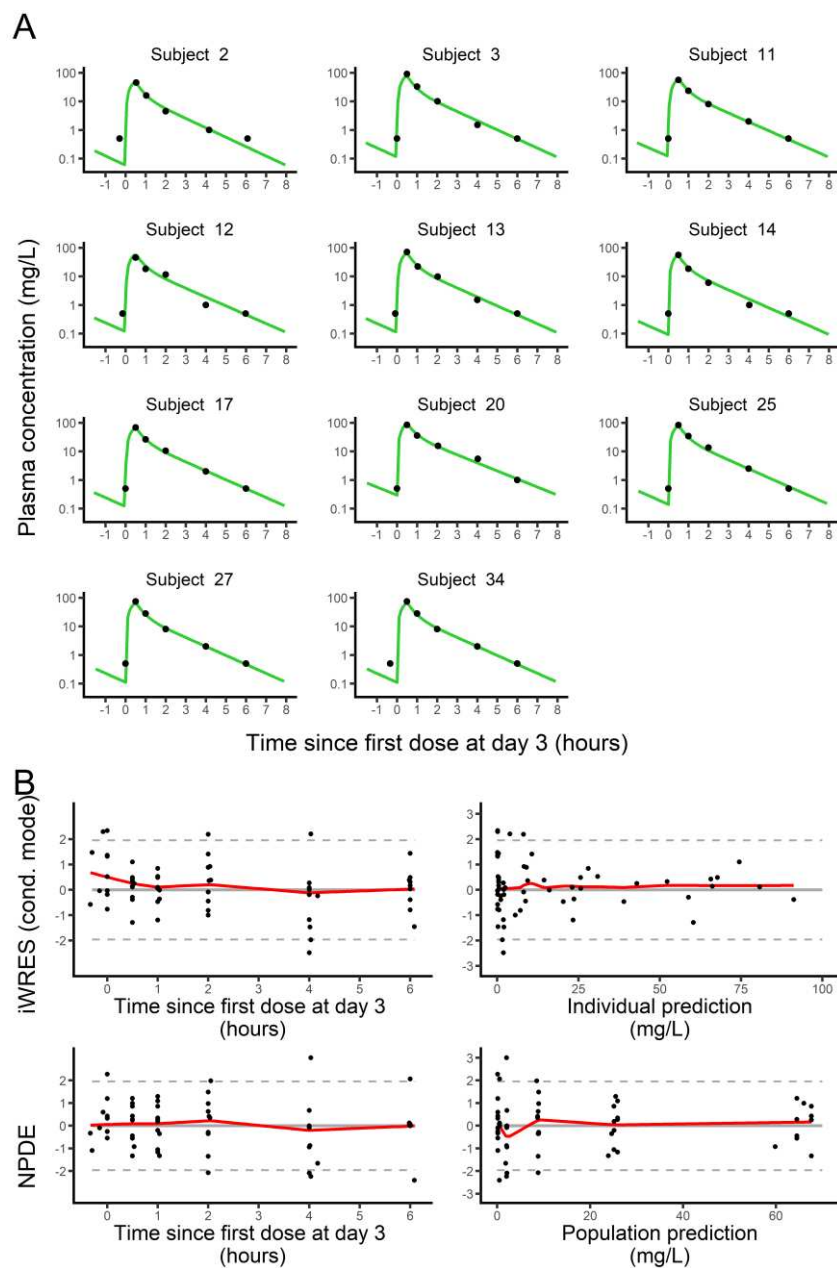
Supplementary Table S3. Spearman rank correlation coefficients of the pharmacokinetic endpoints of ceftriaxone and cefotaxime and of the AUC_{D-1D30} of change from baseline of studied indices of bacterial diversity for the 11 included subjects of each treatment group. P-values refer to the comparison to 0 of the correlation coefficients and are presented in brackets.

	Pharmacokinetic endpoints			
	$fAUC_{0-24,ss}$	fC_{max}	fC_{min}	$fT>1$
Ceftriaxone (n=11)				
Shannon diversity index	0.0 (0.98)	-0.2 (0.58)	-0.1 (0.86)	0.0 (0.91)
Number of OTUs	0.1 (0.80)	-0.1 (0.86)	0.1 (0.71)	0.1 (0.69)
Bray-Curtis dissimilarity	-0.2 (0.56)	0.0 (0.99)	0.0 (0.90)	-0.1 (0.86)
Unweighted UniFrac distance	-0.3 (0.37)	0.0 (0.95)	-0.2 (0.54)	-0.3 (0.36)
Cefotaxime (n=11)				
Shannon diversity index	-0.2 (0.63)	-0.2 (0.63)	-0.2 (0.58)	-0.1 (0.71)
Number of OTUs	-0.3 (0.36)	-0.3 (0.36)	-0.1 (0.78)	-0.3 (0.41)
Bray-Curtis dissimilarity	-0.1 (0.88)	-0.1 (0.88)	0.0 (0.95)	-0.1 (0.80)
Unweighted UniFrac distance	0.5 (0.12)	0.5 (0.12)	0.2 (0.52)	0.5 (0.12)

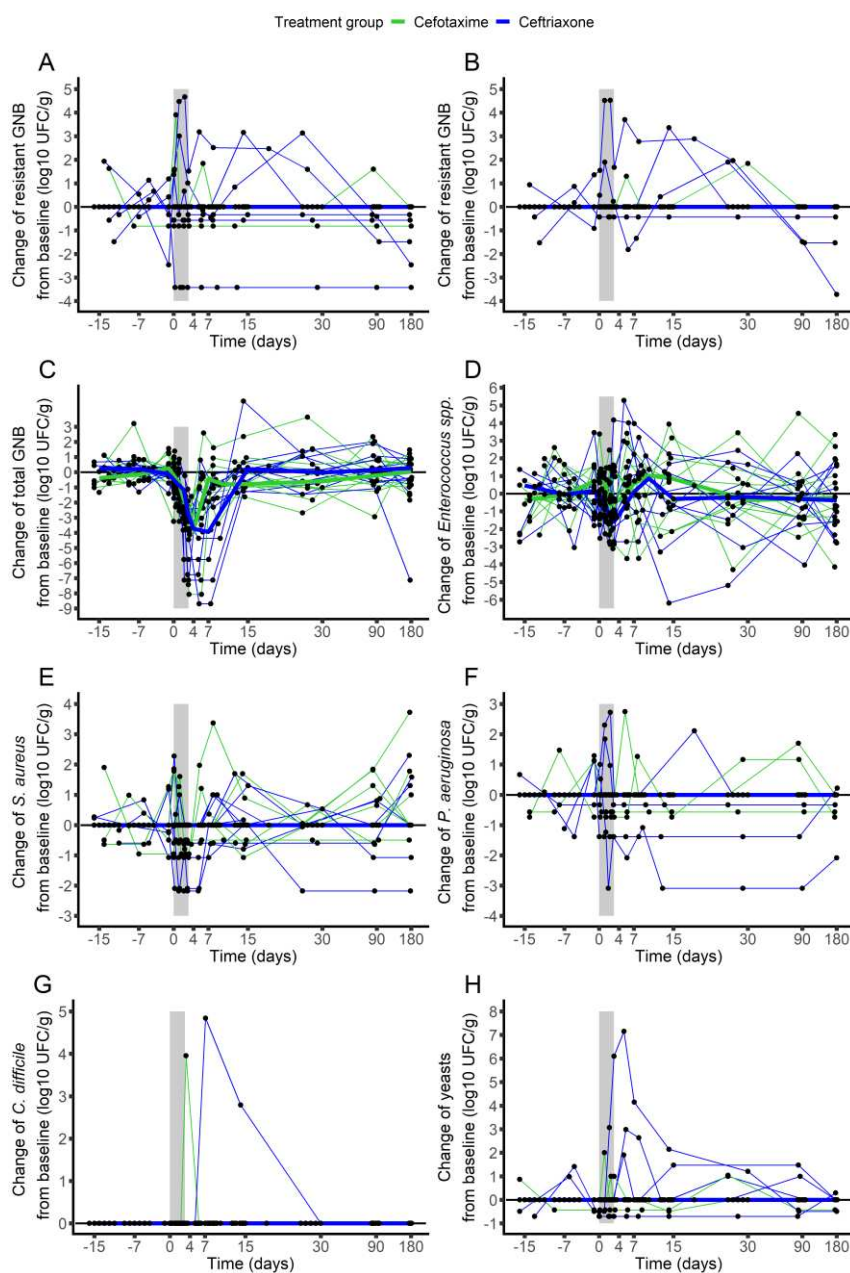
Supplementary Figure S1. Goodness of fit plots for the pharmacokinetic model of ceftriaxone in plasma. Panel A presents the individual fits of the model for the plasma total concentration of ceftriaxone. Black dots represent observed moxifloxacin concentrations. Orange curves represent the individual pharmacokinetic profiles predicted by the model using estimated individual parameters. Panel B presents the plots of the individual weighted residuals (iWRES) and normalized prediction distribution errors (NPDE) over time (left) or model predictions (right) for the total plasma concentrations. The iWRES and NPDE are shown as black points, and spline lines are also added as red curves.



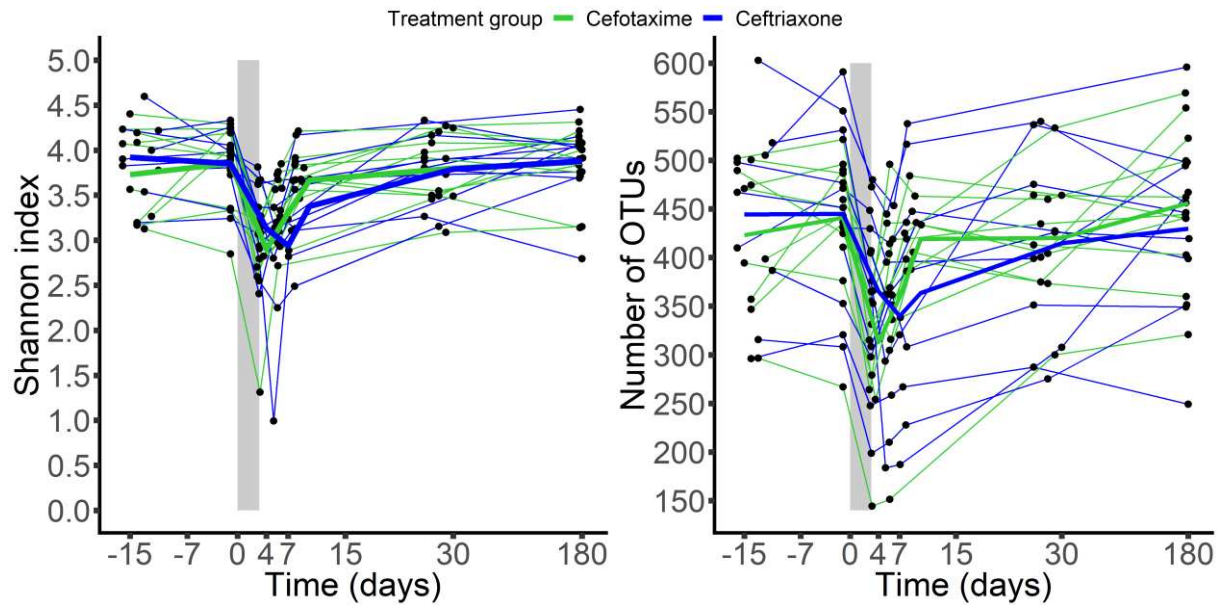
Supplementary Figure S2. Goodness of fit plots for the pharmacokinetic model of cefotaxime in plasma. Panel A presents the individual fits of the model for the plasma total concentration of cefotaxime. Black dots represent observed moxifloxacin concentrations. Green curves represent the individual pharmacokinetic profiles predicted by the model using estimated individual parameters. Panel B presents the plots of the individual weighted residuals (iWRES) and normalized prediction distribution errors (NPDE) over time (left) or model predictions (right) for the total plasma concentrations. The iWRES and NPDE are shown as black points, and spline lines are also added as red curves.



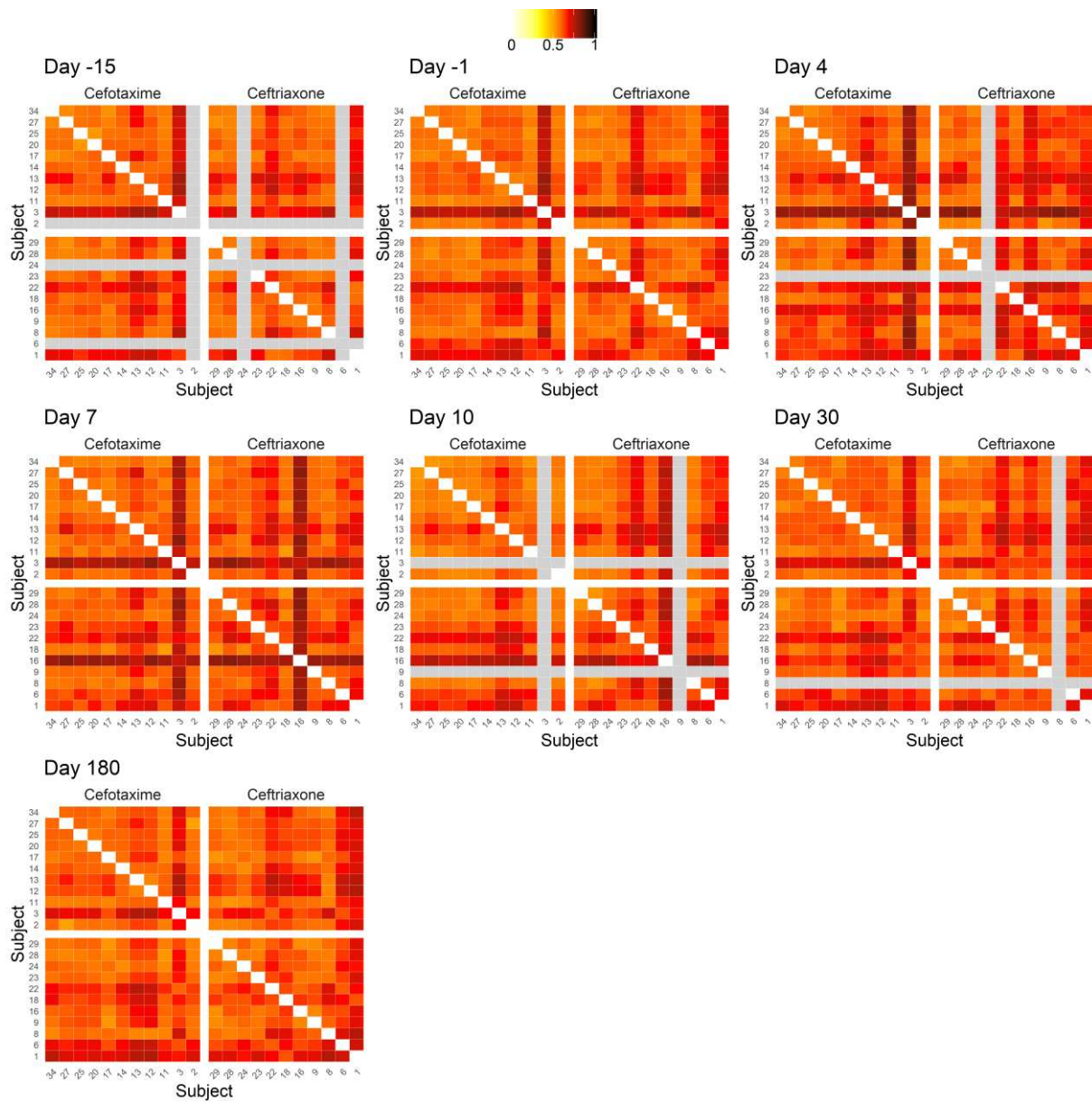
Supplementary Figure S3. Evolution of the change from baseline of the counts of the studied microorganisms in the fecal samples of the 22 included subjects treated by ceftriaxone (n=11, blue) or cefotaxime (n=11, green). Studied microorganisms included 3rd generation cephalosporin Gram-negative enteric bacilli (measured on AES plates, panel A or ChromID ESBL plates, panel B), total Gram-negative enteric bacilli (panel C), *Enterococcus* spp. (panel D), *S. aureus* (panel E), *P. aeruginosa* (panel F), *C. difficile* (panel G) and Yeasts (panel H). The light grey zone represents the treatment period. Thin lines represent individual values, and thick lines represent the median change from baseline of the log-counts at each time. For graphical reasons, baseline counts are represented at day 0. GNB, Gram-negative enteric bacilli.



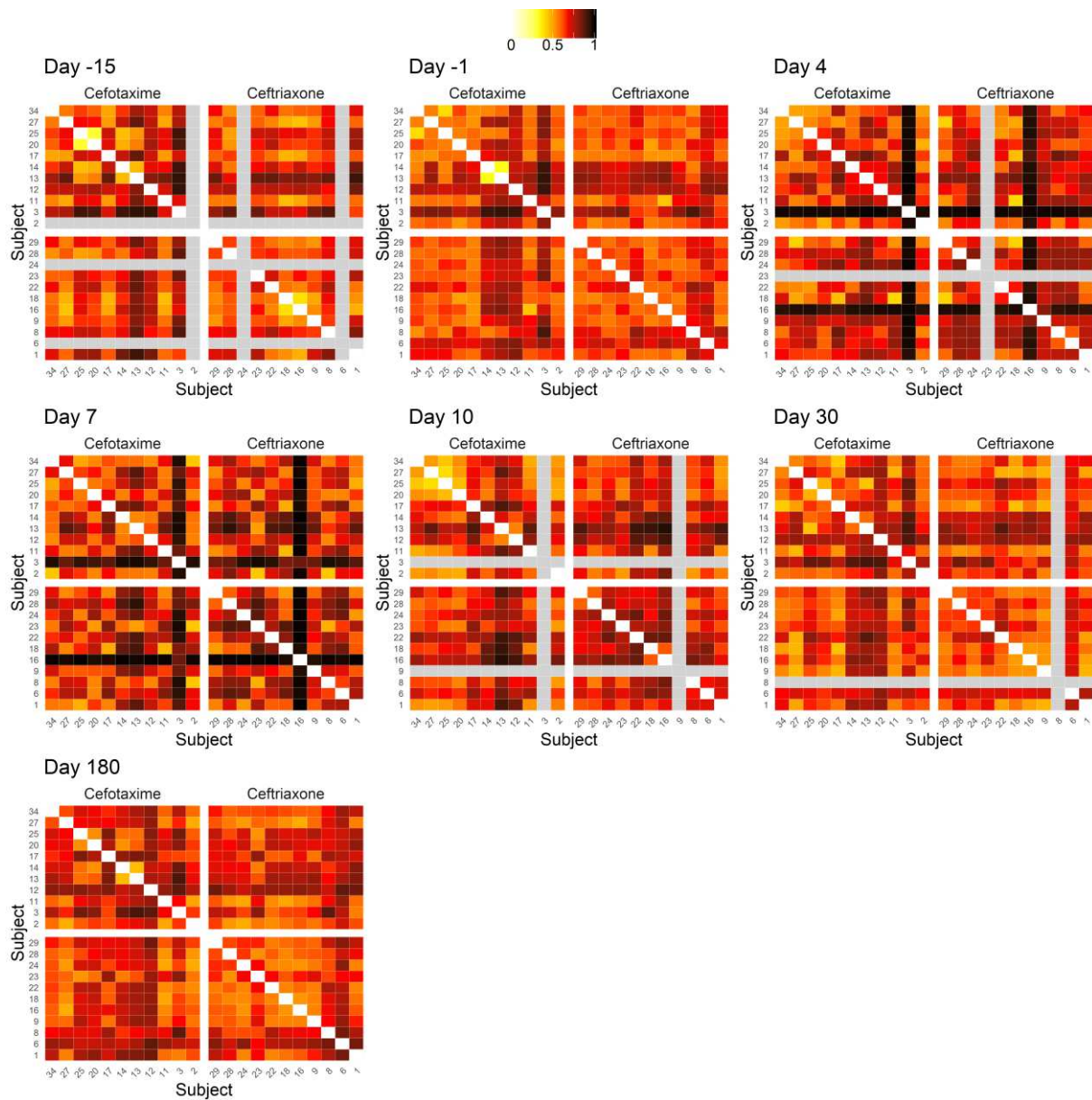
Supplementary Figure S4. Evolution of the Shannon index (panel A) and of the number of OTUs (panel B) in the fecal samples of the 22 included subjects treated by ceftriaxone (n=11, blue) or cefotaxime (n=11, green). The light grey zone represents the treatment period. Thin lines represent individual values, and thick lines represent the mean of the index values at each time.



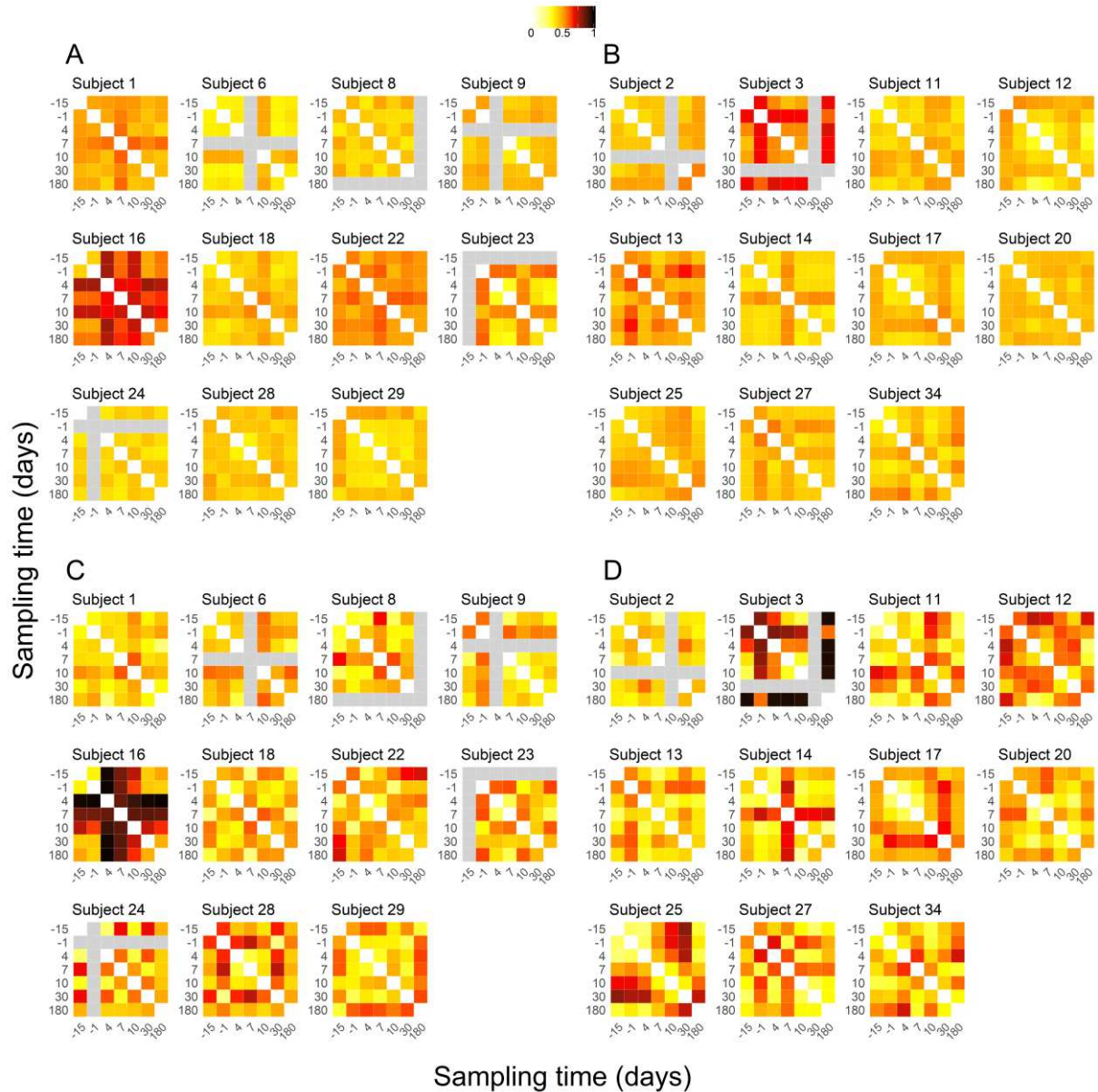
Supplementary Figure S5. Heatmap of the unweighted UniFrac distances matrix at each sampling time for the 22 included subjects. Missing samples are presented in lightgrey.



Supplementary Figure S6. Heatmap of the Bray-Curtis dissimilarity matrix at each sampling time for the 22 included subjects. Missing samples are presented as lightgrey.



Supplementary Figure S7. Heatmap of the intraindividual matrices of β -diversity indices at the various sampling times: unweighted UniFrac distance (panels A & B) and Bray-Curtis dissimilarity (panels C & D) between day -15 and day 180 for the subjects treated with ceftriaxone (n=11, panel A & C) or cefotaxime (n=11, panel B & D). Missing samples are presented in lightgrey.



3.2.3 Principaux résultats de l'article et perspectives

Les principaux résultats de l'étude sont les suivants :

- Tous les sujets sauf deux ont des concentrations indétectables d'antibiotique dans les selles. Cela est probablement dû à la présence de bêta-lactamases endogènes secrétées par les bactéries (principalement les bactéries anaérobies), du microbiote intestinal qui neutralisent les antibiotiques excrétés. Le dosage des bêta-lactamases dans les selles aux différents temps est en cours de réalisation. L'hypothèse est que les deux patients avec des quantités détectables d'antibiotiques à J4 après le début du traitement aient une faible quantité de bêta-lactamases. Ainsi, l'impact d'un antibiotique sur la flore intestinale serait également conditionné par la capacité de l'individu à pouvoir neutraliser la molécule, ceci étant un facteur individu spécifique.
- Dans les deux groupes (céfotaxime et ceftriaxone) il a été observé une réduction du nombre de bacilles à Gram négatif cultivable en post-traitement avec un retour à la normale à J15 mais sans émergence de bacilles à Gram négatif résistants aux CIIG. Cela est probablement dû au choix de la population initiale. Choisir de travailler sur des volontaires sains permet de limiter une altération du microbiote intestinal par des maladies chroniques ou des hospitalisations répétées mais c'est une population avec une faible pression de sélection aux antibiotiques qui limite donc le risque de colonisation par des bactéries résistantes aux antibiotiques.
- L'étude génotypique, par analyse du gène qui code pour l'ARNr 16s, a mis en évidence un impact profond de la ceftriaxone et du céfotaxime sur la diversité bactérienne du microbiote intestinal en post-traitement, mais n'a pas mis en évidence de différence entre les deux groupes. Cela peut s'expliquer car théoriquement si 1 gramme de ceftriaxone est donné par 24h avec une élimination biliaire de 40%, 400mg doivent être excrétés dans les selles et si 3 grammes de céfotaxime sont donnés par 24h avec une élimination biliaire de 10%, 300mg doivent être excrétés dans les selles. La quantité d'antibiotiques excrétée dans les selles est donc théoriquement assez proche dans les deux groupes, pouvant expliquer l'absence de différence observée. Une analyse

métagénomique de ces mêmes échantillons est en cours et permettra de confirmer ces observations et d'étudier d'autres paramètres comme le résistome.

En conclusion, la différence dans le mode d'élimination du céfotaxime et de la ceftriaxone est compensée par leur rythme d'administration, ainsi il n'a pas été observé de différence d'impact sur le microbiote intestinal de ces deux antibiotiques. En revanche, la ceftriaxone a deux avantages (i) le temps au-dessus de la concentration minimale inhibitrice (CMI) est plus important que la cefotaxime, témoignant d'une meilleure efficacité pour lutter contre le germe responsable de la pathologie traitée et (ii) le schéma d'administration (1 fois par jour) est plus confortable pour le patient et pour les infirmières. Ainsi, la prescription de la ceftriaxone paraît s'avérer plus intéressante que le céfotaxime.

3.3 Etude clinique d'impact de la céfotaxime et du ceftriaxone sur le microbiote intestinal chez des volontaires sains : analyse du phageome

3.3.1 Introduction

L'impact des antibiotiques sur le phageome intestinal humain commence à peine à être exploré. Deux études ont montré une modification du phageome sous traitement antibiotique chez un très petit nombre de volontaires sains ^{120,121}. Plus précisément chez un individu, on a observé un shift de l'abondance des phages de la famille des Myoviridae vers des phages de la famille des Siphoviridae après un traitement par quinolone ¹²¹. En 2013, Modi *et al.* ¹⁰⁵ ont publié une étude sur l'impact des antibiotiques sur le phageome intestinal des souris. Les auteurs ont montré que le traitement antibiotique (bêta-lactame et quinolone) a augmenté le nombre de gènes de résistance aux antibiotiques portés par les phages. Ce travail a lancé le débat sur l'implication des phages du phageome intestinal dans la propagation des gènes de résistance aux antibiotiques sous pression de sélection des antibiotiques. Plusieurs études ont observé une augmentation des gènes de résistance aux antibiotiques dans le phageome intestinal après un traitement antibiotique (i) chez des animaux ^{105,117}, (ii) chez des patients hospitalisés recevant des antibiotiques ¹¹⁹, (iii) chez des volontaires sains recevant des antibiotiques ^{120,121}. Cependant, les études menées chez l'homme ne portent que sur un très petit nombre de sujets. Les données publiques disponibles sur les phageomes ont été réanalysées avec un seuil bio-informatique plus restrictif et aucun ou très peu de gènes de résistance ont été retrouvés ¹³³, mettant en doute le rôle des phages dans la transmission des gènes de résistance aux antibiotiques.

Afin de mieux comprendre la complexité du microbiote intestinal et plus précisément l'impact des antibiotiques sur le phageome intestinal, nous avons analysé les selles de l'étude clinique décrite précédemment. Toute la méthodologie d'analyse du phageome intestinal des 22 volontaires ayant reçu du céfotaxime ou de la ceftriaxone, de l'isolement des phages à l'analyse bio-informatique, a été conduite selon notre précédente étude méthodologique publiée ¹⁸².

Les objectifs de cette étude sont (i) la comparaison du phageome intestinal des différents individus avant la prise d'antibiotique, (ii) l'analyse de l'impact de la ceftriaxone et du céfotaxime sur la diversité du phageome intestinal, (iii) la comparaison

des données de 16S bactériens retrouvés dans l'étude précédente avec les données de phages et (iii) le rôle des phages dans le transfert de gènes de résistance aux antibiotiques.

3.3.2 Article en préparation

Impact of antibiotics on human gut phageome

Camille d’Humières^{1,2,3*}, Marie Touchon³, Sara Dion⁴, Remi Denise³, Christiane Bouchier⁵, Laurence Ma⁵, Erick Denamur⁶, Eduardo P. C. Rocha³, Predires group*

¹IAME, UMR 1137, INSERM, Université Paris Diderot, Sorbonne Paris Cité, Laboratoire de Bactériologie, Hôpital Bichat, AP-HP, Paris, France.

²Ecole doctorale Frontières du vivant, Université Paris Diderot, 75013 Paris, France.

³Microbial Evolutionary Genomics, Institut Pasteur, CNRS, UMR3525, Paris, 75015, France.

⁴IAME, UMR 1137, INSERM, Université Paris Diderot, Sorbonne Paris Cité Paris, France.

⁵Institut Pasteur, BIOMICS, 25-28 rue Dr Roux, Paris, 75015, France.

⁶IAME, UMR 1137, INSERM, Université Paris Diderot, Sorbonne Paris Cité, Laboratoire de Génétique Moléculaire, Hôpital Bichat, AP-HP, Paris, France.

*Predires group: France Mentré, Xavier Duval, Antoine Andremont, Bruno Fantin, Laurent Massias, Sarah Tubiana, Charles Burdet, Nathalie Grall, Antoine Andremont, Estelle Marcault, Dusko Ehlich, Emmanuelle Lechatelier, Eduardo Rocha, Marie Touchon, Erick Denamur, Antoine Bridier Nahmias, Mélanie Magnan, Olivier Tenaillon, Jimmy Mullaert.

Impact of antibiotics on humans gut phageome

Camille d'Humières^{1,2,3*}, Marie Touchon³, Sara Dion⁴, Remi Denise³, Christiane Bouchier⁵, Laurence Ma⁵, Erick Denamur⁶, Eduardo P. C. Rocha³, Predires group*

¹IAME, UMR 1137, INSERM, Université Paris Diderot, Sorbonne Paris Cité, Laboratoire de Bactériologie, Hôpital Bichat, AP-HP, Paris, France.

²Ecole doctorale Frontières du vivant, Université Paris Diderot, 75013 Paris, France.

³Microbial Evolutionary Genomics, Institut Pasteur, CNRS, UMR3525, Paris, 75015, France.

⁴IAME, UMR 1137, INSERM, Université Paris Diderot, Sorbonne Paris Cité Paris, France.

⁵Institut Pasteur, BIOMICS, 25-28 rue Dr Roux, Paris, 75015, France.

⁶IAME, UMR 1137, INSERM, Université Paris Diderot, Sorbonne Paris Cité, Laboratoire de Génétique Moléculaire, Hôpital Bichat, AP-HP, Paris, France.

*Predires group: France Mentré, Xavier Duval, Antoine Andremont, Bruno Fantin, Laurent Massias, Sarah Tubiana, Charles Burdet, Nathalie Grall, Antoine Andremont, Estelle Marcault, Dusko Ehlich, Emmanuelle Lechatelier, Eduardo Rocha, Marie Touchon, Erick Denamur, Antoine Bridier Nahmias, Mélanie Magnan, Olivier Tenaillon, Jimmy Mullaert.

Abstract

The microbiota of the human gut sometimes called «the second brain» is a very complex and rich community composed of bacteria, archaea, protozoa, fungi and viruses, coexisting in equilibrium. The study of the dynamics and stability of the human gut microbiota is essential to understand its role in disease, diagnosis, treatment and prevention. Bacteria and their viruses (bacteriophages) are the most important members of this community. In the recent years bacteriophages from gut have started to receive attention, contribution of phages to intestinal microbiota ecology and their effect on human host are just beginning to be highlighted.

In this work, we studied the impact of antibiotics on the human gut phageome. We compared two third generation cephalosporins with distinct pharmacokinetic characteristics, ceftriaxone having higher biliary elimination than cefotaxime. A prospective, monocenter, open, randomized clinical trial in 22 healthy volunteers treated by the intravenous route either with ceftriaxone (1g/24 hours) or cefotaxime (1g/8 hours) for 3 days. The two antibiotics lead to a perturbation of the gut phageome with a stabilisation at day 30. But in terms of phage diversity, the two groups differ: in the ceftriaxone (higher biliary elimination) group there is a loss of phage diversity after the treatment, leaving a few phages that are predominant, a majority of which belongs to podoviridae family. In this study, using stringent bioinformatics cut-off to detect antibiotic resistance genes, phages don't seem to be a vector of antibiotic resistance genes under antibiotic pressure.

Introduction

The microbiota is the set of microorganisms (bacteria, archaea, eukaryotic and prokaryotic viruses, parasites and fungi) that live in a specific environment.. In humans there are different microbiota: oropharyngeal, pulmonary, vaginal, skin and intestinal. The intestinal microbiota is the most large of them with 10^{12} to 10^{14} microorganisms ¹. It plays a major role in (i) acquisition and maintenance of digestive functions (ii) the immune, intestinal and peripheral systems ² (iii) maintaining the natural barrier of mucous membranes (iv) acting against the colonization of the digestive tract by pathogens ². Dysbiosis or disruption of the balance of the intestinal microbiota may be due to various factors such as: unbalanced diet, medication (e.g. antibiotics), stress or radiation. Dysbiosis can have significant functional consequences and may be involved in many digestive diseases (chronic inflammatory bowel diseases, colorectal cancer, etc.) but also in obesity and autism³.

Bacteria have received most of the attention in human microbiome studies. Recent works have also highlighted the role of fungi, parasites and microbial eucaryota that constitute a smaller but important part of the gut microbiome ^{4,5}.

The community of bacteriophages in the intestinal microbiota is called the gut phageome. In recent years, extensive metagenomic evidence has shown that the phageome is very diverse and little recognized component of the intestinal microbial ecosystem. The vast majority of newly characterized phage sequences have no homologue in public DNA databases. In addition, many of the phages detected in metagenomic studies are not related to the currently identified phages or have considerably diverged from known phages⁶ making analysis of the intestinal phageome difficult. Thanks to microscopic observations, between 10^8 and 10^{12} phage particles per gram of stool were observed ⁷. The most abundant phages are thought to be the double-

stranded DNA phages of the order Caudovirales (family Myoviridae, Podoviridae and Siphoviridae) and the single-stranded DNA phages of the Microviridae family⁸. These families of phages infect various bacterial taxa in the intestine such as Firmicutes, Bacteroidetes, Proteobacteria and Actinobacteria. In addition to DNA phages, the intestine may also contain RNA phages, which appear to be transient members and come from food sources⁹. RNA phages are not included in intestinal phage studies, but given the high proportion of uncharacterized phages in the intestine, the study of DNA and RNA phages in the microbiota would be justified. Intestinal phageome has previously been described as individual specific^{10,11} with however a common base present in healthy individuals, called "healthy gut phageome"¹².

Antibiotic treatments disrupt the balance between the different bacterial groups found in the intestinal microbiota. Changes in the quantities and taxonomic composition of the intestinal microbiota following antibiotic administration have been reported regularly^{13,14}. This contributes to an imbalance, dysbiosis, which can lead to negative effects such as *Clostridioides difficile* infection or the emergence of antibiotic-resistant bacteria¹⁵. Currently the impact of an antibiotic on the intestinal microbiota depends on its antibacterial spectrum and route of elimination (and thus on its concentration in the digestive tract). A major effect of antibiotics is to alter the protection effect provided by anaerobic bacteria in the intestinal microbiota¹⁶. This leads to an increase in bacterial density of normally subdominant bacteria such as beta-lactam-resistant enterobacteriaceae or vancomycin-resistant enterococci¹⁵. This increase leads to a higher risk of digestive translocation and thus infection^{16,17} or spread in the environment. Observations of antibiotic-resistant bacteria in the faecal flora under pressure of antibiotic selection are numerous, in animals or humans with conventional culture analysis or metagenomics^{18,19}.

The impact of antibiotics on the human intestinal phageome is just beginning to be explored, two studies have shown a modification of the phageome under antibiotic treatment in a very small number of healthy volunteers^{20,21}. More precisely in one individual, a shift in abundance from Myoviridae to Siphoviridae was found after antibiotic treatment with quinolone²⁰. It is known that antibiotic perturbation might also affect phage populations and promote their propagation rates²². In 2013, Modi *et al.*²³ published a study showed that antibiotic treatment (beta-lactam and quinolone) leads to increase the number on antibiotic resistance genes encoded by phages. It launched the debate on the involvement of phages in antibiotic resistance genes (ARG) propagation in gut phageome under antibiotic selection pressure. Several studies have observed an increase in antibiotic resistance genes in the intestinal phageome after antibiotic treatment (i) in animals^{23,24}, (ii) in hospitalized patients receiving antibiotics²⁵, (iii) in healthy volunteers receiving antibiotics^{20,21}. However, studies carried out in humans include a very small number of subjects. Public available phageome data have been reanalysed with threshold more conservative and none or very few resistance genes have been found, questioning the role of phages in the transmission of antibiotic resistance genes²⁶. This paper warns about the bioinformatics thresholds used for the search for antibiotic resistance genes.

In this study we have worked on the stool samples of a recently published clinical study²⁷. It was a prospective randomized clinical trial performed in healthy volunteers receiving a clinical course of ceftriaxone or cefotaxime for 3 days for the comparison of their impact on the gut microbiota using phenotypic analysis and 16S rRNA gene profiling. Ceftriaxone and cefotaxime are 3rd generation cephalosporins that share the same indication but there have different pharmacokinetics characteristics: ceftriaxone is administered once daily and has a 40% biliary elimination, while cefotaxime needs

thrice daily administrations but has a 10% biliary elimination. It has been shown, in gut microbiota, that emergence of antibiotic resistant bacteria was not significantly different between the two groups and that both antibiotics reduced the counts of total Gram-negative enteric bacilli and decreased bacterial diversity, without significant difference between groups. In this work the gut phageome of these healthy volunteers was studied before and after the antibiotic treatment in order to (i) study the ecology of phages and the impact of antibiotic treatment on the phageome (ii) study the difference of the two antibiotics (cefotaxime/ceftriaxone) with different pharmacokinetics characteristics on the phageome impact (iii) study the phageome antibiotic resistance genes content and his potential implication in bacteria antibioresistance.

Methods

Study design and samples studied

A prospective, open label randomized clinical trial from March 2016 to August 2017 was conducted in adult healthy volunteers in Claude Bernard Hospital (Paris, France). All details about this clinical study were in the Burdet *et al.* paper ²⁷. In brief, 22 healthy volunteers without a hospitalization in the previous 6 months and without an antibiotic treatment in the subsequent 3 months were recruited. There were randomized (1:1 ratio) to receive from day 1 to 3, either ceftriaxone (1g/24 hours) or cefotaxime (1g/8 hours) in intravenous infusions.

For each volunteer a total of 13 faecal samples was obtained at days -15±2, -6±2, -1, 1, 2, 3, 4, 7, 10±1, 15±1, 30±3, 90±7 and 180±7). For the analysis of the phageome, we studied 6 samples per participant, preferably at days -15, -1, 4, 10, 30, 180. Due to technical problems at the time of phage isolation or insufficient amount of extracted DNA, in some case other samples were analyzed: if samples of days -15 or days -7 failed we took sample from day -7, if samples of days 4 or days 10 failed we took sample from day 7, if samples of days 30 failed we took sample from day 15 and if samples of days 180 failed we took sample from day 90. Table 1 resume the fecal samples studied in this paper.

Phage isolation and DNA sequencing

Phage isolation was performed with a step of polyethylen glycol (PEG) concentration, as recommended in this paper ²⁸. One gram from faecal samples were weighed and homogenized in 40 ml of phosphate buffered saline (PBS) (Sigma-Aldrich). The sample was then agitated with a mechanic laboratory agitator during 1 hour at 4°C, centrifuged at 17,000 g for 5 min and the supernatant was filtered at 2 µm and 0.45 µm. Phages were

then concentrated using a PEG method. One molar solid NaCl and 10% (v/v) PEG 8000 (Sigma) were dissolved into the filtered culture fluid and incubated overnight at 4°C as recommended for a constant and stable precipitation. The solutions with the phages were pelleted by centrifugation at 5,250 g for 1h at 4°C and re-suspended in 500 µL of SM buffer (NaCl 100mM, MgSO₄·7H₂O 8mM, Tris-Cl 50mM). Samples were treated with 10 U/ml of DNase (Sigma) for 30 min at 37°C followed by 10 min at 65°C to stop the reaction. DNA was then extracted using the commercial kit "Phage DNA extraction" (Norgen biotek Corp, Ontario, Canada). DNA was purified on a sephadex column (Sigma-Aldrich, France), measured with Qubit dsDNA HS Assay kit (Thermo Fisher Scientific), and sequenced with the Illumina HiSeq2500 PE_250 bases method using the Kit Nextera XT with an input of 1ng DNA.

Bioinformatic analysis: Sequence quality control, contig assembly, and annotation

The sequence reads of the six samples of the same volunteers were pooled. They were trimmed to remove the Illumina adapters and to remove reads with low quality using Atropos (v1.1.18)²⁹ with parameters: atropos trim -m 100 -q 20,20 -trim-n. We reduced the redundancy of the reads with the Khmer software (v2.0)³⁰, using a k-mer size of 20 and a median k-mer coverage level of 20. The resulting reads were assembled using SPAdes 3.10.1³¹ with -meta option. At this step we obtained 22 pools of contigs (1 pool per volunteer). Gene prediction was done using Prodigal (v2.6.2)³² with -p meta option. We excluded genes lacking start and stop codons. In order to focus our analysis on contigs sufficiently large to study genetic contexts, we excluded contigs with less than 3 open reading frames (ORFs).

Bioinformatic analysis: alpha and beta diversity, richness

In order to compare all samples with each other, the 22 pool of contigs 'volunteers were concatenated and clustered with CD-HIT (v4.7, option `-d 0, -g 1, -c 0.95, -aS 0.9`)³³. Then the largest contig of each cluster was selected to constitute a "all contigs reference catalog". We then mapped the samples' reads on this "all contigs reference catalog" using bowtie2 (v2.2.9 `--local --sensitive-local` options)³⁴ and exploit SAM files with samtools (v1.3.1 with the following commands: `views, sort, index, idxstat`)³⁵. As result, we obtained a matrix representing the number of reads of a sample (columns) mapping each contig reference catalog (rows) in the dataset. We computed the pairwise dissimilarity between the samples using the Bray-Curtis compositional dissimilarity measure between columns of the rarefied count matrix. This was done with the skbio python package (v4.2) (<http://scikit-bio.org/>). The count matrix was rarefied to 3 311 186 reads with the "rarefy" function of the vegan package in R³⁶.

For analysis per volunteer, a count matrix was performed for each volunteer with his pool of contigs. All the matrix count were rarefied at 3 003 762 reads. Shannon diversity index, Simpson diversity index and Bray-Curtis dissimilarity were calculated with the skbio python package (v4.2). The richness of a sample was assessed using the number of contigs present in the rarefied sample the number of major contigs was also calculated: it represents the number of contigs representing 80% of all reads of the samples.

Contig classification

Taxonomic attribution of the contigs was made with the algorithm developed in²⁸. Briefly The algorithm for taxonomic attribution of the contigs permitted to classify a contig as : "phage with family", "putative phage", "bacterial", "contamination", "non-attributable (NA)" according to weighted gene repertoire relatedness (wGRR) strategy,

Virsorter software, HMM bacterial markers, 16S profile and NCBI database. In the ceftriaxone group, at day 4, a part of major contigs (i.e. those representing at least 10% of the phageome), were analyzed in detail: we computed wGRR between all pairs and between them and the 2,493 complete bacteriophage genomes from GenBank Refseq (<ftp://ftp.ncbi.nih.gov/genomes>; last accessed April 2019), and the Azobacteroides phage APO17903 presented in ³⁷ but not present in Refseq databases. The wGRR network was visualized with NetworkX python library.

Bioinformatic analysis: antibiotic resistance genes search

Different ARG databases were used for this step: ARG-ANNOT³⁸, CARD³⁹, Resfinder⁴⁰, ResfinderFG⁴¹ and Mustard⁴². Antibiotic resistance genes were searched on the “all contigs reference catalog” against all the ARG databases mentioned above with BLASTn using conservative thresholds: a minimum of 80% sequence identity and 50% of coverage over the ARG. The sequences identified as putative ARGs were extracted. We then clustered ARG sequences using CD-HIT (v4.7, option -d 0, -g 1, -c 0.95, -aS 0.9, -G 0) and removed redundancy by selecting the largest ARG of each cluster. A putative ARG catalogue was created with these sequences. The reads of each sample were mapped on the ARG catalogue using bowtie2 (v2.2.9 --local --very-fast-local options) ³⁴ and exploit SAM files with samtools (v1.3.1 with the following commands: views, sort, index, idxstat) ³⁵.

Missing data handling

In order to make statistics, we managed the missing data with these rules:

- Missing data at day -15 (n=2) should be replaced by data at day -7 when available (n=1).

- Missing data at day -1 (n=2) should be replaced by data at day -7 when available (n=1).
- Missing data at day 4 (n=2) should be replaced by data at day 7 when available (n=1).
- No imputation for missing data at day 30 nor day 180.
- Volunteers for which no data are available at day -14 and day -7 and day -1 will be excluded from statistical analyses (=1 volunteer in the ceftriaxone group)

Statistical methods

For Bray-Curtis distance analyses between all samples, an ANOSIM analysis using skbio python package (v4.2) option `skbio.stats.distance.anosim` was performed on the Bray-Curtis distance matrix count. For each volunteer we computed the variation, from baseline (day -1) and different points after treatment, of this parameters: Bray-Curtis dissimilarity, Shannon diversity index, Simpson diversity index, global richness, major contigs and proportion of ARG. For global richness, major contigs and proportion of ARGs the Log10 value was calculated . Wilcoxon analyses were performed in order to compare the two groups of treatment and these different parameters at different time (change from baseline (day -1) and: day 4, day 10, day 180). For the same parameters, Area Under the Curve (AUC) between day -1 and day 10 (AUC₋₁₋₁₀) was calculated using the trapezoidal method. Wilcoxon test was used to compare AUC₋₁₋₁₀ of the different treatment groups for the different parameters.

Results

The individuality of gut phageome is kept under antibiotic therapy

For the phageome analysis of healthy volunteers before and after taking antibiotics (cefotaxime or ceftriaxone) we had planned to study 6 samples per individual on the following dates: days -15, -1, 4, 10, 30, 180. Due to technical problems at the time of

phage isolation or insufficient amount of extracted DNA, the initially planned samples could not all be processed. Phage isolation and DNA sequencing were performed on 127 samples, details were described in Table 1. Sequencing led to 17 769 047 reads in average [min=3 354 542, max=40 170 834] and the average number of contigs by individual is 25673 [min=12699, max=39604]. All contigs of all healthy volunteers were put together and redundancy was removed in order to create a “all contigs reference catalog” of 297 037 contigs with an average size of 24 898 bp and a median size of 14 675 bp.

In order to compare the global phageome composition within the same individual and between individuals, a pairwise Bray-Curtis dissimilarity was calculated between all samples. The pairwise Bray-Curtis dissimilarity heatmap (Figure 1) shows a clear clustering of the samples by individual. Concerning the overall composition of the samples, it is closer between 2 samples before and after antibiotic treatment of the same individual than between two samples of two different individuals. Interestingly, the importance of the category “individual” is more important than the introduction of antibiotic treatment (Anosim, 999 permutations, $p=0.001$): at the end the sample still has a higher resemblance with the early samples of the individual than the late samples of other individuals. This brings us back to a notion already described in the literature^{10,11} and confirmed here that the phageome is individual-specific even with exposure to antibiotics.

Perturbation of gut phageome with antibiotics and progressive return to initial state.

For each volunteer, the Bray-Curtis distance between the baseline sample (day -1) and other points such as: day -15, day 4, day 10, day 30, day 180 has been calculated (Figure 2A). Before taking any antibiotic we can note that the median of the Bray-Curtis distances (between day -15 and day -1) for the sampling set is 0.50. Both antibiotic treatments have in common the tendency to increase the Bray-Curtis distance between the baseline sample and the day -4 sample ($p=0.0005$, paired Wilcoxon test). This testifies to the global disruption of the phageome by antibiotics. Ten days after treatment the disturbance is still visible and significant ($p=0.004$, paired Wilcoxon test). One month after treatment the phageome has regained an overall composition closer to initial state. Six months after the end of treatment we can observe a statistically significant modification of the phageome due to the effect of time, during 6 months there may have been factors that modified phageome composition (e.g.: feeding)

Taking antibiotics alters the overall composition of the intestinal phageome and makes it even more singular (Figure 2B). Before the antibiotic treatment there are about 25% of the phage contigs that are present in at least 50% of the volunteers, and this number is decreasing to 8.4% after antibiotic treatment and goes back to 15% six months later. By mirror effect we also observe that the number of phage contigs found in one unique volunteer increase from 13.7% before the antibiotic treatment to 20.6% after.

Ceftriaxone antibiotic tends to decrease more than cefotaxime the gut phageome richness and alpha-diversity

The impact of ceftriaxone and cefotaxime on the faecal phageome was analysed based on the analyses of alpha diversity (Shannon and Simpson index) and richness, Figure 3A,B representing evolution of the Shannon and Simpson index and Figure 3C,D representing evolution of the overall richness and major contigs by volunteers. It can be observed that in the group treated with ceftriaxone there is a tendency for a decrease in diversity and richness indices in post-treatment, with an over-representation of some contigs as put in evidence in Figure 3D. This phenomenon is not observed for the cefotaxime group. In order to highlight the effect of antibiotics on the samples, for each volunteer, each sample was compared to a pre-treatment sample (baseline) for the different alpha diversity and richness parameters. Figure 4 shows us the evolution of the change rates from baseline (day -1) of all the parameters and we can observe a greater phageome disruption in ceftriaxone group with a decline in diversity and richness after antibiotic treatment. Statistical comparison of the change in diversity (Shannon and Simpson) and richness (major contigs) before and after treatment (J4, J10, J180) between the two treatment groups are not significant (Table 2). Those for the Simpson index tend to be significant, that is not surprising because Shannon index diversity is more sensitive to the global richness whereas the Simpson index diversity is more sensitive to the predominance of some species.

Contig content description

All the contigs were studied and classified into five categories: "phage with family", "putative phage", "bacterial", "contamination", "non-attributable (NA)" as described in the methods section. Figure 5A and 5B represent the repartition of these categories in

terms of time and treatment. No major fluctuations were observed in terms of relative abundance of each category regarding time and antibiotic treatment. On average there are 43.3% reads mapping putative phage contigs, 25.5% reads mapping phage with family contigs, 19.2% reads mapping bacteria contigs, 12.0% reads mapping NA contigs and 0.1% of other contamination. We found a majority of phage or putative phage reads, which is adequate. When we focus on “phage with family” group, the double stranded DNA phages belonging to Caudovirales family (Podoviridae, Siphovirodae and Myoviridae) are largely dominant (Figure 5C and 5D). After antibiotic treatment, only on ceftriaxone group there is a significant increase of proportion of phage contigs belonging to Podoviridae family (Figure 5F) and a decrease of proportion of phage contigs belonging to Siphoviridae family between day 4 and day 0 (Figure 5E). This phenomena is not observed in the cefotaxime group. Difference between the proportion of phage contigs belonging to Podoviridae family in the two groups are significant from day 4 to day 15 after antibiotic treatment, at day 30 we observe a return to balance (Figure 5F). As described above, in the ceftriaxone group there is a drastic loss of diversity after the antibiotic treatment with few contigs very highly represented. Indeed, at day 4, in 9 volunteers, there are 18 contigs among the major contigs that represents more than 10% of the sample content. These contigs belong to following categories: putative phage (13/18), phage with family (4/18), non attributed (1/18). This means that for almost all patients in the ceftriaxone group, in the post-treatment faecal sample there are on average 2 contigs which represent the vast majority of the phageome (more than 50%). Figure 5G shows us the dynamic of this 18 contigs with a peak at day 4 and then a progressive return to baseline. Network based on wGRR score between 18 major contigs pairs and between 18 contigs and Refseq phages (Figure 6A) shows us that one contig is

very close to the crAssphage and that there are 2 networks around 2 known phages: Cellulophaga phage phi 4 and Azobacteroides phage.

Antibiotic resistance genes

In order to study the potential role of phages in the transfer of antibiotic resistance genes under antibiotic selection pressure, we searched for antibiotic resistance genes on of the "all contigs reference catalog". We used five databases: Resfinder, Argannot, ResfinderFG, Mustard and CARD. We found 2544 non-redundant ARG (Figure 7B), a majority (87%) were only found by Mustard databases, that is not surprising because it is a new database⁴² with ARG specifically found in gut microbiome. The relative abundance of these genes was calculated in each sample (Figure 7A). In the ceftriaxone group there is a decrease of ARG proportion due to the global decrease of phage contigs, and there isn't ARG on the major contigs present. In the cefotaxime group we observed a very small decrease of ARG proportion after treatment. In order to be more exhaustive, we focused on phage contigs defined by contigs tagged phage with family and putative phage but with a good Virsorter prediction. We found 159 contigs with an ARG. Description of origin database and type of these ARG are in figure 7D, 6E. A majority (71%) are only found in mustard database and were genes conferring resistance to trimethoprim and sulfonamide. When we observed the relative abundance of these 159 ARG in all samples, (Figure 7C) we observed the same tendency as all the ARG: there was no increase of ARG after antibiotic treatment. Bacterial reads in samples are a mix of bacterial contamination and generalized transduction. In order to appreciate generalized transduction, we studied the bacterial read proportion in each healthy volunteer over time (Figure 7F). Before antibiotic treatment there was a relative abundance of bacterial read of 20%. After treatment, in the ceftriaxone group there is

decrease of relative abundance of bacterial reads due to phage reads prevalence. In the cefotaxime group there is a small increase of relative abundance of bacterial reads, compatible with a generalized transduction effect. Antibiotic resistance genes were then studied in the bacterial fraction (Figure 7G), a decrease of ARG was observed after antibiotic treatment. It means that the potential induction of generalized transduction after cefotaxime antibiotic treatment isn't concomitant with an augmentation of ARG in the bacterial fraction. Therefore, there is no evidence of an increased frequency of antibiotic resistance after an antibiotic treatment.

Discussion

This work studies, for the first time, the impact of two third-generation-cephalosporins on the intestinal phageome in healthy volunteers in a randomized controlled study. The two antibiotics lead to a perturbation of the gut phageome with a stabilisation at day 30. But in terms of phage diversity, the two groups differ : in the ceftriaxone group there is a loss of phage diversity after the treatment, leaving a few phages that are predominant, a majority of which belongs to podoviridae family.

However, some technical points need to be highlighted. As shown in our previous study on phage isolation, concentration and DNA extraction (reference), ²⁸. PEG concentration was found to be the best technique to isolate phages in terms of bacterial contamination, phage diversity and cost. This technique was applied to the 131 samples we collected, but for some samples there was an insufficient amount of extracted phage DNA. When this happened, we did not use a DNA amplification technique because of the biases described before, we used another sample collected at a close time point. This

problem had not occurred in our previous study but we had processed a much smaller amount of samples. This may be due to the quality of the feces, it may be more difficult to isolate phages on stool with mucus. PEG isolation technique has shown a good reproducibility with a Bray Curtis distance of 0.08 between two replicates²⁸. In this work, we did not include replicates but each volunteer had 2 samples before antibiotic treatment over a period of 15 days. The median BC distance between these two samples was 0.5, it may seem high but it is comparable to Manrique's study that showed a BC distance of 0.47 between two feces samples of the same individual. Relative abundance of bacterial reads was higher than expected. Before treatment, the mean of relative abundance bacterial reads was 20.5% \pm 14%. We have shown that the deeper the DNA is sequenced, the more the proportion of sequenced bacterial DNA increases²⁸. In our previous work, only one sample had been sequenced very deeply (80 millions of reads) and it contained 10% of bacterial reads. The larger number of samples studied here inevitably increased the variability of this parameter.

As shown in our previous study²⁸ gut phageome is individual specific. The impact of antibiotics from the quinolone family on gut phageome has already been studied in two papers^{20,21} but on a very small number of individuals (one and two respectively). In this work, we demonstrated that the weight of the individual is stronger than the disturbance caused by the antibiotic. Antibiotic treatment changes the composition of gut phageome, the BC distance decreases progressively until day 30. Between day 30 and day 180, the Bray-Curtis distance increases again, this is easily explained by the time factor and the set of intra-individual variability that it can generate. In a study, on the same cohort of healthy volunteers²⁷, we showed, using a 16S analysis, that there was no difference between the cefotaxime and ceftriaxone groups: both altered the diversity of the bacterial intestinal microbiota with a return to baseline at day 30. Regarding the

phage intestinal microbiota, we observed a drastic decrease in the diversity of the gut phageome in more than half of the volunteers in the ceftriaxone group, this is not statistically significant due to the low number of healthy volunteers and the variability between groups. Predominant contigs, after ceftriaxone treatment, seem to belong to the podoviridae family and one is the crAssphage. Observations on the impact of the two antibiotics on microbial diversity are not the same depending on the angle of study. On the bacterial front, we did not observe any difference. However, when focusing on phages, a difference seems to appear. This is not surprising considering the difference in granularity of the two analyses: phages are specific to a bacterial strain, it can bring to light trends not seen at the Operational Taxon Unit level studied by the 16S.

The role of phages in the transfer of antibiotic resistance genes remains controversial. Several studies have highlighted an increase in antibiotic resistance genes found in the phage fraction after antibiotic treatment^{21,23,25} and others have rejected this phenomenon^{26,43,44}. When using stringent bioinformatic cut-offs, as recommended²⁶, we did not show an increase in antibiotic resistance genes in the phage fraction after antibiotic treatment. In order to evaluate the implication of generalized transduction in this central question, we studied the bacterial fraction of our samples. The relative abundance of bacterial reads tends to increase after taking antibiotics, it may indicate a processes of specialized transduction. But no increase in the number of antibiotic resistance genes has been found. In this study phages don't seem to be a vector of antibiotic resistance genes under antibiotic pressure.

Samples from this clinical study will also be analyzed to study the metagenome, resistome and mycobiome. Our phageome data can thus be enriched with metagenomic data in order to confirm the lifestyle of certain phages and this way, indentifying the phage host. Antibiotic resistance genes present in the metagenome can also be

compared to those of the phageome, this will help to better analyse the role of phages in transfer of antibiotic resistance.

Table

Table 1: Distribution of the samples processed for phageome analyses

Healthy volunteer number	treatment	J -15	J -7	J -1	J4	J7	J10	J15	J30	J90	J180
1	Ceftriaxone										
2	Cefotaxime										
3	Cefotaxime										
6	Ceftriaxone										
8	Ceftriaxone										
9	Ceftriaxone										
11	Cefotaxime										
12	Cefotaxime										
13	Cefotaxime										
14	Cefotaxime										
16	Ceftriaxone										
17	Cefotaxime										
18	Ceftriaxone										
20	Cefotaxime										
22	Ceftriaxone										
23	Ceftriaxone										
24	Ceftriaxone										
25	Cefotaxime										
27	Cefotaxime										
28	Ceftriaxone										
29	Ceftriaxone										
34	Cefotaxime										

Table 2: statistical table for the comparison of cefotaxime group and ceftriaxone group. Wilcoxon test were performed

	Shannon	Simpson	Major contigs (80p)
AUC day -1 day 10	P=0.32	P=0.058	P=0.121
Change from baseline day - 1 day 4	P=0.257	P=0.069	P=0.173
Change from baseline day - 1 day 10	P=0.231	P=0.011 *	P=0.181
Change from baseline day - 1 day 180	P=0.290	P=0.95	P=0.199

Figures

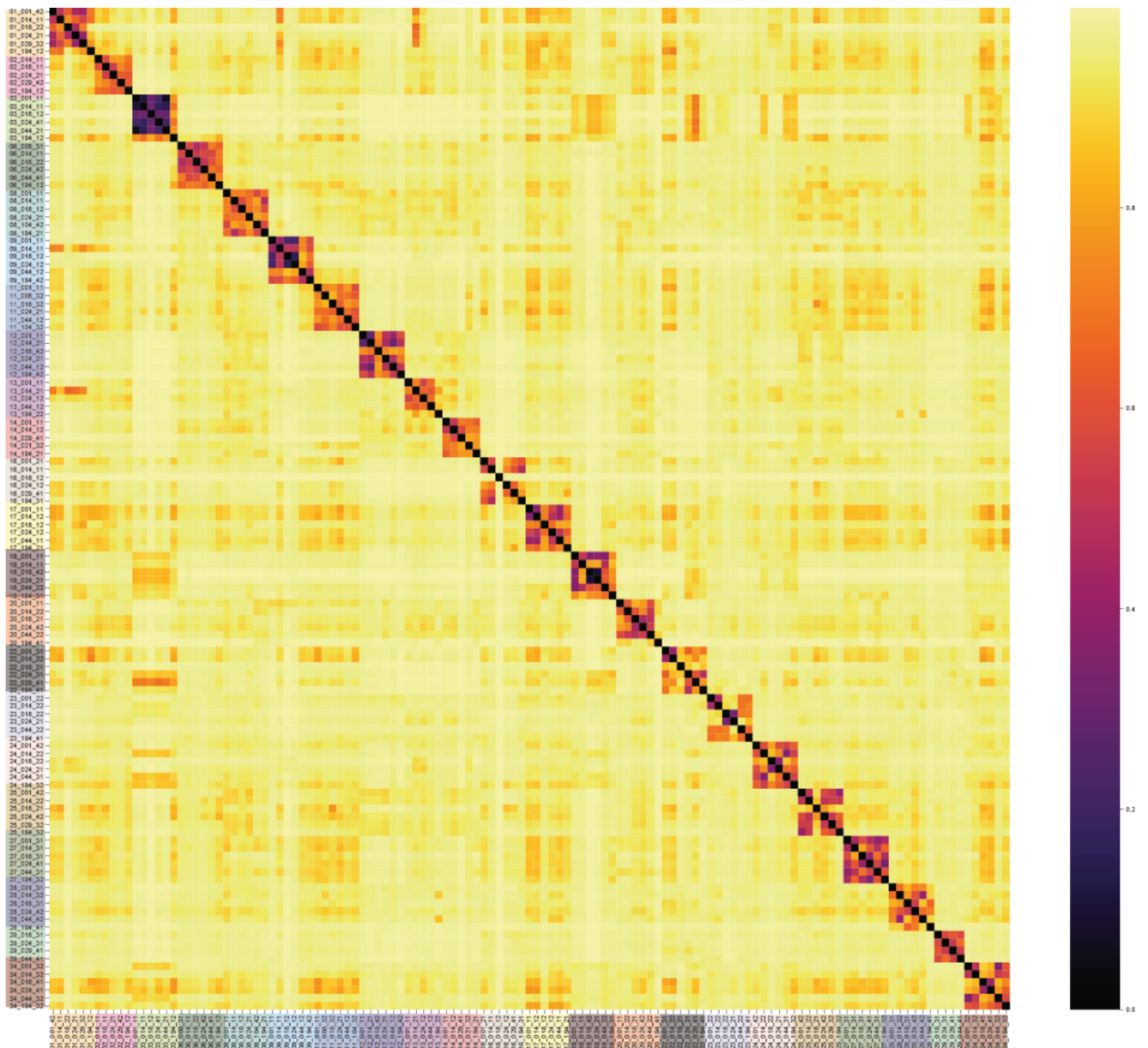


Figure 1 : Heatmap of the Bray-Curtis dissimilarity between the 127 samples of 22 healthy volunteers.

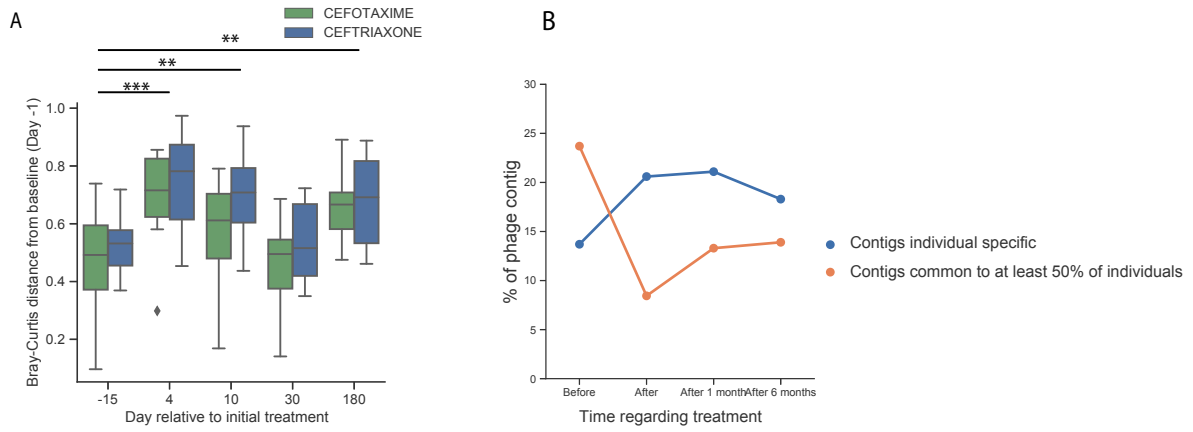


Figure 2(A): Boxplot of the distribution of the Bray-Curtis distance from day -15, 4, 10, 30 or 180 versus day -1 for the two groups (cefotaxime in green and ceftriaxone in blue). Statistical tests were performed with Wilcoxon test ($** < 0.01$, $*** < 0.001$)(B): Description of percentage of shared contigs between individual before and after (until 6 months) treatment.

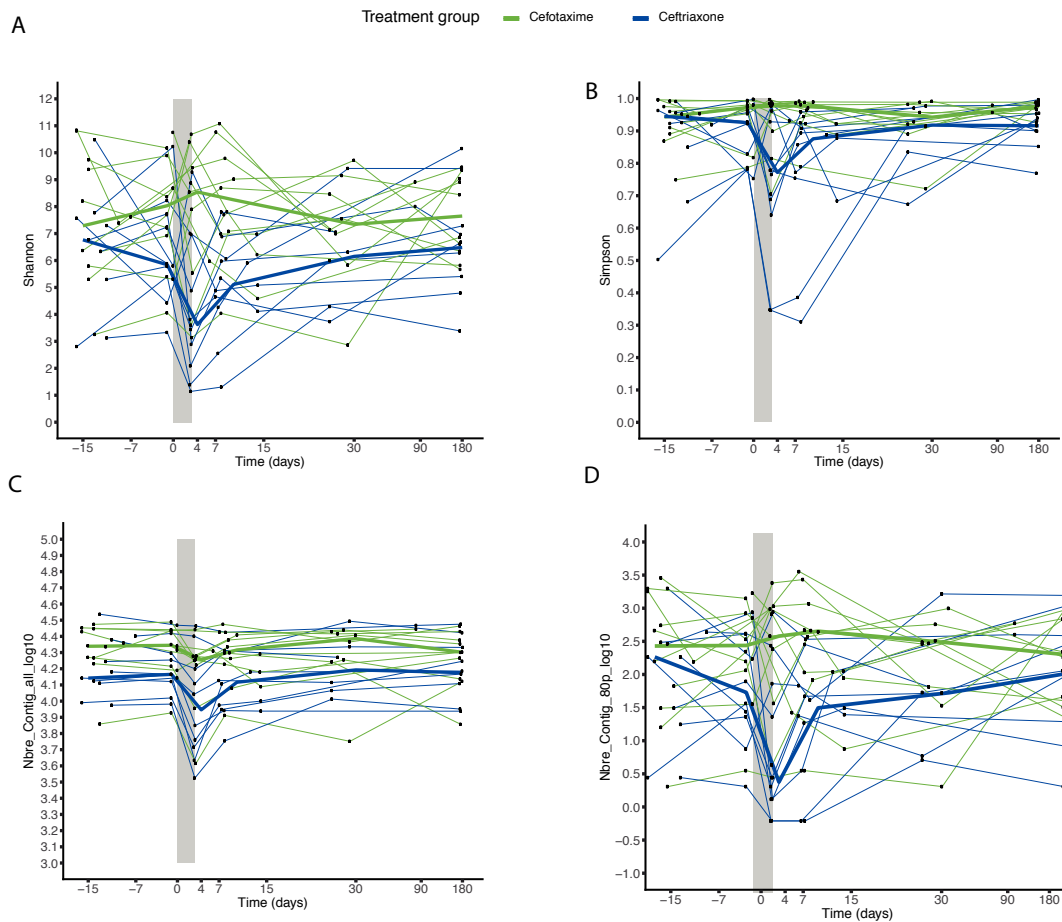


Figure 3: Evolution of different variable in faecal samples of the 22 volunteers, 11 treated by ceftriaxone in blue and 11 treated by cefotaxime in green. (A) Evolution of Shannon index (B) Evolution of Simpson index (C) Evolution of the number of total contigs (D) Evolution of the number of major contigs: number of contigs representing 80% of the reads content of the sample. The light grey zone represents the treatment period. Thin lines represent individual values, and thick lines represent the median at each time.

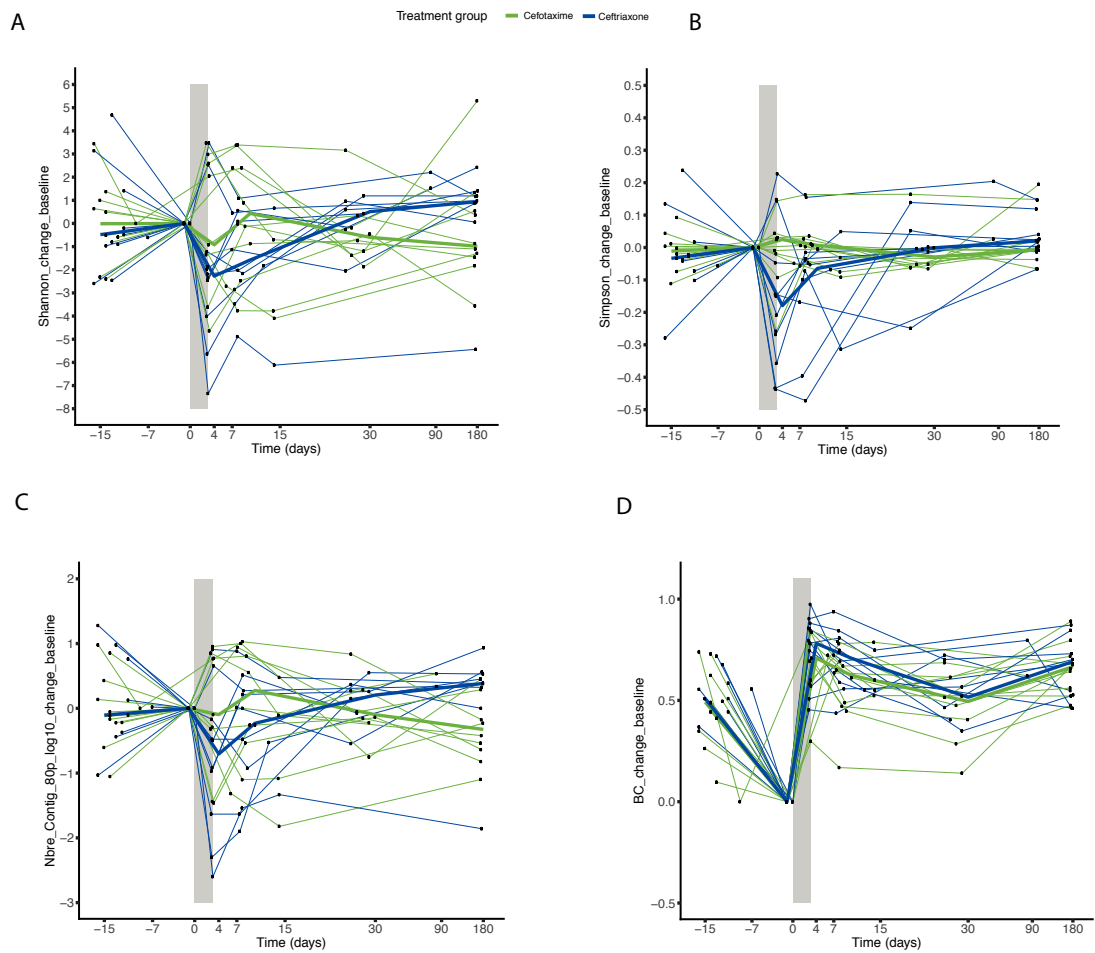


Figure 4: Evolution of the change from baseline of (A) Shannon index, (B) Simpson index, (C) the number of major contigs (D) Bray-Curtis distance. The light grey zone represents the treatment period. Thin lines represent individual values, and thick lines represent the median at each time.

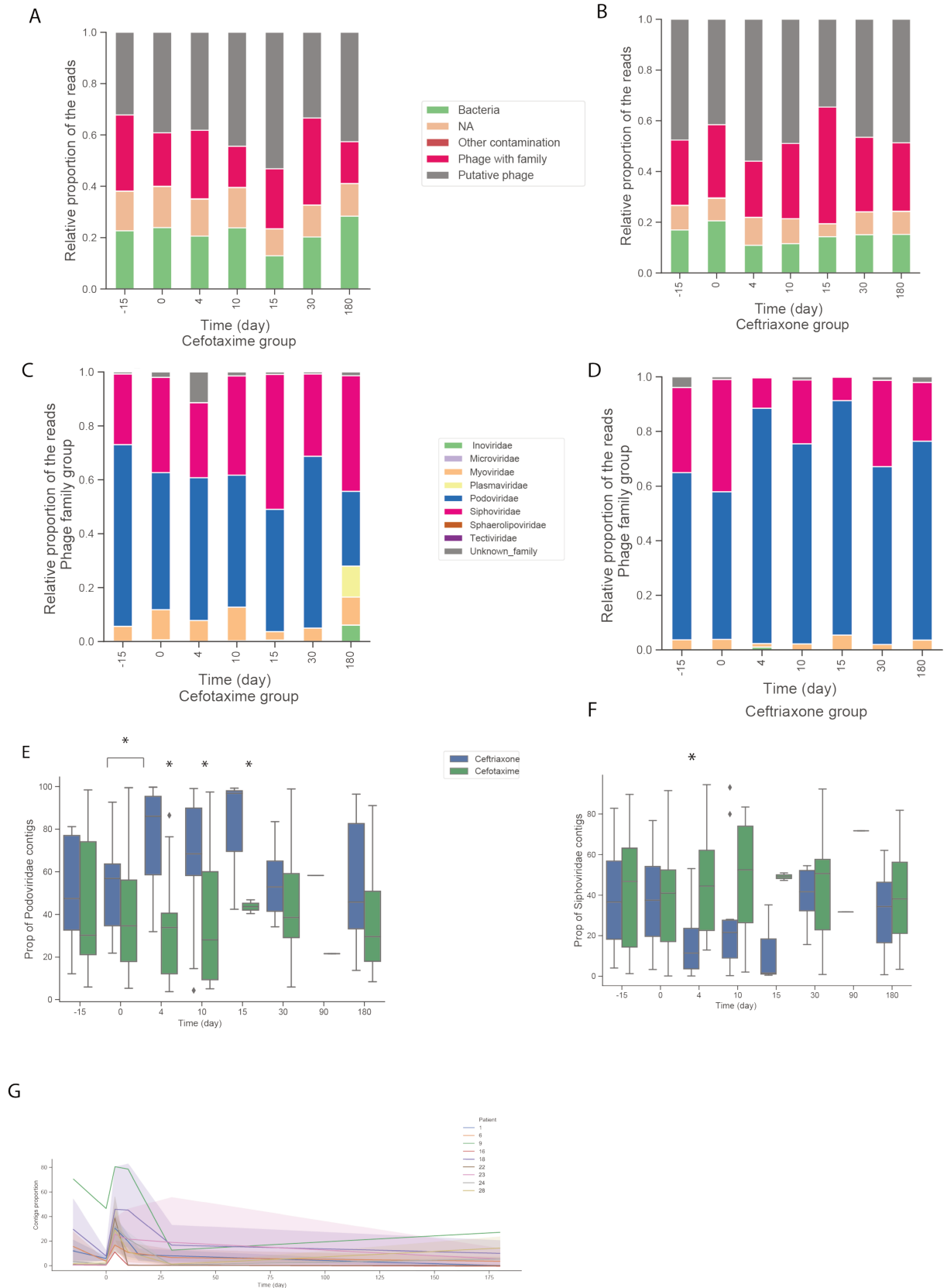


Figure 5: Phageome content. Frequency of reads mapping the different classes of contigs (“phage with family”, “putative phage”, “bacteria”, “contamination”, “non-attributable (NA)”) for samples in cefotaxime group (A) and ceftriaxone group (B). Frequency of the different family of reads from “phage with family” group for samples in cefotaxime group (C) and ceftriaxone group (D). (E) Boxplot of the proportion of reads

mapping Siphoviridae contigs (in the phage with family group) in terms of time and traitement (cefotaxime in green and ceftriaxone in blue). (F) Boxplot of the proportion of reads mapping Podoviridae contigs (in the phage with family group) in terms of time and traitement (cefotaxime in green and ceftriaxone in blue). (G) Evolution of the proportion of major contigs presents a day 4 in ceftriaxone group.

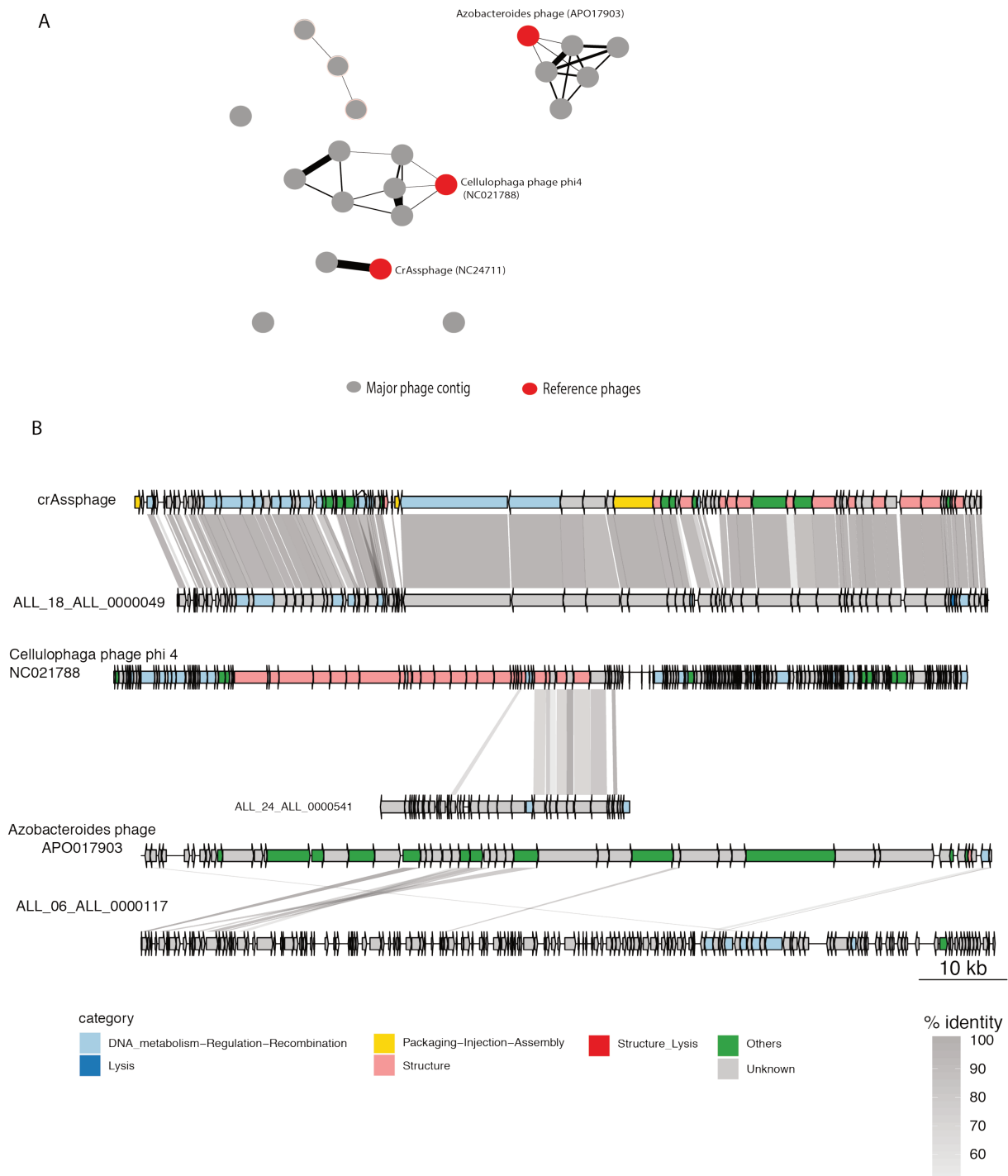


Figure 6: (A) Network based on wGRR score between 18 major contigs pairs and between 18 contigs and Refseq phages. Grey nodes are contigs and red nodes are reference phages. Edges link pairs with a wGRR > 5% were represented (the thickness of the edge is proportional to the score) (B) Genomic organization of 3 contigs and their closest reference phages. Arrows represent predicted ORFs and are oriented according to transcriptional direction. Colors indicate assigned functional categories. This figure was generated using the R package GenoPlotR v0.8.

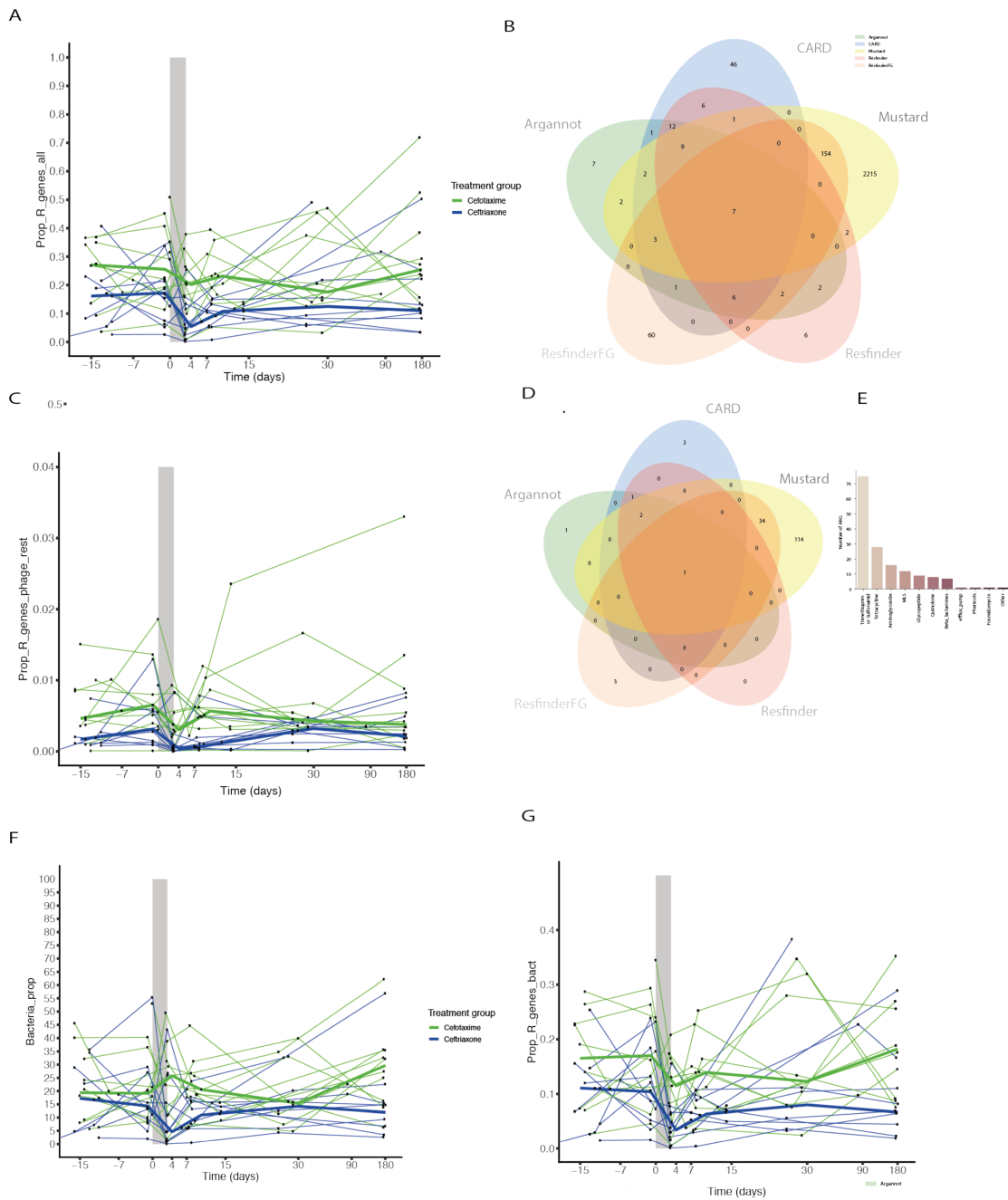


Figure 7: Phageome antibiotic resistance genes (A) : Evolution of all antibiotic resistance genes found in faecal samples of the 22 volunteers, 11 treated by ceftriaxone in blue and 11 treated by cefotaxime in green (B) Venn diagram of the non-redundant ARG found in the "all contigs reference catalog" identified with the 5 different ARG databases (C) Evolution of phage ARG (n=159) found in faecal samples of the 22 volunteers (D) Venn diagram of phage ARG found in phage contigs identified with the 5 different ARG databases (E) Description of the 159 ARG category (F) Evolution of relative abundance of bacterial reads in faecal samples of the 22 volunteers (G) Evolution of the proportion of ARG found in bacteria fraction in faecal samples of the 22 volunteers.

Bibliography

1. Sekirov, I., Russell, S. L., Antunes, L. C. M. & Finlay, B. B. Gut microbiota in health and disease. *Physiol. Rev.* **90**, 859–904 (2010).
2. Belkaid, Y. & Hand, T. W. Role of the microbiota in immunity and inflammation. *Cell* **157**, 121–141 (2014).
3. Landman, C. & Quévrain, E. [Gut microbiota: Description, role and pathophysiological implications]. *Rev Med Interne* **37**, 418–423 (2016).
4. Laforest-Lapointe, I. & Arrieta, M.-C. Microbial Eukaryotes: a Missing Link in Gut Microbiome Studies. *mSystems* **3**, (2018).
5. Hoffmann, C. *et al.* Archaea and fungi of the human gut microbiome: correlations with diet and bacterial residents. *PLoS ONE* **8**, e66019 (2013).
6. Minot, S. *et al.* Rapid evolution of the human gut virome. *Proceedings of the National Academy of Sciences* **110**, 12450–12455 (2013).
7. Hoyles, L. *et al.* Characterization of virus-like particles associated with the human faecal and caecal microbiota. *Research in Microbiology* **165**, 803–812 (2014).
8. Reyes, A. *et al.* Gut DNA viromes of Malawian twins discordant for severe acute malnutrition. *Proc Natl Acad Sci U S A* **112**, 11941–11946 (2015).
9. Zhang, T. *et al.* RNA viral community in human feces: prevalence of plant pathogenic viruses. *PLoS Biol.* **4**, e3 (2006).
10. Minot, S. *et al.* The human gut virome: Inter-individual variation and dynamic response to diet. *Genome Research* **21**, 1616–1625 (2011).
11. Reyes, A., Semenkovich, N. P., Whiteson, K., Rohwer, F. & Gordon, J. I. Going viral: next-generation sequencing applied to phage populations in the human gut. *Nature Reviews Microbiology* **10**, 607–617 (2012).
12. Manrique, P. *et al.* Healthy human gut phageome. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 10400–10405 (2016).
13. Sullivan, A., Edlund, C. & Nord, C. E. Effect of antimicrobial agents on the ecological balance of human microflora. *Lancet Infect Dis* **1**, 101–114 (2001).
14. Willing, B. P., Russell, S. L. & Finlay, B. B. Shifting the balance: antibiotic effects on host-microbiota mutualism. *Nat. Rev. Microbiol.* **9**, 233–243 (2011).
15. Bhalla, A. *et al.* Antianaerobic Antibiotic Therapy Promotes Overgrowth of Antibiotic-Resistant, Gram-Negative Bacilli and Vancomycin-Resistant Enterococci in the Stool of Colonized Patients. *Infection Control & Hospital Epidemiology* **24**, 644–649 (2003).
16. Taur, Y. *et al.* Intestinal Domination and the Risk of Bacteremia in Patients Undergoing Allogeneic Hematopoietic Stem Cell Transplantation. *Clin Infect Dis* **55**, 905–914 (2012).
17. Ruppé, E. *et al.* Relative Fecal Abundance of Extended-Spectrum- β -Lactamase-Producing *Escherichia coli* Strains and Their Occurrence in Urinary Tract Infections in Women. *Antimicrobial Agents and Chemotherapy* **57**, 4512–4517 (2013).
18. Murray, B. E., Mathewson, J. J., DuPont, H. L., Ericsson, C. D. & Reves, R. R. Emergence of resistant fecal *Escherichia coli* in travelers not taking prophylactic antimicrobial agents. *Antimicrob. Agents Chemother.* **34**, 515–518 (1990).
19. Fantin, B. *et al.* Ciprofloxacin dosage and emergence of resistance in human commensal bacteria. *J. Infect. Dis.* **200**, 390–398 (2009).
20. Fernández-Orth, D. *et al.* Faecal phageome of healthy individuals: presence of antibiotic resistance genes and variations caused by ciprofloxacin treatment. *Journal of*

Antimicrobial Chemotherapy **74**, 854–864 (2019).

21. Górska, A. *et al.* Dynamics of the human gut phageome during antibiotic treatment. *Comput Biol Chem* **74**, 420–427 (2018).
22. Goerke, C., Köller, J. & Wolz, C. Ciprofloxacin and trimethoprim cause phage induction and virulence modulation in *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* **50**, 171–177 (2006).
23. Modi, S. R., Lee, H. H., Spina, C. S. & Collins, J. J. Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome. *Nature* **499**, 219–222 (2013).
24. Wang, M. *et al.* Estimating the contribution of bacteriophage to the dissemination of antibiotic resistance genes in pig feces. *Environ. Pollut.* **238**, 291–298 (2018).
25. Abeles, S. R., Ly, M., Santiago-Rodriguez, T. M. & Pride, D. T. Effects of Long Term Antibiotic Therapy on Human Oral and Fecal Viromes. *PLoS ONE* **10**, e0134941 (2015).
26. Enault, F. *et al.* Phages rarely encode antibiotic resistance genes: a cautionary tale for virome analyses. *ISME J* (2016) doi:10.1038/ismej.2016.90.
27. Burdet, C. *et al.* Ceftriaxone and cefotaxime have similar effects on the intestinal microbiota in human volunteers treated by standard doses regimens. *Antimicrob. Agents Chemother.* (2019) doi:10.1128/AAC.02244-18.
28. d’Humières, C. *et al.* A simple, reproducible and cost-effective procedure to analyse gut phageome: from phage isolation to bioinformatic approach. *Sci Rep* **9**, 11331 (2019).
29. Didion, J. P., Martin, M. & Collins, F. S. Atropos: specific, sensitive, and speedy trimming of sequencing reads. *PeerJ* **5**, e3720 (2017).
30. Crusoe, M. R. *et al.* The khmer software package: enabling efficient nucleotide sequence analysis. *F1000Research* (2015) doi:10.12688/f1000research.6924.1.
31. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
32. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
33. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
34. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
35. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
36. Oksanen, J. *et al.* *vegan: Community Ecology Package.* (2018).
37. Pramono, A. K. *et al.* Discovery and Complete Genome Sequence of a Bacteriophage from an Obligate Intracellular Symbiont of a Cellulolytic Protist in the Termite Gut. *Microbes Environ.* **32**, 112–117 (2017).
38. Gupta, S. K. *et al.* ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob. Agents Chemother.* **58**, 212–220 (2014).
39. Jia, B. *et al.* CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **45**, D566–D573 (2017).
40. Zankari, E. *et al.* Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* **67**, 2640–2644 (2012).
41. Munk, P. *et al.* Abundance and diversity of the faecal resistome in slaughter pigs and broilers in nine European countries. *Nature Microbiology* **3**, 898 (2018).
42. Ruppé, E. *et al.* Prediction of the intestinal resistome by a three-dimensional

structure-based method. *Nat Microbiol* **4**, 112–123 (2019).

43. Allen, H. K. *et al.* Antibiotics in feed induce prophages in swine fecal microbiomes. *MBio* **2**, (2011).

44. Johnson, T. A. *et al.* The In-Feed Antibiotic Carbadox Induces Phage Gene Transcription in the Swine Gut Microbiome. *MBio* **8**, (2017).

3.3.3 Principaux résultats de l'article et perspectives

Pour la première fois dans la littérature, un travail porte sur l'impact de deux céphalosporines de troisième génération sur le phageome intestinal chez des volontaires sains, dans une étude clinique randomisée contrôlée. Les deux antibiotiques provoquent une perturbation du phageome intestinal avec un retour progressif à l'état pré-antibiotique à J30 post traitement. Lorsqu'on s'intéresse à la diversité des phages on observe que dans le groupe ceftriaxone à J4 post traitement il y a une baisse drastique du nombre de phages présents dans l'échantillon. Les phages prédominants sont prédits comme appartenant à la famille des Podoviridae et l'un d'entre eux correspond au crAssphage. Une étude en 2007 a montré, *in vitro*, que des céphalosporines (cefotaxime, ceftriaxone, ceftazidime) utilisées à de faibles concentrations augmentent la production de phages virulents par les bactéries¹⁸³. La ceftriaxone et la cefotaxime ont une pharmacocinétique différente, la ceftriaxone est peut-être présente dans les selles avec des concentrations qui favorisent l'expression de certains phages, ce qui expliquerait notre observation. Le dosage microbiologique des antibiotiques dans les selles a été effectué dans l'étude précédente ¹⁸⁴, seulement deux individus (un dans chaque groupe) sur la selle à J4 post traitement avaient une quantité détectable d'antibiotique. Mais les faibles concentrations pouvant induire les phages, de l'ordre du nanogramme¹⁸³ se situent en dessous de la limite de détection du dosage microbiologiques des antibiotiques (1.25 µg/g).

Lors de l'analyse des données génomiques bactérienne de séquences de 16S sur les mêmes échantillons, il avait été montré un effet global des antibiotiques sur la perte de diversité des bactéries du microbiote intestinal mais sans différence entre les deux groupes. Nous avons donc des observations différentes entre le 16S et le phageome intestinal. Cela peut s'expliquer car nous n'analysons pas les données à la même échelle : les phages sont spécifiques d'une souche bactérienne alors que le 16S se situe au mieux au niveau de l'espèce bactérienne.

La place des phages dans le transfert de gènes de résistance aux antibiotiques est actuellement discutée et controversée. L'analyse réalisée dans ce travail montre que, en

choisissant des seuils bio-informatiques conservateurs, on ne retrouve pas plus de gènes de résistance dans les contigs phagiques après la prise d'antibiotiques. Cela semble être en accord avec une étude publiée récemment ¹³³ .

L'intégration de ces données avec les données de métagénomique globale sur les mêmes échantillons va permettre d'affiner notre analyse et de mieux caractériser les phages (hôte, style de vie). Les gènes de résistance retrouvés dans les phages vont également pouvoir être comparés au résistome global.

4. Conclusions et perspectives

L'objectif global de cette thèse était d'étudier l'impact de deux antibiotiques appartenant à la famille des céphalosporines de troisième génération sur le microbiote intestinal (bactéries et phages) des volontaires sains. Pour l'étude des phages, un développement méthodologique a été nécessaire, aussi bien expérimental (isolement des phages dans les selles) que bioinformatique. Ce nouveau protocole a été mis en œuvre pour analyser l'impact de deux antibiotiques (céfotaxime et ceftriaxone) sur le phageome intestinal.

Le microbiote intestinal est un milieu complexe. Jusqu'à présent c'est surtout la composition de la fraction bactérienne qui a été reliée à plusieurs maladies comme les maladies inflammatoires chroniques de l'intestin (MICI), le diabète, l'obésité ou encore l'autisme. Or, le microbiote intestinal est composé principalement de bactéries et de phages. Afin de pouvoir étudier l'implication du microbiote intestinal dans la santé humaine, il est donc nécessaire d'étudier également les phages et les étroites interactions qu'ils entretiennent avec les bactéries. La comparaison d'ADN phagique d'un échantillon de selle extrait et séquencé par les méthodes de métagénomiques classiques et de celui séquencé après une étape spécifique de concentration de phages et d'exclusion des bactéries sur la même selle révèle que 4 à 17% (cela dépend de la profondeur de séquençage) de l'ADN phagique isolé après concentration est retrouvé dans les données de métagénomique⁸⁴. Cela montre l'importance de traiter la selle avec une méthode adaptée, experte, si l'on souhaite étudier spécifiquement l'ADN de phage. De plus il est important, et intéressant, d'avoir les deux jeux de données (phageome et métagénomique globale) afin d'observer (i) les prophages intégrés aux bactéries dans les données de métagénomique globale et (ii) les particules virales sous forme de virion dans les données de métagénomique de phage. La question du choix de la procédure d'isolement des phages présents dans les selles est donc une première étape nécessaire à leur étude.

Nous avons montré que la méthode d'isolement des phages des selles, avec une étape de concentration des phages au polyéthylène glycol, était la plus performante. Cette technique a montré ses preuves en termes de reproductibilité, de simplicité et de faible coût. Cela impose une limitation à la rapidité d'exécution de la technique. Une piste possible d'amélioration qui n'a pas été explorée est la possibilité de réduire ce temps d'incubation. De plus, lorsque nous avons utilisé cette méthode à grande échelle, sur plus d'une centaine de prélèvements, nous avons observé quelques failles : pour certains échantillons de selles, l'ADN extrait n'était pas en quantité ou en qualité suffisante. Est-ce un problème de limites de la technique choisie ou un problème de composition de selles? Certaines selles on peut être trop de mucus rendant l'étape de filtration difficile, ainsi il faudrait rajouter une étape supplémentaire afin de mieux dissoudre le mucus. A présent, il serait intéressant de comparer, sur un grand nombre d'échantillons, la technique du PEG avec différents temps d'incubation et une étape supplémentaire pour dissoudre le mucus et ainsi mieux homogénéiser la selle. En revanche, il me paraît peu pertinent de continuer à utiliser la technique de concentration par ultracentrifugation au chlorure de césium, plusieurs papiers ayant montré qu'elle n'était pas reproductible et que la diversité des phages extraits était moindre. De plus elle est laborieuse à réaliser. Afin d'apprécier finement les performances de chaque méthodologie, certains contrôles pourraient être ajoutés tels que des réplicats d'extraction ou l'utilisation d'une communauté artificielle (« mock community ») avec quelques phages connus (ADN simple et double brin).

Au-delà des détails techniques de l'extraction, il serait intéressant de pouvoir visualiser les phages avant l'extraction de leur ADN. Une entreprise appelée Myriade a mis au point un appareil nommé Videodrop, qui permet la visualisation et le comptage instantané des phages dans 5µL de solution ¹⁸⁵. La visualisation de plusieurs culots de phages isolés de selles permettrait de comparer les diversités de morphologie et de faire le dénombrement de particules phagiques.

Un deuxième point abordé dans notre méthodologie concerne l'utilisation d'une polymérase (type phi 29) pour amplifier l'ADN de phage avant le séquençage. Nous avons confirmé et quantifié les biais d'analyse et d'interprétation générés par l'ajout d'une telle étape. La polymérase phi 29 amplifie préférentiellement et de façon non

reproductible l'ADN simple brin. Pendant longtemps, la quantité d'ADN demandé pour préparer les librairies de séquençage était telle que cette étape était indispensable. Dans les études utilisant une polymérase, il est difficile de tirer des conclusions sur l'abondance relative des familles de phages entre elles car les familles de phage à ADN simple brin (ex : Microviridae) ont été préférentiellement amplifiées. Désormais, le nouveau kit de préparation de librairie Illumina requiert très peu d'ADN (1 ng) et l'étape d'amplification n'a plus lieu d'être. Des études très récentes ont pourtant encore recours à ce procédé, il faut donc être très prudent sur leurs conclusions. D'autre part, les kits commerciaux Illumina, de préparation de librairie de séquençage, ne prennent en compte que l'ADN double brin. Ainsi, avec ces techniques, il est difficile d'avoir une vue d'ensemble exhaustive des phages du microbiote intestinal car les phages à ADN simple brin ne sont pas bien représentés. Pour pallier à ce problème, l'utilisation d'un autre kit commercial (Swift Biosciences 1S Plus kit) (Roux et al en 2016) permet de préparer des librairies Illumina en tenant compte de l'ADN simple brin et double brins grâce à une étape préliminaire. A ce kit, il serait possible d'ajouter une étape de reverse transcriptase avant l'étape de séquençage afin de séquencer également les phages à ARN. Il est à noter que l'utilisation d'une polymérase type phi 29, amplifiant préférentiellement les phages à ADN simple brin, peut être intéressante pour l'étude des phages à ADN simple brin avec une approche qualitative et non quantitative.

Un troisième point étudié dans notre partie méthodologique est la profondeur de séquençage. Nous avons déterminé la profondeur de séquençage optimale pour une bonne analyse des phages du microbiote intestinal. Grâce au séquençage d'un même échantillon deux fois (faible/forte profondeur) et à des prédictions de différentes profondeurs *in silico*, nous avons pu observer une diminution du nombre de nouveaux contigs de phages lorsque la profondeur augmente. Le nombre de nouveaux contigs de phage commence à stagner lorsque la profondeur de 20 000 000 de reads est atteinte, au-delà, la majorité des nouveaux contigs ne sont pas des phages. Notre travail a permis de déterminer un ordre de grandeur d'une bonne profondeur de séquençage pour une bonne représentativité de la diversité des contigs de phage.

Les améliorations expérimentales réalisées ont été accompagnées de nouvelles approches en bioinformatique. L'analyse bio-informatique de l'ADN des phages est difficile car les bases de données publiques sont pauvres en phages et encore plus en phages du microbiote intestinal. Par ailleurs, il existe une grande diversité phagique

expliquée par l'existence d'un ensemble de gènes très divers et un taux élevé d'évolution nucléotidique. Les bases de données publiques de phages étant très pauvres, il est important d'insérer une étape d'assemblage des reads dans le pipeline bioinformatique d'analyse des reads de métagénome de phage. Si cette étape n'est pas réalisée, les reads sont mappés directement sur les bases de données de phages, la majorité d'entre eux ne mappera pas et sera alors perdue pour la suite des analyses. Après l'assemblage des reads en contigs, il est possible de comparer ceux-ci aux bases de données connues et de rechercher des signaux "de phage" afin d'identifier des contigs s'apparentant à des phages. Nous avons mis au point un algorithme bio-informatique qui permet de classer les contigs. La première étape de cet algorithme est basée sur l'analyse d'un score de similarité appelé wGRR (« weighted gene repertoire relatedness »). Le wGRR est calculé entre chaque contig et chaque référence d'une base de donnée (de phage ou de prophage). Il permet d'identifier les contigs proches de phages connus et ainsi de connaître leur famille, leur hôte et leur style de vie. Nous avons déterminé des cut-off du wGRR calculés en fonction (i) des bases de données utilisées, (ii) de la question posée et (iii) de la taille du contig. La prise en compte de la taille du contig est indispensable en métagénomique au vue de leur grande variabilité.

Dans notre travail de méthodologie, nous avons recherché les prophages dans deux banques de bactéries : dans les génomes des bactéries de Refseq (disponible publiquement) grâce au logiciel Phage Finder et dans des bactéries retrouvées dans le microbiote intestinal (457 bactéries venant du Human Metagenome Project), 1243 prophages ont été identifiés par Virsorter. Nous avons le projet de construire une base de données publique de prophages du microbiote intestinal à partir des bactéries du (i) Human Metagenome Project, (ii) de Metahit et (iii) des bactéries du microbiote intestinal retrouvées par technique de culturomique. Les prophages seraient alors recherchés dans ces différents génomes de bactéries (ou draft) avec différents logiciels (ex: Virsorter, MARVEL, Prophage Hunter). Les prophages retrouvés seront vérifiés manuellement et annotés. Cela permettrait d'avoir une base de données originale de prophages du microbiote intestinal et de pouvoir les caractériser plus précisément. Concernant les phages virulents, il est plus difficile de construire une base de données. Seule l'analyse d'études métagénomiques de phageome peut permettre de caractériser de nouveaux phages virulents présents dans le microbiote intestinal.

Forts de la mise au point de notre pipeline bio-informatique, nous avons pu étudier et caractériser l'ADN séquencé. Cela nous a permis d'observer que les phages présents dans le microbiote intestinal sont individu-spécifiques et sont majoritairement des caudovirales tempérés ayant comme hôte les Firmicutes. Notre technique ne permet pas une bonne représentation des phages à ADN simple brin, les conclusions concernant la composition en phages du microbiote intestinal doivent donc être analysées dans cette limite.

Lors de l'analyse des séquences d'ADN issues des phageomes intestinaux, il est toujours retrouvé de l'ADN bactérien. Cet ADN peut avoir plusieurs origines : cela peut être dû à de la contamination bactérienne lors du processus d'isolation des phages, à de la transduction généralisée si de l'ADN bactérien a été accidentellement encapsidé dans un virion, ou à la présence de vésicules dans lesquelles il peut y avoir de l'ADN bactérien. Il est impossible de faire la part des choses entre ces trois phénomènes. Pour autant, il est important de ne pas éliminer cet ADN bactérien car il peut être informatif, notamment pour estimer la transduction généralisée lors de l'analyse des données longitudinales.

Cette méthodologie d'isolement et d'analyse des phages a ensuite été appliquée aux échantillons de selles d'une étude clinique dont l'objectif était de comparer l'impact de deux céphalosporines de troisième génération au spectre commun mais au mode d'élimination différent sur le microbiote intestinal. La résistance bactérienne aux antibiotiques constitue une menace pour l'Homme et l'utilisation d'antibiotiques favorise l'émergence de bactéries résistantes dans le microbiote intestinal. C'est dans ce contexte que la bonne prescription antibiotique, c'est à dire (i) un spectre étroit sur le germe visé et (ii) un moindre impact sur le microbiote, est primordiale. L'étude clinique CEREMI a été créée dans le but de comparer l'impact de la ceftriaxone et du céfotaxime sur l'émergence de bactéries résistantes dans le microbiote intestinal. L'étude clinique a été conduite chez 22 volontaires sains (11 dans chaque bras : ceftriaxone versus céfotaxime) et les doses d'antibiotiques ainsi que le rythme d'administration étaient ceux réalisés en pratique clinique (céfotaxime, 1gramme/8 heures et ceftriaxone 1 gramme/24 heures). Des échantillons de selles et de plasma ont été prélevés à 13 points de temps différents au cours de l'étude. Je me suis personnellement impliquée dans la phase de culture bactérienne et nous avons constaté une réduction du nombre de bacilles à Gram négatif post traitement dans les deux bras, puis un retour à la normale

quinze jours plus tard, sans émergence de bacilles à Gram négatif résistants aux CIIG. Il n'est pas étonnant de ne pas observer l'émergence de bactéries résistantes car l'étude a été faite sur des volontaires sains, population peu colonisée par des bactéries productrices de BLSE (entre 5 et 10% en Europe ^{186,187}). L'étude du gène 16S a également permis de constater une chute de la diversité bactérienne dans les deux groupes mais sans différence significative entre les deux antibiotiques. Si l'on considère les analyses bactériennes (culture et étude du gène 16S), la différence entre les modes d'élimination du cefotaxime et de la ceftriaxone, est probablement compensée par le rythme d'administration de ces molécules. Au total, une quantité proche d'antibiotique arrive dans le tube digestif mais avec une pharmacocinétique différente, c'est la raison pour laquelle les différentes analyses n'ont pas révélé de différence d'impact sur le microbiote intestinal entre les deux molécules.

Nous avons ensuite pu analyser, sur ces mêmes échantillons de selles, le phageome intestinal. C'est la première fois qu'une étude randomisée contrôlée analyse l'impact d'antibiotiques sur le phageome intestinal humain. Deux études ^{120,121} s'étaient intéressées à l'impact des quinolones sur le phageome intestinal mais le nombre d'individus était très restreint (respectivement 1 et 2).

La méthodologie d'isolement des phages des selles et d'analyse bio-informatique des séquences mise au point précédemment a été appliquée aux échantillons de l'étude clinique CEREMI. Nous avons observé que plus de la moitié des reads des échantillons étaient des reads phagiques. Ce ratio est plutôt élevé en comparaison avec d'autres études confrontées au problème « du dark virome matter » ¹⁴⁶.

Cette étude confirme que le phageome intestinal est individu-spécifique et que la variable « individu » est plus forte que la « variable prise antibiotique ». L'identité phagique de l'individu est même encore plus forte après le traitement : le pourcentage des contigs phages retrouvé chez seulement un individu passe de 13% avant la prise d'antibiotique à 20% après et revient à 18% 6 mois plus tard. Dans les deux groupes, la prise d'antibiotique a modifié le phageome intestinal avec un retour progressif à l'état initial à J30. Mais en regardant plus précisément nous avons observé que dans le groupe ceftriaxone, il y a une chute drastique du nombre de contigs sur l'échantillon post-traitement chez plus de la moitié des volontaires sains. Cette chute de diversité n'est pas observée dans le groupe cefotaxime. Cet effet n'est pas significatif en raison du faible

nombre de volontaires dans chaque bras et de la grande variabilité inter-individuelle. Lorsque l'on regarde plus précisément les contigs de phages qui sont très prédominants dans les échantillons post-traitement du groupe ceftriaxone, il est difficile de les caractériser mais une majorité est proche des phages appartenant à la famille Podoviridae et un correspond au crAssphage. Plusieurs hypothèses pourraient expliquer cette observation : cela peut correspondre à des prophages excisés puis répliqués en nombre, ou encore à des phages virulents participant aux régulations du changement d'abondance des phyla bactériens. La poursuite du projet devrait nous aider à affiner notre réponse. Les deux antibiotiques semblent donc avoir un effet différent sur le phageome intestinal, effet qui n'a pas été vu dans les analyses bactérienne de 16S. Ce n'est pas une contradiction car les populations étudiées et les niveaux d'analyses sont différents : les phages étant souvent capables d'infecter un nombre réduit de souches d'une espèce, l'analyse se situe au niveau de la souche bactérienne alors que pour le 16S l'analyse se situe au mieux au niveau de l'espèce bactérienne. Cette différence de granularité peut expliquer les différences observées.

Une question importante à laquelle ce travail s'est intéressé, est la place des phages dans le transfert de gènes de résistance aux antibiotiques sous la pression de sélection antibiotique. Certaines études ont montré une augmentation des gènes de résistance dans la fraction phagique des selles après un traitement antibiotique, ces travaux ont été ensuite remis en question par le travail d'Enault *et al.*¹³³ qui insistait sur les cut-off bio-informatiques choisis pour détecter ces gènes de résistance afin d'éviter les faux positifs. Dans notre étude, nous avons recherché des gènes de résistance aux antibiotiques sur le catalogue complet des contigs grâce à plusieurs bases de données de gènes de résistance et des seuils bio-informatiques restrictifs. Au total 159 contigs de phages avec un gène de résistance aux antibiotiques ont été détectés. Soixante-dix pour cent des gènes de résistance ont été uniquement trouvés par la base de données très récente MUSTARD¹²⁷. Cela est cohérent car cette base de données a été construite avec une méthode originale de modélisation 3D et uniquement à partir de drafts de génomes bactériens provenant du microbiote intestinal. Cependant, sa robustesse n'a pas encore pu être testée dans d'autres études. Sur les 6 gènes de résistance appartenant à la famille des beta-lactamases retrouvés, 5 ont été identifiés grâce à la base de données MUSTARD et sont des beta-lactamases de classe B ou métallo-beta-lactamases. Ces gènes

de résistance sont particulièrement inquiétants du fait de leur large spectre. La présence de ces métallo-betalactamases dans les contigs de phage mérite d'être confirmée par une approche de génomique fonctionnelle. D'un point de vue global, nous ne mettons pas en évidence une augmentation des gènes de résistance aux antibiotiques dans la fraction phagique post-traitement antibiotique. L'effet est même contraire, particulièrement dans le groupe ceftriaxone où les contigs phagiques très représentés dans l'échantillon post-traitement ne comportent pas de gène de résistance. Etudier les gènes de résistance aux antibiotiques sur les contigs phagiques revient à explorer le rôle de la transduction spécialisée dans le transfert de gènes de résistance. Afin d'apprécier le rôle de la transduction généralisée nous avons étudié la fraction bactérienne de chacun des échantillons. L'abondance relative des reads bactériens a tendance à augmenter dans l'échantillon de selles post-antibiotique, évoquant un processus de transduction généralisée. Nous avons ensuite analysé la proportion de gènes de résistance aux antibiotiques dans la fraction bactérienne et nous n'avons pas retrouvé plus de gènes de résistance dans les échantillons post-traitement. Dans l'hypothèse où un processus de transduction généralisée est favorisé par le traitement antibiotique, cela n'inclut pas une encapsidation préférentielle de gène de résistance aux antibiotiques. Notre travail ne met donc pas en évidence un rôle des phages dans le transfert de gènes de résistance aux antibiotiques sous traitement par céphalosporine de troisième génération. Plus généralement la question d'une contre sélection des gènes de résistance par les phages se pose car il a été décrit plusieurs prophages avec des gènes de résistance mais ayant perdu leurs capacités lytiques^{101,188}. Cela expliquerait leur absence de prolifération après le traitement antibiotique. D'autres études doivent être conduites afin de mieux comprendre ce phénomène.

Prochainement, les données de métagénomique vont être disponibles sur les échantillons de selles avec une analyse du résistome qui correspond aux gènes de résistance aux antibiotiques présents dans les données de métagénomique. De plus, l'ensemble des champignons ou mycobiome va également être étudié ainsi que la population d'*E. coli*. Les données métagénomiques vont pouvoir être confrontées aux données du phageome et ainsi nous allons pouvoir confirmer, pour certains phages, leur style de vie et leur hôte. Une étude de co-abondance entre les bactéries et les phages nous permettra également d'étudier la dynamique phages/bactérie dans le microbiote

intestinal. Nous pourrions également constater les différences entre les deux jeux de données et à quel point la méthode d'isolement spécifique des phages apporte une information supplémentaire. En ce qui concerne le résistome, nous pourrions voir si certains gènes de résistance sont portés par des prophages intégrés aux bactéries et leur évolution sous pression antibiotique.

Face à l'augmentation du nombre de bactéries multi-résistantes et à l'inefficacité des antibiotiques actuels, d'autres modes de traitement sont envisagés. La phagothérapie c'est-à-dire l'usage des phages pour traiter des maladies bactériennes se développe rapidement. Cependant, cette voie thérapeutique fait face à de nombreux obstacles : (i) de trouver le phage spécifique de la souche bactérienne à éliminer (ii) de décider du mode d'administration car les données pharmacodynamiques et pharmacocinétiques sont presque inexistantes (iii) les différents mécanismes mis en place par les bactéries pour lutter contre les phages. La question du transfert de gènes de résistance aux antibiotiques via les phages est particulièrement préoccupante dans les cas de la phagothérapie. Les phages utilisés dans la phagothérapie sont uniquement virulents donc la seule transduction possible est la transduction généralisée. Ce mode de transduction est le plus difficile à évaluer. L'association antibiotique-phage a démontré, *in vitro*, son efficacité en potentialisant l'activité lytique du phage ¹⁸³. Au regard (i) de la faible probabilité de la transduction généralisée (ii) du phage choisi qui aura un spectre d'hôte le plus étroit possible et (iii) des résultats de notre étude qui ne montre pas d'implication des phages dans le transfert de gène de résistance sous pression antibiotique, la probabilité du transfert de gène de résistance dans le cas de la phagothérapie paraît très faible voire négligeable.

L'importance de l'équilibre du microbiote intestinal pour une bonne santé n'est plus à démontrer. Mais l'importance des phages, en particulier, du microbiote intestinal est en train d'être dévoilée, notamment dans le traitement des infections à *Clostridioides difficile*.

Récemment, afin de remédier aux récurrentes infections à *C. difficile* potentiellement mortelles, la transplantation fécale a fait ses preuves. Des études ont montré que la composition en phage du donneur pouvait avoir une influence sur la réponse du receveur^{139,189}. Une autre étude pionnière¹⁹⁰ a réalisé des transplantations fécales, chez

5 patients, uniquement avec les phages extraits du microbiote intestinal d'un donneur. Cette approche a montré son efficacité pour le traitement des infections à *C. difficile*, à l'image de la transplantation fécale classique⁴⁷. Cela met en évidence le rôle central des phages dans l'équilibre du microbiote et soulève de nombreuses questions : l'évolution du phageome chez le receveur est-il facteur pronostic de guérison définitive de la maladie ? La restauration du phageome intestinal est-il aussi important que la restauration du bactériome ? La composition du phageome du donneur est-elle un facteur pronostic ? La composition du phageome doit-elle intervenir dans sélection des donneurs ? Toutes ces questions témoignent de l'importance de l'intégration des données de phages du microbiote intestinal.

En conclusion, ce travail démontre l'importance de ne pas limiter l'étude du microbiote intestinal aux bactéries. L'analyse des phages du microbiote est indispensable bien que difficile tant sur la méthodologie d'isolement que l'analyse bio-informatique. Ce travail propose une méthodologie d'analyse et son application dans l'exploration de l'impact de deux antibiotiques sur le bactériome et le phageome.

5. Bibliographie

1. <http://www.who.int/mediacentre/news/releases/2014/amr-report/fr/>.
2. Colomb-Cotinat M *et al.* Morbidité et mortalité des infections à bactéries multi-résistantes aux antibiotiques en France en 2012. (2015).
3. Cassini, A. *et al.* Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. *The Lancet Infectious Diseases* **19**, 56–66 (2019).
4. Andremont, A. Role of the commensal flora in the spread of bacterial resistance : the invisible collateral damage of antibiotic treatments. *ASM News* 601–7 (2003).
5. Carlet, J. The gut is the epicentre of antibiotic resistance. *Antimicrob Resist Infect Control* **1**, 39 (2012).
6. Perrier-Gros, J. D. *et al.* Entérocoques résistants aux glycopeptides dans les viandes. *Bull Epidemiol Hebd* 50–5 (1998).
7. Holmes, A. H. *et al.* Understanding the mechanisms and drivers of antimicrobial resistance. *The Lancet* **387**, 176–187 (2016).
8. Høiby, N. *et al.* Excretion of ciprofloxacin in sweat and multiresistant *Staphylococcus epidermidis*. *Lancet* **349**, 167–169 (1997).
9. Prevot, M. H., Andremont, A., Sancho-Garnier, H. & Tancrede, C. Epidemiology of intestinal colonization by members of the family Enterobacteriaceae resistant to cefotaxime in a hematology-oncology unit. *Antimicrob. Agents Chemother.* **30**, 945–947 (1986).
10. de Lastours, V., Chau, F., Roy, C., Larroque, B. & Fantin, B. Emergence of quinolone resistance in the microbiota of hospitalized patients treated or not with a fluoroquinolone. *J. Antimicrob. Chemother.* **69**, 3393–3400 (2014).
11. Woerther, P.-L. *et al.* Emergence and dissemination of extended-spectrum beta-lactamase-producing *Escherichia coli* in the community: lessons from the study of a remote and controlled population. *J. Infect. Dis.* **202**, 515–523 (2010).
12. Fantin, B. *et al.* Ciprofloxacin dosage and emergence of resistance in human commensal bacteria. *J. Infect. Dis.* **200**, 390–398 (2009).
13. Tacão, M., Correia, A. & Henriques, I. Resistance to broad-spectrum antibiotics in aquatic systems: anthropogenic activities modulate the dissemination of bla(CTX-M)-like genes. *Appl. Environ. Microbiol.* **78**, 4134–4140 (2012).
14. AKIBA T, KOYAMA K, ISHIKI Y, KIMURA S & FUKUSHIMA T. On the mechanism of the development of multiple-drug-resistant clones of *Shigella*. *Jpn J Microbiol* 219–227 (1960).
15. Brabban, A. D., Hite, E. & Callaway, T. R. Evolution of foodborne pathogens via temperate bacteriophage-mediated gene transfer. *Foodborne Pathog. Dis.* **2**, 287–303 (2005).
16. Lupo, A., Coyne, S. & Berendonk, T. U. Origin and evolution of antibiotic resistance: the common mechanisms of emergence and spread in water bodies. *Front Microbiol* **3**, 18 (2012).
17. Sekirov, I., Russell, S. L., Antunes, L. C. M. & Finlay, B. B. Gut microbiota in health

and disease. *Physiol. Rev.* **90**, 859–904 (2010).

18. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
19. Miller TL & Wolin PJ. Enumeration of *Methanobrevibacter smithii* in human feces. *Arch Microbiol* **131**, 14–8 (1982).
20. Miller TL & Wolin MJ. *Methanosphaera stadtmaniae* gen. nov., sp. nov.: a species that forms methane by reducing methanol with hydrogen. *Arch Microbiol* **141**, 116–22 (1985).
21. Minot, S. *et al.* Rapid evolution of the human gut virome. *Proceedings of the National Academy of Sciences* **110**, 12450–12455 (2013).
22. Nash, A. K. *et al.* The gut mycobiome of the Human Microbiome Project healthy cohort. *Microbiome* **5**, 153 (2017).
23. Hallen-Adams, H. E. & Suhr, M. J. Fungi in the healthy human gastrointestinal tract. *Virulence* **8**, 352–358 (2017).
24. Shkoporov, A. N. & Hill, C. Bacteriophages of the Human Gut: The ‘Known Unknown’ of the Microbiome. *Cell Host Microbe* **25**, 195–209 (2019).
25. Landman, C. & Quévrain, E. [Gut microbiota: Description, role and pathophysiological implications]. *Rev Med Interne* **37**, 418–423 (2016).
26. Grall, N., Andremont, A. & Ruppé, E. Microbiote intestinale. *EMC* (2017).
27. Turnbaugh, P. J. *et al.* The human microbiome project. *Nature* **449**, 804–810 (2007).
28. Browne, H. P. *et al.* Culturing of ‘unculturable’ human microbiota reveals novel taxa and extensive sporulation. *Nature* **533**, 543–546 (2016).
29. Lagier, J.-C. *et al.* Culture of previously uncultured members of the human gut microbiota by culturomics. *Nat Microbiol* **1**, 16203 (2016).
30. Goodrich, J. K. *et al.* Conducting a Microbiome Study. *Cell* **158**, 250–262 (2014).
31. Jumpstart Consortium Human Microbiome Project Data Generation Working Group. Evaluation of 16S rDNA-based community profiling for human microbiome research. *PLoS ONE* **7**, e39315 (2012).
32. Yarza, P. *et al.* Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nature Reviews Microbiology* **12**, 635–645 (2014).
33. Le Chatelier, E. *et al.* Richness of human gut microbiome correlates with metabolic markers. *Nature* **500**, 541–546 (2013).
34. Costea, P. I. *et al.* Towards standards for human fecal sample processing in metagenomic studies. *Nat. Biotechnol.* **35**, 1069–1076 (2017).
35. Sullivan, A., Edlund, C. & Nord, C. E. Effect of antimicrobial agents on the ecological balance of human microflora. *Lancet Infect Dis* **1**, 101–114 (2001).
36. Willing, B. P., Russell, S. L. & Finlay, B. B. Shifting the balance: antibiotic effects on host-microbiota mutualism. *Nat. Rev. Microbiol.* **9**, 233–243 (2011).
37. Hall, I. C. & O’toole, E. INTESTINAL FLORA IN NEW-BORN INFANTS: WITH A DESCRIPTION OF A NEW PATHOGENIC ANAEROBE, *BACILLUS DIFFICILIS*. *Am J Dis Child* **49**, 390–402 (1935).
38. Theriot, C. M. & Young, V. B. Interactions Between the Gastrointestinal Microbiome and *Clostridium difficile*. *Annu. Rev. Microbiol.* **69**, 445–461 (2015).
39. Heimann, S. M. *et al.* Economic burden of *Clostridium difficile* associated diarrhoea: a cost-of-illness study from a German tertiary care hospital. *Infection* **43**, 707–714 (2015).
40. Bhalla, A. *et al.* Antianaerobic Antibiotic Therapy Promotes Overgrowth of

Antibiotic-Resistant, Gram-Negative Bacilli and Vancomycin-Resistant Enterococci in the Stool of Colonized Patients. *Infection Control & Hospital Epidemiology* **24**, 644–649 (2003).

41. Taur, Y. *et al.* Intestinal Domination and the Risk of Bacteremia in Patients Undergoing Allogeneic Hematopoietic Stem Cell Transplantation. *Clin Infect Dis* **55**, 905–914 (2012).

42. Ruppé, E. *et al.* Relative Fecal Abundance of Extended-Spectrum- β -Lactamase-Producing *Escherichia coli* Strains and Their Occurrence in Urinary Tract Infections in Women. *Antimicrobial Agents and Chemotherapy* **57**, 4512–4517 (2013).

43. Perez, F., Pultz, M. J., Endimiani, A., Bonomo, R. A. & Donskey, C. J. Effect of Antibiotic Treatment on Establishment and Elimination of Intestinal Colonization by KPC-Producing *Klebsiella pneumoniae* in Mice. *Antimicrob Agents Chemother* **55**, 2585–2589 (2011).

44. Murray, B. E., Mathewson, J. J., DuPont, H. L., Ericsson, C. D. & Reves, R. R. Emergence of resistant fecal *Escherichia coli* in travelers not taking prophylactic antimicrobial agents. *Antimicrob. Agents Chemother.* **34**, 515–518 (1990).

45. Dethlefsen, L. & Relman, D. A. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proc. Natl. Acad. Sci. U.S.A.* **108 Suppl 1**, 4554–4561 (2011).

46. Ruppé, E. *et al.* Impact of antibiotics on the intestinal microbiota needs to be re-defined to optimize antibiotic usage. *Clin. Microbiol. Infect.* **24**, 3–5 (2018).

47. van Nood, E. *et al.* Duodenal infusion of donor feces for recurrent *Clostridium difficile*. *N. Engl. J. Med.* **368**, 407–415 (2013).

48. Tavoukjian, V. Faecal microbiota transplantation for the decolonization of antibiotic-resistant bacteria in the gut: a systematic review and meta-analysis. *J. Hosp. Infect.* **102**, 174–188 (2019).

49. Otawa, K. *et al.* Abundance, diversity, and dynamics of viruses on microorganisms in activated sludge processes. *Microb. Ecol.* **53**, 143–152 (2007).

50. Ashelford, K. E., Day, M. J. & Fry, J. C. Elevated abundance of bacteriophage infecting bacteria in soil. *Appl. Environ. Microbiol.* **69**, 285–289 (2003).

51. Letarov, A. & Kulikov, E. The bacteriophages in human- and animal body-associated microbial communities. *Journal of Applied Microbiology* **107**, 1–13 (2009).

52. Weinbauer, M. G. Ecology of prokaryotic viruses. *FEMS Microbiol. Rev.* **28**, 127–181 (2004).

53. Ackermann, H.-W. Classification of bacteriophages. in *the bacteriophages* (ed. Calendar Richard) (2006).

54. Lefkowitz, E. J. *et al.* Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Res* **46**, D708–D717 (2018).

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

56. Duckworth, D.H. Phage Ecology. in *History and basic properties of bacterial viruses*.

57. Bertozzi Silva, J., Storms, Z. & Sauvageau, D. Host receptors for bacteriophage adsorption. *FEMS Microbiol Lett* **363**, (2016).

58. Abedon, Stephen T. Phage Ecology. in *The Bacteriophages* (ed. Calendar Richard) (2006).

59. Wagner, P. L. & Waldor, M. K. Bacteriophage control of bacterial virulence. *Infect. Immun.* **70**, 3985–3993 (2002).

60. Bobay, L.-M., Touchon, M. & Rocha, E. P. C. Manipulating or Superseding Host Recombination Functions: A Dilemma That Shapes Phage Evolvability. *PLoS Genetics* **9**,

e1003825 (2013).

61. Bobay, L.-M., Touchon, M. & Rocha, E. P. C. Pervasive domestication of defective prophages by bacteria. *Proceedings of the National Academy of Sciences* **111**, 12127–12132 (2014).
62. Touchon, M., Bernheim, A. & Rocha, E. P. Genetic and life-history traits associated with the distribution of prophages in bacteria. *ISME J* **10**, 2744–2754 (2016).
63. Ohnishi, M., Kurokawa, K. & Hayashi, T. Diversification of *Escherichia coli* genomes: are bacteriophages the major contributors? *Trends Microbiol.* **9**, 481–485 (2001).
64. Łoś, M. & Węgrzyn, G. Pseudolysogeny. *Adv. Virus Res.* **82**, 339–349 (2012).
65. Wilson, D. R. & Finlay, B. B. Phage display: applications, innovations, and issues in phage and host biology. *Can. J. Microbiol.* **44**, 313–329 (1998).
66. Roach, D. R. & Debarbieux, L. Phage therapy: awakening a sleeping giant. *Emerging Topics in Life Sciences* **1**, 93–103 (2017).
67. Zinder, N. D. & Lederberg, J. GENETIC EXCHANGE IN *SALMONELLA*1. *J Bacteriol* **64**, 679–699 (1952).
68. Touchon, M., Moura de Sousa, J. A. & Rocha, E. P. Embracing the enemy: the diversification of microbial gene repertoires by phage-mediated horizontal gene transfer. *Curr. Opin. Microbiol.* **38**, 66–73 (2017).
69. Haaber, J. *et al.* Bacterial viruses enable their host to acquire antibiotic resistance genes from neighbouring cells. *Nature Communications* **7**, 13333 (2016).
70. Chen, J. *et al.* Genome hypermobility by lateral transduction. *Science* **362**, 207–212 (2018).
71. Popa, O., Landan, G. & Dagan, T. Phylogenomic networks reveal limited phylogenetic range of lateral gene transfer by transduction. *ISME J* **11**, 543–554 (2017).
72. Hyman, P. & Abedon, S. T. Chapter 7 - Bacteriophage Host Range and Bacterial Resistance. in *Advances in Applied Microbiology* vol. 70 217–248 (Academic Press, 2010).
73. Sullivan, M. B., Waterbury, J. B. & Chisholm, S. W. Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature* **424**, 1047–1051 (2003).
74. Chen, J. & Novick, R. P. Phage-mediated intergeneric transfer of toxin genes. *Science* **323**, 139–141 (2009).
75. Labrie, S. J., Samson, J. E. & Moineau, S. Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.* **8**, 317–327 (2010).
76. Samson, J. E., Magadán, A. H., Sabri, M. & Moineau, S. Revenge of the phages: defeating bacterial defences. *Nat. Rev. Microbiol.* **11**, 675–687 (2013).
77. Hoyles, L. *et al.* Characterization of virus-like particles associated with the human faecal and caecal microbiota. *Research in Microbiology* **165**, 803–812 (2014).
78. Reyes, A. *et al.* Gut DNA viromes of Malawian twins discordant for severe acute malnutrition. *Proc Natl Acad Sci U S A* **112**, 11941–11946 (2015).
79. Zhang, T. *et al.* RNA viral community in human feces: prevalence of plant pathogenic viruses. *PLoS Biol.* **4**, e3 (2006).
80. Reyes, A. *et al.* Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* **466**, 334–338 (2010).
81. Waller, A. S. *et al.* Classification and quantification of bacteriophage taxa in human gut metagenomes. *The ISME journal* **8**, 1391–1402 (2014).
82. Knowles, B. *et al.* Lytic to temperate switching of viral communities. *Nature* **531**, 466–470 (2016).
83. Silveira, C. B. & Rohwer, F. L. Piggyback-the-Winner in host-associated microbial communities. *NPJ Biofilms Microbiomes* **2**, 16010 (2016).

84. Minot, S. *et al.* The human gut virome: Inter-individual variation and dynamic response to diet. *Genome Research* **21**, 1616–1625 (2011).
85. Reyes, A., Semenkovich, N. P., Whiteson, K., Rohwer, F. & Gordon, J. I. Going viral: next-generation sequencing applied to phage populations in the human gut. *Nature Reviews Microbiology* **10**, 607–617 (2012).
86. Stern, A., Mick, E., Tirosh, I., Sagy, O. & Sorek, R. CRISPR targeting reveals a reservoir of common phages associated with the human gut microbiome. *Genome Research* **22**, 1985–1994 (2012).
87. Manrique, P. *et al.* Healthy human gut phageome. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 10400–10405 (2016).
88. Dutilh, B. E. *et al.* A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat Commun* **5**, 4498 (2014).
89. Guerin, E. *et al.* Biology and Taxonomy of crAss-like Bacteriophages, the Most Abundant Virus in the Human Gut. *Cell Host Microbe* **24**, 653-664.e6 (2018).
90. Shkoporov, A. N. *et al.* ΦCrAss001 represents the most abundant bacteriophage family in the human gut and infects *Bacteroides intestinalis*. *Nature Communications* **9**, 4781 (2018).
91. Lim, E. S. *et al.* Early life dynamics of the human gut virome and bacterial microbiome in infants. *Nature Medicine* **21**, 1228–1234 (2015).
92. McCann, A. *et al.* Viromes of one year old infants reveal the impact of birth mode on microbiome diversity. *PeerJ* **6**, (2018).
93. Norman, J. M. *et al.* Disease-Specific Alterations in the Enteric Virome in Inflammatory Bowel Disease. *Cell* **160**, 447–460 (2015).
94. Breitbart, M. *et al.* Viral diversity and dynamics in an infant gut. *Research in Microbiology* **159**, 367–373 (2008).
95. Turnbaugh, P. J., Bäckhed, F., Fulton, L. & Gordon, J. I. Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell Host Microbe* **3**, 213–223 (2008).
96. Ma, Y., You, X., Mai, G., Tokuyasu, T. & Liu, C. A human gut phage catalog correlates the gut phageome with type 2 diabetes. *Microbiome* **6**, 24 (2018).
97. Barr, J. J., Youle, M. & Rohwer, F. Innate and acquired bacteriophage-mediated immunity. *Bacteriophage* **3**, e25857 (2013).
98. Legendre, D., Soumillon, P. & Fastrez, J. Engineering a regulatable enzyme for homogeneous immunoassays. *Nat. Biotechnol.* **17**, 67–72 (1999).
99. Mazaheri Nezhad Fard, R., Barton, M. D. & Heuzenroeder, M. W. Bacteriophage-mediated transduction of antibiotic resistance in enterococci. *Lett. Appl. Microbiol.* **52**, 559–564 (2011).
100. Schuch, R. & Fischetti, V. A. Detailed genomic analysis of the Wbeta and gamma phages infecting *Bacillus anthracis*: implications for evolution of environmental fitness and antibiotic resistance. *J. Bacteriol.* **188**, 3037–3051 (2006).
101. Billard-Pomares, T. *et al.* Characterization of a P1-like bacteriophage carrying an SHV-2 extended-spectrum β-lactamase from an *Escherichia coli* strain. *Antimicrob. Agents Chemother.* **58**, 6550–6557 (2014).
102. Colomer-Lluch, M., Jofre, J. & Muniesa, M. Quinolone resistance genes (qnrA and qnrS) in bacteriophage particles from wastewater samples and the effect of inducing agents on packaged antibiotic resistance genes. *J. Antimicrob. Chemother.* **69**, 1265–1274 (2014).
103. Colomer-Lluch, M. *et al.* Antibiotic resistance genes in bacterial and bacteriophage fractions of Tunisian and Spanish wastewaters as markers to compare

- the antibiotic resistance patterns in each population. *Environ Int* **73**, 167–175 (2014).
104. Parsley, L. C. *et al.* Identification of diverse antimicrobial resistance determinants carried on bacterial, plasmid, or viral metagenomes from an activated sludge microbial assemblage. *Appl. Environ. Microbiol.* **76**, 3753–3757 (2010).
 105. Modi, S. R., Lee, H. H., Spina, C. S. & Collins, J. J. Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome. *Nature* **499**, 219–222 (2013).
 106. Quiros, P. *et al.* Antibiotic Resistance Genes in the Bacteriophage DNA Fraction of Human Fecal Samples. *Antimicrobial Agents and Chemotherapy* **58**, 606–609 (2014).
 107. Fancello, L., Desnues, C., Raoult, D. & Rolain, J. M. Bacteriophages and diffusion of genes encoding antimicrobial resistance in cystic fibrosis sputum microbiota. *J. Antimicrob. Chemother.* **66**, 2448–2454 (2011).
 108. Blanchard, T. J., Poston, S. M. & Reynolds, P. J. Recipient characteristics in the transduction of methicillin resistance in *Staphylococcus epidermidis*. *Antimicrob Agents Chemother* **29**, 539–541 (1986).
 109. Pereira, M. S., Barreto, V. P. & Siqueira-Júnior, J. P. Phage-mediated transfer of tetracycline resistance in *Staphylococcus aureus* isolated from cattle in Brazil. *Microbios* **92**, 147–155 (1997).
 110. Rocha, E. P. C. An Appraisal of the Potential for Illegitimate Recombination in Bacterial Genomes and Its Consequences: From Duplications to Genome Reduction. *Genome Res* **13**, 1123–1132 (2003).
 111. De Paepe, M. *et al.* Carriage of λ Latent Virus Is Costly for Its Bacterial Host due to Frequent Reactivation in Monoxenic Mouse Intestine. *PLoS Genet.* **12**, e1005861 (2016).
 112. Feiss, M. & Siegele, D. A. Packaging of the bacteriophage lambda chromosome: Dependence of cos cleavage on chromosome length. *Virology* **92**, 190–200 (1979).
 113. Jardine, P. J. & Anderson, D. L. DNA packaging in double-stranded DNA phages. in *The bacteriophages* (eds. Abedon, S. T. & Calendar, R. D.) 49–65 (2005).
 114. Schicklmaier, P., Moser, E., Wieland, T., Rabsch, W. & Schmieger, H. A comparative study on the frequency of prophages among natural isolates of *Salmonella* and *Escherichia coli* with emphasis on generalized transducers. *Antonie Van Leeuwenhoek* **73**, 49–54 (1998).
 115. de Vargas, C. *et al.* Structure and function of the global ocean microbiome.
 116. Kenzaka, T., Tani, K. & Nasu, M. High-frequency phage-mediated gene transfer in freshwater environments determined at single-cell level. *ISME J* **4**, 648–659 (2010).
 117. Wang, M. *et al.* Estimating the contribution of bacteriophage to the dissemination of antibiotic resistance genes in pig feces. *Environ. Pollut.* **238**, 291–298 (2018).
 118. Brown-Jaque, M. *et al.* Antibiotic resistance genes in phage particles isolated from human faeces and induced from clinical bacterial isolates. *Int. J. Antimicrob. Agents* **51**, 434–442 (2018).
 119. Abeles, S. R., Ly, M., Santiago-Rodriguez, T. M. & Pride, D. T. Effects of Long Term Antibiotic Therapy on Human Oral and Fecal Viromes. *PLoS ONE* **10**, e0134941 (2015).
 120. Górska, A. *et al.* Dynamics of the human gut phageome during antibiotic treatment. *Comput Biol Chem* **74**, 420–427 (2018).
 121. Fernández-Orth, D. *et al.* Faecal phageome of healthy individuals: presence of antibiotic resistance genes and variations caused by ciprofloxacin treatment. *Journal of Antimicrobial Chemotherapy* **74**, 854–864 (2019).
 122. Jia, B. *et al.* CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **45**, D566–D573 (2017).
 123. Gibson, M. K., Forsberg, K. J. & Dantas, G. Improved annotation of antibiotic

- resistance determinants reveals microbial resistomes cluster by ecology. *ISME J* **9**, 207–216 (2015).
124. Gupta, S. K. *et al.* ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob. Agents Chemother.* **58**, 212–220 (2014).
125. Zankari, E. *et al.* Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* **67**, 2640–2644 (2012).
126. Munk, P. *et al.* Abundance and diversity of the faecal resistome in slaughter pigs and broilers in nine European countries. *Nature Microbiology* **3**, 898 (2018).
127. Ruppé, E. *et al.* Prediction of the intestinal resistome by a three-dimensional structure-based method. *Nat Microbiol* **4**, 112–123 (2019).
128. Wallace, J. C., Port, J. A., Smith, M. N. & Faustman, E. M. FARME DB: a functional antibiotic resistance element database. *Database (Oxford)* **2017**, (2017).
129. Enault, F. *et al.* Phages rarely encode antibiotic resistance genes: a cautionary tale for virome analyses. *ISME J* **11**, 237–247 (2017).
130. Kim, M.-S. & Bae, J.-W. Spatial disturbances in altered mucosal and luminal gut viromes of diet-induced obese mice. *Environ. Microbiol.* **18**, 1498–1510 (2016).
131. Allen, H. K. *et al.* Antibiotics in feed induce prophages in swine fecal microbiomes. *MBio* **2**, (2011).
132. Johnson, T. A. *et al.* The In-Feed Antibiotic Carbadox Induces Phage Gene Transcription in the Swine Gut Microbiome. *MBio* **8**, (2017).
133. Enault, F. *et al.* Phages rarely encode antibiotic resistance genes: a cautionary tale for virome analyses. *ISME J* (2016) doi:10.1038/ismej.2016.90.
134. Shkoporov, A. N. *et al.* Reproducible protocols for metagenomic analysis of human faecal phageomes. *Microbiome* **6**, 68 (2018).
135. Cardona, S. *et al.* Storage conditions of intestinal microbiota matter in metagenomic analysis. *BMC Microbiol.* **12**, 158 (2012).
136. Conceição-Neto, N. *et al.* Modular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis. *Sci Rep* **5**, 16532 (2015).
137. Kleiner, M., Hooper, L. V. & Duerkop, B. A. Evaluation of methods to purify virus-like particles for metagenomic sequencing of intestinal viromes. *BMC Genomics* **16**, (2015).
138. Castro-Mejía, J. L. *et al.* Optimizing protocols for extraction of bacteriophages prior to metagenomic analyses of phage communities in the human gut. *Microbiome* **3**, 64 (2015).
139. Zuo, T. *et al.* Bacteriophage transfer during faecal microbiota transplantation in *Clostridium difficile* infection is associated with treatment outcome. *Gut* (2017) doi:10.1136/gutjnl-2017-313952.
140. Yilmaz, S., Allgaier, M. & Hugenholz, P. Multiple displacement amplification compromises quantitative analysis of metagenomes. *Nature Methods* **7**, 943–944 (2010).
141. Džunková, M. *et al.* Direct sequencing from the minimal number of DNA molecules needed to fill a 454 picotiterplate. *PLoS ONE* **9**, e97379 (2014).
142. Kim, K.-H. & Bae, J.-W. Amplification Methods Bias Metagenomic Libraries of Uncultured Single-Stranded and Double-Stranded DNA Viruses. *Appl. Environ. Microbiol.* **77**, 7663–7668 (2011).
143. Roux, S. *et al.* Towards quantitative viromics for both double-stranded and single-stranded DNA viruses. *PeerJ* **4**, e2777 (2016).
144. Marine, R. *et al.* Caught in the middle with multiple displacement amplification:

- the myth of pooling for avoiding multiple displacement amplification bias in a metagenome. *Microbiome* **2**, 3 (2014).
145. Pedulla, M. L. *et al.* Origins of Highly Mosaic Mycobacteriophage Genomes. *Cell* **113**, 171–182 (2003).
 146. Aggarwala, V., Liang, G. & Bushman, F. D. Viral communities of the human gut: metagenomic analysis of composition and dynamics. *Mob DNA* **8**, 12 (2017).
 147. Marcel Martin, M. M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
 148. Modolo, L. & Lerat, E. UrQt: an efficient software for the Unsupervised Quality trimming of NGS data. *BMC Bioinformatics* **16**, 137 (2015).
 149. Atropos: specific, sensitive, and speedy trimming of sequencing reads [PeerJ]. <https://peerj.com/articles/3720/>.
 150. Sutton, T. D. S., Clooney, A. G., Ryan, F. J., Ross, R. P. & Hill, C. Choice of assembly software has a critical impact on virome characterisation. *Microbiome* **7**, 12 (2019).
 151. Roux, S., Hallam, S. J., Woyke, T. & Sullivan, M. B. Viral dark matter and virus-host interactions resolved from publicly available microbial genomes. *Elife* **4**, (2015).
 152. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
 153. Zhu, W., Lomsadze, A. & Borodovsky, M. Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res* **38**, e132–e132 (2010).
 154. Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucl. Acids Res.* **44**, D279–D285 (2016).
 155. Haft, D. H., Selengut, J. D. & White, O. The TIGRFAMs database of protein families. *Nucleic Acids Res* **31**, 371–373 (2003).
 156. Grazziotin, A. L., Koonin, E. V. & Kristensen, D. M. Prokaryotic Virus Orthologous Groups (pVOGs): a resource for comparative genomics and protein family annotation. *Nucleic Acids Res.* **45**, D491–D498 (2017).
 157. Fouts, D. E. Phage_Finder: Automated identification and classification of prophage regions in complete bacterial genome sequences. *Nucleic Acids Res* **34**, 5839–5851 (2006).
 158. Roux, S., Enault, F., Hurwitz, B. L. & Sullivan, M. B. VirSorter: mining viral signal from microbial genomic data. *PeerJ* **3**, e985 (2015).
 159. Ren, J., Ahlgren, N. A., Lu, Y. Y., Fuhrman, J. A. & Sun, F. VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome* **5**, 69 (2017).
 160. Arndt, D. *et al.* PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Research* **44**, W16–W21 (2016).
 161. Akhter, S., Aziz, R. K. & Edwards, R. A. PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. *Nucleic Acids Res* **40**, e126 (2012).
 162. Jang, H. B. *et al.* Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nature Biotechnology* **1** (2019).
 163. Amgarten, D., Braga, L. P. P., da Silva, A. M. & Setubal, J. C. MARVEL, a Tool for Prediction of Bacteriophage Sequences in Metagenomic Bins. *Front. Genet.* **9**, (2018).
 164. Zheng, T. *et al.* Mining, analyzing, and integrating viral signals from metagenomic data. *Microbiome* **7**, 42 (2019).
 165. Fang, Z. *et al.* PPR-Meta: a tool for identifying phages and plasmids from metagenomic fragments using deep learning. *Gigascience* **8**, (2019).
 166. Song, W. *et al.* Prophage Hunter: an integrative hunting tool for active prophages.

Nucleic Acids Res **47**, W74–W80 (2019).

167. Zhao, G. *et al.* VirusSeeker, a computational pipeline for virus discovery and virome composition analysis. *Virology* **503**, 21–30 (2017).

168. Roux, S., Tournayre, J., Mahul, A., Debroas, D. & Enault, F. Metavir 2: new tools for viral metagenome comparison and assembled virome analysis. *BMC bioinformatics* **15**, 76 (2014).

169. Wommack, K. E. *et al.* VIROME: a standard operating procedure for analysis of viral metagenome sequences. *Stand Genomic Sci* **6**, 427–439 (2012).

170. Rampelli, S. *et al.* ViromeScan: a new tool for metagenomic viral community profiling. *BMC Genomics* **17**, 165 (2016).

171. Lorenzi, H. A. *et al.* TheViral MetaGenome Annotation Pipeline(VMGAP):an automated tool for the functional annotation of viral Metagenomic shotgun sequencing data. *Stand Genomic Sci* **4**, 418–429 (2011).

172. Shannon, CE. A Mathematical Theory of Communication. (1948).

173. Simpson, E. H. Measurement of Diversity. *Nature* **163**, 688 (1949).

174. Bray, JR & Curtis, JT. A New Phylogenetic Method For Comparing Microbial Communities; **71**, 8228–8235 (2005).

175. Hill, T. C. J., Walsh, K. A., Harris, J. A. & Moffett, B. F. Using ecological diversity measures with bacterial communities. *FEMS Microbiol. Ecol.* **43**, 1–11 (2003).

176. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

177. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

178. Résumé des caractéristiques du produit - CEFTRIAXONE ALMUS 1 g/3,5 ml, poudre et solvant pour solution injectable (IM) - Base de données publique des médicaments. <http://base-donnees-publique.medicaments.gouv.fr/affichageDoc.php?specid=62676989&typedoc=R#RcpPropPharmacocinetiques>.

179. Résumé des caractéristiques du produit - CEFOTAXIME MYLAN 1 g, poudre pour solution injectable (IM-IV) - Base de données publique des médicaments. <http://base-donnees-publique.medicaments.gouv.fr/affichageDoc.php?specid=65053285&typedoc=R#RcpPropPharmacocinetiques>.

180. Michéa-Hamzhepour, M., Auckenthaler, R., Kunz, J. & Pechère, J. C. Effect of a single dose of cefotaxime or ceftriaxone on human faecal flora. A double-blind study. *Drugs* **35 Suppl 2**, 6–11 (1988).

181. Bräutigam, H. H., Knothe, H. & Rangoonwala, R. Impact of cefotaxime and ceftriaxone on the bowel and vaginal flora after single-dose prophylaxis in vaginal hysterectomy. *Drugs* **35 Suppl 2**, 163–168 (1988).

182. d’Humières, C. *et al.* A simple, reproducible and cost-effective procedure to analyse gut phageome: from phage isolation to bioinformatic approach. *Sci Rep* **9**, 11331 (2019).

183. Comeau, A. M., Tétart, F., Trojet, S. N., Prère, M.-F. & Krisch, H. M. Phage-Antibiotic Synergy (PAS): beta-lactam and quinolone antibiotics stimulate virulent phage growth. *PLoS ONE* **2**, e799 (2007).

184. Burdet, C. *et al.* Ceftriaxone and cefotaxime have similar effects on the intestinal microbiota in human volunteers treated by standard doses regimens. *Antimicrob. Agents Chemother.* (2019) doi:10.1128/AAC.02244-18.

185. Roose-Amsaleg, C. *et al.* Utilization of interferometric light microscopy for the

- rapid analysis of virus abundance in a river. *Res. Microbiol.* **168**, 413–418 (2017).
186. Wickramasinghe, N. H. *et al.* High community faecal carriage rates of CTX-M ESBL-producing *Escherichia coli* in a specific population group in Birmingham, UK. *J. Antimicrob. Chemother.* **67**, 1108–1113 (2012).
187. Nicolas-Chanoine, M.-H. *et al.* 10-Fold increase (2006-11) in the rate of healthy subjects with extended-spectrum β -lactamase-producing *Escherichia coli* faecal carriage in a Parisian check-up centre. *J. Antimicrob. Chemother.* **68**, 562–568 (2013).
188. Wipf, J. R. K., Schwendener, S. & Perreten, V. The novel macrolide-Lincosamide-Streptogramin B resistance gene *erm(44)* is associated with a prophage in *Staphylococcus xylosus*. *Antimicrob. Agents Chemother.* **58**, 6133–6138 (2014).
189. Draper, L. A. *et al.* Long-term colonisation with donor bacteriophages following successful faecal microbial transplantation. *Microbiome* **6**, 220 (2018).
190. Ott, S. J. *et al.* Efficacy of Sterile Fecal Filtrate Transfer for Treating Patients With *Clostridium difficile* Infection. *Gastroenterology* **152**, 799-811.e7 (2017).
-

Titre

Impact des antibiotiques sur le phageome intestinal humain: de la méthode à l'application.

Résumé

L'antibiorésistance est un problème majeur de santé publique. Mieux comprendre l'impact des antibiotiques sur le microbiote intestinal et l'acquisition de résistances est essentiel. Dans le cadre d'une étude clinique randomisée contrôlée (CEREMI), deux antibiotiques (ceftriaxone et céfotaxime) aux spectres d'activité antimicrobienne similaires, mais aux modes d'élimination différents, ont été administrés à 22 volontaires sains. Les selles de ces derniers ont été collectées à différents temps sur une période de 6 mois afin d'évaluer l'impact des antibiotiques sur la composition et l'évolution des communautés bactériennes et phagiques (virus bactériens) du microbiote intestinal ainsi que la présence de gènes de résistance aux antibiotiques.

L'étude du phageome intestinal a nécessité le développement de méthodologies adaptées tant au niveau expérimental que bioinformatique. Suite à la comparaison de différentes méthodes d'isolement de phages issus de fèces, la technique de concentration au polyéthylène glycol s'est révélée être la plus adaptée à notre étude au regard du grand nombre d'échantillons (>100) à analyser par la suite, de par sa simplicité, sa reproductibilité, son faible coût et de la grande diversité des phages obtenus. Un pipeline bioinformatique d'identification et d'annotation des phages, combinant des approches de métagénomique et de génomique comparative a été spécifiquement développé. Il permet, à partir des reads issus du séquençage métagénomique des phages, d'assembler, d'annoter syntaxiquement et fonctionnellement les différents contigs et enfin de prédire automatiquement s'ils sont des phages. Ces méthodologies ont ainsi été appliquées à l'ensemble des échantillons de selles issues de l'étude CEREMI.

Les analyses de la communauté phagique ont révélé que l'usage de ces deux antibiotiques entraîne une forte perturbation du phageome intestinal avec un retour progressif à l'équilibre au 30ème jour. L'usage de ceftriaxone semble cependant avoir un effet plus important sur la baisse de diversité phagique. Malgré cette perturbation, le phageome intestinal reste hautement individu-spécifique et ne semble pas impliqué dans le transfert de gènes de résistance aux antibiotiques.

En parallèle, l'étude de la communauté bactérienne, par des comptages bactériens et l'analyse des 16S, a montré que les deux antibiotiques entraînent une diminution rapide et drastique de la diversité bactérienne, sans pour autant sélectionner de bacilles à Gram négatifs résistants aux céphalosporines de troisième génération. Contrairement au phageome, les deux traitements semblent avoir un impact similaire sur le bactériome.

Mots clefs

Microbiote intestinal, antibiotiques, phages, phageome intestinal, gènes de résistance aux antibiotiques.