



HAL
open science

Bactériophages et vésicules extracellulaires du microbiote intestinal : développement de nouveaux outils d'étude et rôles dans les Maladies Inflammatoires Chroniques de l'Intestin (MICI)

Romain Sausset

► **To cite this version:**

Romain Sausset. Bactériophages et vésicules extracellulaires du microbiote intestinal : développement de nouveaux outils d'étude et rôles dans les Maladies Inflammatoires Chroniques de l'Intestin (MICI). Microbiologie et Parasitologie. Université Paris-Saclay, 2022. Français. NNT : 2022UPASB037 . tel-03935343

HAL Id: tel-03935343

<https://pastel.hal.science/tel-03935343>

Submitted on 11 Jan 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.

Bactériophages et vésicules extracellulaires du microbiote intestinal : développement de nouveaux outils d'étude et rôles dans les Maladies Inflammatoires Chroniques de l'Intestin (MICI)

Bacteriophages and extracellular vesicles of the gut microbiota: development of new study tools and roles in Inflammatory Bowel Disease (IBD)

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 581, agriculture, alimentation, biologie, environnement et santé (ABIES)

Spécialité de doctorat : Microbiologie

Graduate School : GS Biosphera. Référent : AgroParisTech

Thèse préparée dans l'unité **MICALIS Institute** (Université Paris-Saclay, INRAE, AgroParisTech), sous la direction de **Marianne DE PAEPE**, Chargée de recherche (HDR), le co-encadrement de **Luisa DE SORDI**, Maître de conférences (HDR) et de **Marie-Agnès PETIT**, Directrice de recherche, et la co-supervision de **Luc TALINI**, CEO

Thèse soutenue à Paris-Saclay, le 30 Juin 2022, par

Romain SAUSSET

Composition du Jury

Pascale BOULANGER Directrice de recherche, CNRS (Université Paris-Saclay)	Présidente
Amanda SILVA Chargée de recherche (HDR), CNRS (Université Paris Cité)	Rapporteur & Examinatrice
Claire LE HENAFF-LE MARREC Professeure, Bordeaux INP (Université de Bordeaux)	Rapporteur & Examinatrice
Benoit CHASSAING Directeur de recherche, INSERM (Université Paris Cité)	Examineur
Henri DUBOC MCU-PH (Université Paris Cité)	Examineur
Marianne DE PAEPE Chargée de recherche (HDR), INRAE (Université Paris-Saclay)	Directrice de thèse

Titre : Bactériophages et vésicules extracellulaires du microbiote intestinal : développement de nouveaux outils d'étude et rôles dans les Maladies Inflammatoires Chroniques de l'Intestin (MICI)

Mots clés : phages, vésicules, Crohn, microbiote, ILM

Bien que répandues dans les pays développés, les maladies inflammatoires chroniques de l'intestin (MICI), qui incluent la maladie de Crohn, ne sont que partiellement comprises. Si leurs symptômes sont désormais bien connus, ces pathologies sont difficiles à diagnostiquer chez les patients, du fait de l'absence de biomarqueur et de notre méconnaissance concernant l'étiologie de celles-ci. Il est cependant communément admis que ces pathologies s'accompagnent de déséquilibres du microbiote intestinal, appelés dysbioses. Qu'elles soient à l'origine ou la conséquence des MICI, ces dysbioses, majoritairement étudiées par métagénomique, ne concernent pas uniquement les bactéries du microbiote intestinal. En effet, les phages, prédateurs naturels des bactéries, pourraient, comme quelques publications en font l'hypothèse, jouer un rôle dans la maladie de Crohn.

L'objectif principal de ma thèse aura été de dénombrer les phages, ainsi que les vésicules extracellulaires (VE), également potentiellement impactées au cours des MICI, dans les selles d'individus sains et de patients atteints de la maladie de Crohn, afin de mieux comprendre leurs rôles dans cette maladie.

Dans un premier temps, j'ai optimisé un protocole basé sur l'utilisation de la microscopie par épifluorescence pour le dénombrement des Virus Like Particles (VLPs), qui, dans le microbiote intestinal, correspondent en majorité à des phages. Les selles d'individus sains et de patients atteints de la maladie de Crohn en période de poussée ou de rémission ont ainsi été analysées. Le ratio VLPs/bactérie s'est ainsi révélé être 100 plus élevé chez les patients en poussée par rapport aux individus sains, suggérant que les phages intestinaux sont beaucoup plus nombreux chez les patients en poussée.

J'ai également mis au point un protocole de dénombrement des vésicules extracellulaires, en utilisant le Videodrop (Myriade), un nouveau dispositif optique basé sur l'Interferometric Light Microscopy (ILM). La mise au point de ce protocole de dénombrement aura été l'occasion de comparer les performances du Videodrop à deux appareils reposant sur l'utilisation d'une technologie proche, le Nanoparticle Tracking Analysis (NTA). Bien qu'ayant un seuil de détection plus élevé que les appareils NTA, ce qui fait qu'une partie des vésicules ne peut pas être détectée, le Videodrop s'est révélé être le plus pratique et le plus rapide d'utilisation des trois appareils testés. Une augmentation d'un facteur trois de la médiane des ratio VE/bactérie a été observé dans le groupe des patients en phase de poussée par rapport aux individus sains.

Les résultats obtenus au cours de cette thèse mettent en lumière l'augmentation du nombre de particules virales dans les selles des patients atteints de la maladie de Crohn au cours des phases de poussée. Plusieurs hypothèses, qui restent à être vérifiées, pourraient expliquer ces résultats. Premièrement, il est possible que l'inflammation intestinale soit inductrice de prophages initialement présents au sein du microbiote. D'autre part, une soudaine multiplication de virus dans le microbiote intestinal pourrait conduire à une poussée inflammatoire. Si cette augmentation du ratio VLPs/bactérie est spécifique de la maladie de Crohn, ce qui reste à investiguer, cela pourrait permettre, à terme, la mise au point d'un nouveau marqueur de cette pathologie.

Title : Bacteriophages and extracellular vesicles of the gut microbiota: development of new study tools and roles in Inflammatory Bowel Disease (IBD)

Keywords : phages, vesicles, Crohn, microbiota, ILM

Although widespread in developed countries, chronic inflammatory bowel diseases (IBD), which include Crohn's disease, are only partially understood. Although their symptoms are now well known, these pathologies are difficult to diagnose in patients, due to the absence of biomarkers and our lack of knowledge regarding their etiology. However, it is commonly accepted that these pathologies are accompanied by imbalances in the intestinal microbiota, called dysbiosis. Whether they are the origin or the consequence of IBD, these dysbiosis, mostly studied by metagenomics, do not only concern the bacteria of the intestinal microbiota. Indeed, phages, natural predators of bacteria, could, as some publications hypothesize, play a role in Crohn's disease.

The main objective of my thesis was to enumerate phages, as well as extracellular vesicles (EVs), also potentially impacted during IBD, in the stools of healthy individuals and patients with Crohn's disease, in order to better understand their roles in this disease.

First, I optimized a protocol based on the use of epifluorescence microscopy for the enumeration of Virus Like Particles (VLPs), which, in the intestinal microbiota, correspond in majority to phages. The stools of healthy individuals and of patients with Crohn's disease in relapse or remission were analyzed. The VLPs/bacteria ratio was found to be 100 times higher in the stools of patients facing a flare than in healthy individuals, suggesting that intestinal phages are much more numerous during the crises.

I also developed a protocol for enumerating extracellular vesicles, using the Videodrop (Myriade), a novel optical device based on Interferometric Light Microscopy (ILM). The development of this enumeration protocol was an opportunity to compare the performance of the Videodrop to two devices based on a similar technology, the Nanoparticle Tracking Analysis (NTA). Although it has a higher detection threshold than the NTA devices, which means that some of the vesicles cannot be detected, the Videodrop proved to be the most practical and fastest to use of the three devices tested. A three-fold increase in the median EVs/bacteria ratio was observed in the group of patients in the flare phase compared to healthy individuals.

The results obtained during this thesis highlight the increase in the number of viral particles in the stools of patients with Crohn's disease during the flare phases. Several hypotheses, which remain to be verified, could explain these results. First, it is possible that the intestinal inflammation is an inducer of prophages initially present in the microbiota. Second, a sudden increase in viruses in the gut microbiota could lead to an inflammatory phase. If this increase in the VLPs/bacteria ratio is specific to Crohn's disease, which remains to be investigated, this could eventually lead to the development of a new marker for this pathology.

Thèse suivie par l'école doctorale n°581 : agriculture, alimentation,
biologie, environnement, santé (ABIES)

université
PARIS-SACLAY

ÉCOLE DOCTORALE

Agriculture, alimentation,
biologie, environnement,
santé (ABIES)

Cette thèse a été financée par

Myriade

The logo for Myriade, featuring the word "myriade" in a lowercase, sans-serif font. The letters are filled with a pattern of small dots, giving it a textured, digital appearance.

L'Association Nationale de la Recherche et de la
Technologie



L'Agence Nationale de la Recherche



Membres invités du Jury

Luisa DE SORDI

Maître de conférences (HDR), CRSA (Sorbonne Université)

Marie-Agnès PETIT

Directrice de recherche (HDR), INRAE (Université de Paris-Saclay)

Luc TALINI

CEO, Myriade

Remerciements

Je souhaite remercier toutes les personnes mentionnées ci-dessous, sans qui cette thèse n'aurait pas été possible :

Merci à Marianne De Paepe, ma directrice de thèse, pour ton encadrement au cours de toutes ces années. Merci pour ta confiance, ta patience, ta grande disponibilité et toutes nos discussions scientifiques. J'ai beaucoup appris à tes côtés et ce fut un immense plaisir de travailler avec toi.

Merci à Luisa De Sordi, co-encadrement de cette thèse, pour ta positivité à toute épreuve, ta disponibilité et ton accueil au sein des locaux du CRAS. Merci pour tous les échanges que nous avons pu avoir ensemble, qui furent très enrichissants.

Merci à Marie-Agnès Petit, co-encadrante de ma thèse, pour ta disponibilité et tes bons conseils. Merci pour les innombrables échanges que nous avons eus et merci de partager la science avec autant de passion.

Merci à Luc Talini, co-encadrant de thèse du côté de Myriade, pour ta confiance, tes conseils et nos grands moments vécus ensemble. Je n'oublierai jamais notre déplacement à Athènes et les situations rocambolesques dans lesquelles nous nous sommes retrouvés !

Merci à mon jury : Claire Le Hénaff, Amanda Silva, Pascale Boulanger, Benoît Chassaing et Henri Duboc d'avoir accepté d'évaluer mon travail.

Merci à mon comité de suivi de thèse : Benoît Chassaing, Christophe Gantzer et Gregory Caignard pour m'avoir accompagné au cours de ce projet. Merci pour vos remarques et nos discussions annuelles.

Merci à l'ensemble de l'équipe Phages : Marie-Agnès Petit, François Lecoïnte, Marianne De Paepe, Julien Lossouarn, Elisabeth Moncault, Olivier Son, Ilias Theodorou, Alice Eon-Bertho, Maud Billaud et Pauline Misson pour tous ces bons moments partagés ensemble. Si j'ai pris autant de plaisir à venir quotidiennement au laboratoire, c'est en particulier grâce à vous tous et à la super ambiance qui règne dans l'équipe.

Merci à l'équipe de Myriade : Luc Talini, David Poizat, Philippe Garabedian, François Mazuel, Matthieu Greffet, Marie Berger et Donatien Ramiandrisoa pour avoir suivi une partie de mes travaux

et pour m'avoir intégré au sein de votre équipe. Je souhaite remercier plus particulièrement Matthieu, Marie et Donatien pour les longues heures passées au téléphone à répondre à mes innombrables questions sur le Videodrop.

Merci à l'équipe de l'Institut Langevin : Claude Boccara, Martine Boccara, Ignacio Izzedin, Samer Alhadda et Houda Bey. Un grand merci à Samer pour le temps que tu as consacré à répondre à toutes mes questions ayant trait à la physique !

Merci à l'équipe Microbiote, Intestin et Inflammation du Centre de Recherche de Saint Antoine pour votre accueil chaleureux.

Merci à toutes les personnes m'ayant aidé à réaliser mes expérimentations et à analyser mes travaux :

- Ilias Theodorou, dont le projet était directement lié au mien. Échanger avec toi a toujours été simple et plaisant !
- Maud Billaud, pour tes astuces de bioinformatique,
- Youssra Boukrim, qui était venu faire son stage de M1 parmi nous, pour avoir réalisé les dénombrements de phages et de vésicules sur une dizaine d'échantillons
- Christine Longin, dans un premier temps pour l'acquisition de supers photos de mes phages et filtrats fécaux ; puis pour m'avoir appris à piloter cet incroyable engin qu'est le MET,
- Zuzana Krupova, pour nos échanges à propos du ZetaView,
- Alice Grangier, Eric Guedon et Sandrine Peron pour nos échanges à propos du NanoSight, et les expérimentations réalisées pour moi,
- Céline Henry et Aaron Millan-Oropeza, de la plateforme PAPPSO, pour avoir choyé mes échantillons et pour votre aide concernant l'analyse protéomique de ceux-ci,
- Cyrille Billaudeau, pour m'avoir partagé tes connaissances concernant la microscopie par épifluorescence et m'avoir aidé à de nombreuses reprises à ce sujet,
- Elisabeth Moncault, pour ta bonne humeur et tes excellents conseils à la paillasse. Les journées de paillasse passaient plus lentement quand tu ne manipulais pas à mes côtés !

Merci à mes deux super cothésardes, Maud et Pauline, ainsi qu'à Jeanne, notre coloc de bureau. Merci pour tous ces fous rires, les goûters, les soirées chez les uns et les autres, les barbecues, les week-ends et j'en passe ! Vivement nos retrouvailles !

Merci à tous mes autres copains du labo : Seb, Angelo, Aaron, Simon, Marine pour nos moments de détente et nos soirées passées ensemble. J'ai hâte de continuer à vous retrouver en dehors du laboratoire !

Merci à mes amis de très (trop) longue date, Cyrielle, Laura, Lucie, Mathilde, Quentin, Victorine et Virigine. Vous avez tous contribué à mon épanouissement personnel et professionnel. Merci de m'avoir offert un toit quand j'en avais le plus besoin, si j'ai pu aller au bout de mes études, c'est aussi grâce à vous !

Enfin, merci à famille : mes frères Mathieu, Theo et Julien, ainsi qu'à ma Maman pour vos encouragements et votre soutien depuis toujours. Merci d'être là quand j'ai besoin de recharger mes batteries et merci de faire partie de mon équipe contre les aléas de la vie !

Table des matières

Liste des Figures	9
Liste des Tableaux	11
Liste des abréviations	12
Introduction bibliographique	15
<i>Avant-propos</i>	16
Les Maladies Inflammatoires Chroniques de l'Intestin	16
Objectifs de la thèse	19
1. Les phages	21
1.1. Historique de l'étude des phages	21
1.2. Classification des phages	21
1.3. Cycles de vie des phages.....	24
1.4. Le dénombrement et la caractérisation des phages.....	25
2. Les Vésicules Extracellulaires (VE)	37
2.1. Composition et classification des vésicules extracellulaires	37
2.2. Rôles des VE	40
2.3. Techniques de caractérisation des VE.....	41
3. Les phages et les vésicules extracellulaires du microbiote intestinal	43
3.1. Composition du microbiote intestinal humain	43
3.2. Facteurs impactant la diversité du microbiote intestinal	46
3.3. Les fonctions du microbiote intestinal.....	47
3.4. Les virus du microbiote intestinal	48
3.5. Les VE du microbiote intestinal.....	52
4. Les dysbioses intestinales associées à la maladie de Crohn	54
4.1. Modifications du microbiote intestinal dans la maladie de Crohn	54
4.2. Altérations de la fraction virale du microbiote	57
4.3. Les rôles des VE dans la maladie de Crohn	58
Résultats	61
<i>Avant-propos</i>	62
1. Détermination des limites et des performances du Videodrop	63
1.1. Les gradients d'iodixanol	63
1.2. Dénombrement de populations homogènes de phages et de vésicules	64
2. Dénombrement spécifique de phages et de VE	101
2.1. Détection des VE et des phages par microscopie par épifluorescence	101

2.2. Destruction spécifique des VE.....	105
2.3. Conclusion.....	112
3. <i>Quantification de phages et de VE dans les selles d'individus sains et de patients atteints de la maladie de Crohn</i>	114
3.1. Constitution des cohortes.....	115
3.2. Dénombrement des bactéries dans les selles.....	115
3.3. Préparation des filtrats fécaux.....	119
3.4. Détection et dénombrement des phages dans les filtrats fécaux	120
3.5. Dénombrement et caractérisation des VE dans les selles	130
3.6. Nombre de VLPs et de VE par bactérie dans les selles	150
4. <i>Recherche de phages porteurs de gènes d'antibiorésistance dans des viromes de lapins d'élevage</i>	155
4.1. Introduction	155
4.2. Méthodes.....	155
4.3. Résultats.....	157
4.4. Conclusion.....	159
Matériel et Méthodes	163
<i>Destruction des VE</i>	164
<i>Dénombrement des bactéries dans les selles</i>	164
<i>Préparation des filtrats fécaux</i>	165
Le protocole dit « classique »	165
Le protocole dit « RNA later ».....	165
<i>Dénombrement des VLP dans les filtrats fécaux</i>	166
<i>Passage des filtrats fécaux sur colonnes SEC</i>	166
<i>Extraction des peptides des VE</i>	167
<i>Spectrométrie de masse en tandem par chromatographie liquide</i>	168
Discussion	169
Autres projets	176
<i>Article de Revue</i>	177
<i>Manuscrit de l'article 2</i>	189
Références bibliographiques	242

Liste des Figures

FIGURE 1. LES DIFFERENTS TYPES DE MICI. TIRE DE (YESHI ET AL. 2020).....	17
FIGURE 2. CLASSIFICATION DES PHAGES BASEE SUR LA MORPHOLOGIE DE LEURS VIRIONS ET LA NATURE DE LEUR MATERIEL GENOMIQUE. TIRE DE (DION, OECHSLIN, AND MOINEAU 2020).....	23
FIGURE 3. CYCLES DE VIE DES PHAGES. ADAPTE DE (SAUSSET ET AL. 2020).....	25
FIGURE 4. PRINCIPE DE LA TRPS. TIRE DE (COUMANS ET AL. 2014).....	29
FIGURE 5. VARIATIONS D'INTENSITES DE PARTICULES DE DIFFERENTES TAILLES, UTILISEES PAR LA DLS. TIRE DE (FALKE S. 2019).	30
FIGURE 6. EXPRESSION DE L'INTENSITE DE LA LUMIERE DIFFUSEE I.....	30
FIGURE 7. EQUATION DE STOKES-EINSTEIN.....	31
FIGURE 8. PRESENTATION DU VIDEODROP. ADAPTE DE (TURKKI ET AL. 2021).....	33
FIGURE 9. SCHEMA DE LA BIOGENESE DES VE EUCARYOTES. ADAPTE DE (DIAZ-GARRIDO ET AL. 2021).	38
FIGURE 10. ADAPTATION DES MODELES DE SECRETION DES VE BACTERIENNES PROPOSE PAR (YANG ET AL. 2018).	40
FIGURE 11. ADAPTATION DU MODELE DE MARQUEURS PROTEIQUES DE VE PROPOSE PAR (KOWAL ET AL. 2016).	43
FIGURE 12. COMPOSITION DU MICROBIOTE INTESTINAL HUMAIN. TIRE DE (ARUMUGAM ET AL. 2011).	44
FIGURE 13. DIVERSITE BACTERIENNE AU SEIN DU TRACTUS GASTRO INTESTINAL. ADAPTE DE (SEKIROV ET AL. 2010).....	45
FIGURE 14. DEVELOPPEMENT DU MICROBIOTE INTESTINAL HUMAIN. TIRE DE (CLEMENTE ET AL. 2012).	47
FIGURE 15. COMPOSITION DES VIROMES DE 7 INDIVIDUS SAINS. ADAPTE DE (SHKOPOROV ET AL. 2019).....	49
FIGURE 16. QUANTIFICATIONS ABSOLUES DE GENOMES VIRAUX PAR GRAMME, TOUT AU LONG DU TRACTUS GASTRO-INTESTINAL DE MACAQUES. TIRE DE (SHKOPOROV ET AL. 2021).....	51
FIGURE 17. DETERMINATION DE L'ORIGINE BACTERIENNE DE VESICULES FECALES DANS CHEZ 10 INDIVIDUS SAINS BASEE SUR LA METAGENOMIQUE DES VESICULES. TIRE DE (KAMELI ET AL. 2021).	53
FIGURE 18. COMPOSITION DU MICROBIOTE FECAL PAR ABONDANCE RELATIVE DES READS. ADAPTE DE (SABINO ET AL. 2016).....	54
FIGURE 19. SCHEMA PRESENTANT LA DENSITE DES FRACTIONS PRELEVEES ET LES NANOPARTICULES ATTENDUES DANS LES FRACTIONS	64
FIGURE 20. Φ X174 OBSERVES PAR MICROSCOPIE PAR EPIFLUORESCENCE.....	104
FIGURE 21. IMPACT DE L'ETAPE DE CENTRIFUGATION SUR LA QUANTITE DE PHAGES DANS LE SURNAGEANT	107
FIGURE 22. EFFETS DES DIFFERENTS TRAITEMENTS SUR LES VE ET LES PHAGES	108
FIGURE 23. IMPACT DU TRAITEMENT PAR LE NP-40 SUR LES DISTRIBUTIONS DE TAILLES DES PHAGES	109
FIGURE 24. CONCENTRATIONS DE VLPS.....	110
FIGURE 25. EFFETS DU TRAITEMENT PAR LE NP-40 SUR DES VE PURIFIEES, OBSERVEES PAR MET	112
FIGURE 26. METHODE DE GATING EMPLOYEE LORS DES ANALYSES DE CYTOMETRIE POUR DES DETERMINER LA CONCENTRATION DES SELLES EN BACTERIES.....	117
FIGURE 27. CONCENTRATION DES SELLES EN BACTERIES	118
FIGURE 28. MET DE FILTRATS FECAUX D'INDIVIDUS SAINS	120
FIGURE 29. NOMBRE DE VLPS DETECTEES DANS LES FRACTIONS	123
FIGURE 30. MET DE FRACTIONS VIRALES DE FILTRATS FECAUX D'INDIVIDUS SAINS	124
FIGURE 31. INOCULATION DE PHAGES EN QUANTITES CONNUES DANS DES FILTRATS FECAUX	125
FIGURE 32. FILTRAT FECAL R2 OBSERVE PAR MICROSCOPIE PAR EPIFLUORESCENCE	126
FIGURE 33. FILTRAT FECAL R9 OBSERVE PAR MICROSCOPIE PAR EPIFLUORESCENCE	127

FIGURE 34. DENOMBREMENT DES VLPS DANS LES FILTRATS FECAUX.....	128
FIGURE 35. COMPARAISON DE LA SEC ET DE LA CENTRIFUGATION.....	132
FIGURE 36. QUANTITES DE NANOPARTICULES AU COURS DU PROCESSUS DE PURIFICATION DES VE.....	133
FIGURE 37. OBSERVATION DES FRACTIONS ISSUES DU GRADIENT D'IODIXANOL PAR MET.....	134
FIGURE 38. CLUSTERING DES ECHANTILLONS SELON LEUR CONTENU, EN SOUS-GROUPES.....	135
FIGURE 39. REPARTITION DES SPECTRES EN FONCTION DE L'ORIGINE DE LEURS PROTEINES BACTERIENNES ASSOCIEES.....	136
FIGURE 40. REPARTITION DES PROTEINES EN GENRES BACTERIENS DANS LA FRACTION PURIFIEE DE VE LE FILTRAT FECAL DE L'INDIVIDU SAIN S2	137
FIGURE 41. CONCENTRATIONS EN NP DANS LES FRACTIONS 2 (VE PURIFIEES) DE SELLES D'INDIVIDUS SAINS.....	139
FIGURE 42. EFFETS DE LA PROTEINASE K SUR LES FLAGELLES.....	142
FIGURE 43. PROPORTIONS DE PHAGES ET DE VLPS RESISTANTS A LA PROTEINASE K.....	143
FIGURE 44. PROPORTIONS DE VE RESISTANTES A LA PROTEINASE K.....	144
FIGURE 45. EFFET DE LA DUREE DE L'INCUBATION LORS DU TRAITEMENT DE FILTRATS FECAUX PAR LA PROTEINASE K.....	145
FIGURE 46. EFFETS DE LA PROTEINASE K SUR LES FILTRATS FECAUX.....	146
FIGURE 47. FORMULE DU CALCUL DU NOMBRE DE VE DANS LES FILTRATS FECAUX.....	147
FIGURE 48. DENOMBREMENT DES VE DANS LES FILTRATS FECAUX.....	148
FIGURE 49. NOMBRE DE VE PAR BACTERIE.....	150
FIGURE 50. NOMBRE DE VLPS PAR BACTERIE.....	151
FIGURE 51. CONCENTRATIONS INDIVIDUELLES DE BACTERIES, VLPS ET VE.....	153
FIGURE 52. MET DE FILTRATS FECAUX D'INDIVIDUS DONT LA CONCENTRATION EN VLPS EST ELEVEE.....	154

Liste des Tableaux

TABEAU I. RENDEMENTS QUANTIQUES DE FLUORESCENCE	102
TABEAU II. CARACTERISTIQUES GENERALES DES FLUOROCHROMES UTILISES.....	103
TABEAU III. CARACTERISTIQUES DES ECHANTILLONS DE VIROMES DE LAPIN.....	160
TABEAU IV. CARACTERISTIQUES DES ARGs DANS LES ECHANTILLONS ANALYSES	161
TABEAU V. CONCENTRATIONS EN VLPs/G DES ECHANTILLONS ANALYSES.....	191

Liste des abréviations

ADN : acide désoxyribonucléique

AhR : ligand des récepteurs de l'aryle hydrocarboné

AIEC : *Escherichia coli* adhérent et invasif

ARN : acide ribonucléique

CD : cluster de différenciation

DLS : Dynamic Light Scattering

DSS : dextran sodium sulfate

EPI : microscopie par épifluorescence

Hsp : heat-shock protein

IL : interleukine

ILM : Interferometric Light Microscopy

MET : Microscopie Electronique à Transmission

MICI : Maladies Inflammatoires Chroniques de l'Intestin

NP : nanoparticule

NP-40 : Nonidet P-40

NTA : Nanoparticle Tracking Analysis

PBS : phosphate buffered saline

PCR : Polymerase Chain Reaction

pH : potentiel hydrogène

qPCR : Polymerase Chain Reaction quantitative

SEC : Chromatographie par exclusion de taille

TRPS : Tunable Resistive Pulse Sensing

UV : ultraviolet

VE : vésicule extracellulaire

Introduction bibliographique

Avant-propos

Les Maladies Inflammatoires Chroniques de l'Intestin

Bien qu'elles touchent près de 200 000 personnes en France (Observatoire National des MICI, page web, 21/3/22) et malgré leur prévalence dans les pays développés, les maladies inflammatoires chroniques de l'intestin (MICI), constituées de la maladie de Crohn et de la rectocolite hémorragique, ne restent encore aujourd'hui que partiellement comprises.

La maladie de Crohn

La maladie de Crohn, d'origine encore mal connue, est une affection handicapante qui peut toucher tous les segments du tube digestif, ainsi que certaines localisations extra digestives.

Sa physiopathologie repose sur un ensemble de facteurs génétiques et de facteurs environnementaux. Bien que de nombreux gènes de prédisposition soient connus (la mutation de NOD2-CARD15 est la plus fréquente) et que le tabagisme soit un facteur aggravant de la maladie (Berkowitz et al. 2018, Ni et al. 2017), l'origine de cette dernière n'est pas encore élucidée. Sa survenue est cependant corrélée au style de vie occidental et aux types de régimes qui lui sont associés, plus riches en aliments gras, en sucres simples, en aliments transformés et plus pauvres en fruits et légumes (Dixon et al. 2015, Dolan and Chang 2017, Reddavid et al. 2018).

Survenant à tout âge, la maladie de Crohn est le plus souvent diagnostiquée entre 20 et 30 ans, avec une légère prédominance féminine. Caractérisée par une forte inflammation, dont la localisation peut être iléale, colique ou iléocolique dans la plupart des cas, cette maladie évolue par crises avec des périodes de rémissions incomplètes (**Figure 1**). De plus, cette pathologie, comme toutes les MICI, se caractérise par la survenue de dysbioses, c'est-à-dire de déséquilibres observés dans le microbiote intestinal par rapport à un état de santé stable.

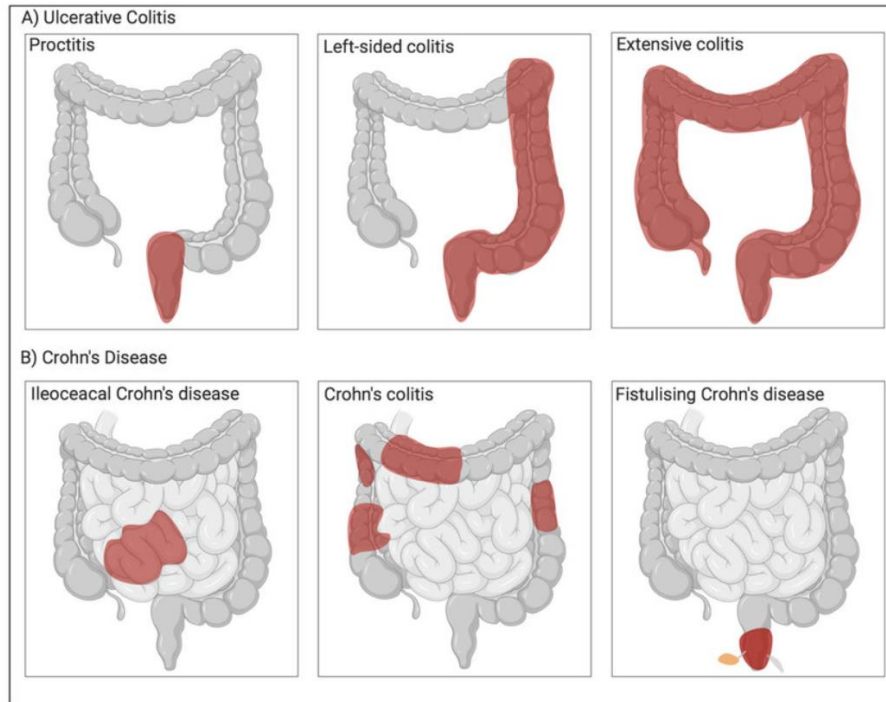


Figure 1. Les différents types de MICI. Tiré de (Yeshi et al. 2020).

(A) Différentes formes de la rectocolite hémorragique. (B) Différentes formes de la maladie de Crohn.

De nombreux symptômes peuvent caractériser la maladie de Crohn : des douleurs abdominales, des diarrhées, un syndrome dysentérique, un syndrome abdominal aigu pouvant simuler une appendicite aiguë et d'autres signes généraux : fièvre, fatigue, perte de poids, anorexie, pâleur, dyspnée, retard de croissance ou retard pubertaire chez les enfants et les adolescents, etc. Cependant, le diagnostic ne repose pas sur un test spécifique mais sur la combinaison de données cliniques, biologiques, endoscopiques (iléocoloscopie, endoscopie supérieure), histologiques et d'imagerie. Les biopsies peuvent révéler l'état inflammatoire du tube digestif. L'absence de test rapide spécifique au diagnostic de la maladie de Crohn rend le parcours de diagnostic parfois très long pour les patients, et retarde ainsi la mise en place d'un traitement adapté à ces derniers. Le dosage de la calprotectine fécale (une glycoprotéine notamment produite par les polynucléaires neutrophiles au cours de l'inflammation), permet cependant, en complément aux examens endoscopiques, de détecter l'inflammation intestinale au cours des crises. La concentration de cette protéine étant corrélée à l'activité inflammatoire (D'Haens et al. 2012, Mosli et al. 2015, Sipponen et al. 2008), elle être utilisée pour prédire la survenue de crises chez les patients (Mao et al. 2012).

Les complications de l'inflammation sont nombreuses : la malnutrition, les anémies inflammatoires et/ou carencielles et des retards de croissance chez l'enfant sont régulièrement

constatés chez les patients. Des complications telles que les occlusions intestinales aiguës peuvent également être observées. De plus, des formes perforantes ou fistulisantes peuvent provoquer des péritonites localisées ou généralisées, ainsi que des abcès intra-abdominaux, anopérinéaux et des fistules anopérinéales, entérales, vasculaires et entérocutanées. La survenue d'une d'atteinte colique étendue évoluant depuis plus de 8 ans est également considérée comme un facteur de risque pour la survenue d'un cancer colorectal.

Le premier traitement proposé est basé sur les corticoïdes, qui vont permettre une amélioration rapide de la crise, en réduisant la douleur et la diarrhée, améliorant ainsi l'état général du patient. Le traitement doit être poursuivi pendant plusieurs semaines et son arrêt doit être progressif.

En cas de résistance aux traitements par les corticoïdes dans les cas modérés et sévères de la maladie de Crohn, des anticorps monoclonaux comme l'infliximab et l'adalimumab (tous deux anti-TNF) ou l'ustekinumab (anti-IL) peuvent être préconisés pour modérer l'inflammation. Ces anticorps sont administrés par injection et peuvent provoquer des infections et des douleurs intenses. Le méthotrexate, un immunosuppresseur, peut également être administré par injection.

Une intervention chirurgicale peut être nécessaire en cas de sténose, de perforation de l'intestin, d'abcès dans la cavité abdominale ou de fistule. Cette chirurgie consiste à retirer les parties de l'intestin les plus touchées. La maladie de Crohn pouvant toucher toutes les parties de l'intestin, des récurrences peuvent se produire malgré l'intervention chirurgicale (VIDAL 2021).

La rectocolite hémorragique

La rectocolite hémorragique présente de nombreuses similitudes avec la maladie de Crohn, dont son occurrence corrélée au style de vie occidental (Reddavid et al. 2018) et l'occurrence de dysbioses intestinales liées. Sa principale différence porte sur sa localisation, moins diffuse dans l'intestin et plus localisée vers la partie distale du colon (**Figure 1**). Son diagnostic reste également difficile à établir.

L'évolution de la rectocolite hémorragique est imprévisible. Elle est caractérisée par des crises successives de durée variable, entrecoupées de phases de rémission plus ou moins complète. L'inflammation peut rester localisée ou s'étendre plus en amont avec le temps (environ 50% des cas après 10 ans) (**Figure 1**). L'une des principales complications de la rectocolite hémorragique est la

diarrhée aiguë sévère, qui se caractérise par des évacuations sanguinolentes plus de six fois par jour, le plus souvent associées à un amaigrissement, une fièvre, une anémie et un syndrome inflammatoire.

Le diagnostic est déterminé par une rectosigmoïdoscopie, voire une iléocoloscopie avec biopsies. Certaines formes peuvent être difficiles à différencier de la maladie de Crohn et la gastroscopie permet alors l'établissement d'un diagnostic différentiel. Comme les patients atteints de la maladie de Crohn, les patients atteints de rectocolite hémorragique sont traités par des corticoïdes, des immunosuppresseurs et/ou des anticorps.

Le traitement chirurgical de la rectocolite hémorragique consiste à retirer la totalité du gros intestin. En fonction de l'état de sa paroi interne, le rectum est conservé ou enlevé. En cas d'ablation, les symptômes de la rectocolite hémorragique disparaissent définitivement. Après l'ablation, la partie terminale de l'iléon est rattachée au rectum ou à l'anus (si le rectum a été amputé). Dans ce cas, le médecin crée une poche avant l'anus (appelée "réservoir iléal" ou "poche iléo-anale") où les selles s'accumulent avant d'être émises. Ces traitements sont handicapants pour les patients. Après l'ablation, des complications peuvent survenir. En effet, les personnes qui ont eu l'iléon rattaché au rectum ont un risque plus élevé de cancer du rectum et d'infection.

Les patients ayant eu l'iléon rattaché directement à l'anus peuvent souffrir d'une infection de la poche iléo-anale. Lorsque cette infection ne peut être endiguée par la prise d'antibiotiques, il peut être nécessaire de faire communiquer directement l'iléon avec la peau du ventre ("stomie" ou "anus artificiel", qui nécessite le port de poches externes pour recueillir les selles). La stomie est nécessaire chez environ 10% des personnes ayant subi une ablation du côlon et du rectum (VIDAL 2022).

Objectifs de la thèse

Il est aujourd'hui admis que les MICI sont liées à des états dysbiotiques du microbiote intestinal. Qu'elles soient à l'origine ou la conséquence de ces pathologies, ces dysbioses, majoritairement étudiées d'un point de vue bactérien, pourraient également s'accompagner d'importants changements au sein des communautés de bactériophages (ou « phages »), prédateurs naturels des bactéries. En conséquence de l'inflammation et d'une concentration en phages potentiellement plus élevées dans ces écosystèmes, la production et le relargage de vésicules extracellulaires (VE) dans le microbiote intestinal pourrait également être favorisée.

Les différences de concentrations en phages et en VE entre les selles d'individus sains et de patients atteints de la maladie de Crohn n'ont que rarement été étudiées. Ce travail de thèse a pour objectif de dénombrer les phages et les VE au sein de selles de patients atteints de la maladie de Crohn afin d'y observer d'éventuelles différences avec un groupe d'individus sains. Un protocole de dénombrement des phages et des VE dans les filtrats fécaux humains a ainsi dû être mis au point. Pour ce faire, le Videodrop, un nouvel appareil basé sur l'Interferometric Light Microscopy (ILM) et capable de détecter et de dénombrer certaines nanoparticules, a été utilisé. Développé par la société Myriade, qui finance cette thèse CIFRE, cet appareil n'avait encore jamais été testé et comparé à ses concurrents sur des phages et des VE au commencement de ce projet. Au cours de ce travail de thèse, le Videodrop a donc été comparé aux technologies déjà existantes pour le dénombrement des phages et des VE.

1. Les phages

1.1. Historique de l'étude des phages

La présence de phages dans l'intestin n'a été décrite que deux ans après leur découverte par Frederick Twort (Twort 1915), lorsque Félix d'Hérelle (d'Herelle 1917) a indépendamment découvert les phages ainsi que leur potentiel thérapeutique dans des selles de patients atteints de dysenterie. Quelques années plus tard en Union Soviétique, les phages ont été utilisés pour traiter diverses infections intestinales principalement causées par *Vibrio cholerae* (D'Herelle 1929) et *Shigella spp* (Babalova et al. 1968). Cependant, le succès de ces traitements ne fut que partiel et l'avènement des antibiotiques, qui se révélèrent à la fois plus efficaces et plus rentables, conduisit au quasi-abandon de la phagothérapie dans les pays occidentaux (Kutter et al. 2010). Avec l'augmentation de la résistance bactérienne aux antibiotiques, la phagothérapie a récemment regagné en intérêt, alimentant les recherches sur la biologie fondamentale et appliquée des phages. La découverte relativement récente de l'influence des phages dans les écosystèmes bactériens aquatiques explique également l'essor actuel des études sur les phages (Weinbauer and Rassoulzadegan 2004), ainsi que la prise de conscience de la possible importance du rôle des phages au sein du microbiote intestinal et par conséquent, en santé humaine.

1.2. Classification des phages

Les phages présentent un diamètre de capsidie compris entre 30 nm et 120 nm. La plupart des phages appartiennent à l'ordre des *Caudoviricetes* et possèdent une queue pouvant mesurer jusqu'à 250 nm. La taille de leurs génomes varie entre ~3,5 kb (comme le phage MS2 d'*Escherichia coli*, un *Leviviridae* à ARN simple brin (Fiers et al. 1976)) et ~540 kb (dans le cas du génome à ADN double brin des LAK phages, des myophages de *Prevotella spp.* (Devoto et al. 2019)). La classification traditionnelle des phages était basée sur la morphologie de leur capsidie, la présence ou non d'une queue et la composition de leur matériel nucléaire (**Figure 2**). Cette classification permet de distinguer les phages caudés en fonction de la morphologie de leur queue. On y retrouve ainsi les myophages, dont la queue est rigide ; les siphophages, pour lesquels la queue est souple ; ainsi que les podophages, qui possèdent une courte queue. Cependant, cette classification traditionnelle ne permet pas l'obtention de groupes monophylétiques, contrairement aux approches basées sur la séquence des génomes. Ainsi, suite à l'essor des méthodes d'analyse génomiques et la découverte de nombreuses nouvelles espèces phagiques, la classification traditionnelle basée sur la morphologie des virions fut progressivement abandonnée et remplacée par une méthode moléculaire de

classification, basée sur le pourcentage de gènes et de protéines partagés entre les espèces (Turner, Kropinski, and Adriaenssens 2021). Dans ce manuscrit, les familles taxonomiques définies sur des critères moléculaires sont désignées en italique. Celles qui ne le sont pas encore sont simplement désignées par leur morphotype.

La plupart des phages intestinaux chez l'Humain sont des *Microviridae* (à génome ADN simple brin) et des podophages (principalement des Crassphages, à ADN double brin) (Dutilh et al. 2014, Guerin et al. 2018, Shkoporov and Hill 2019, Shkoporov, Khokhlova, et al. 2018). D'autres familles phagiques sont également présentes dans le microbiote intestinal, comme les myophages, les siphophages, les *Tectiviridae*, ainsi que les *Inoviridae*.

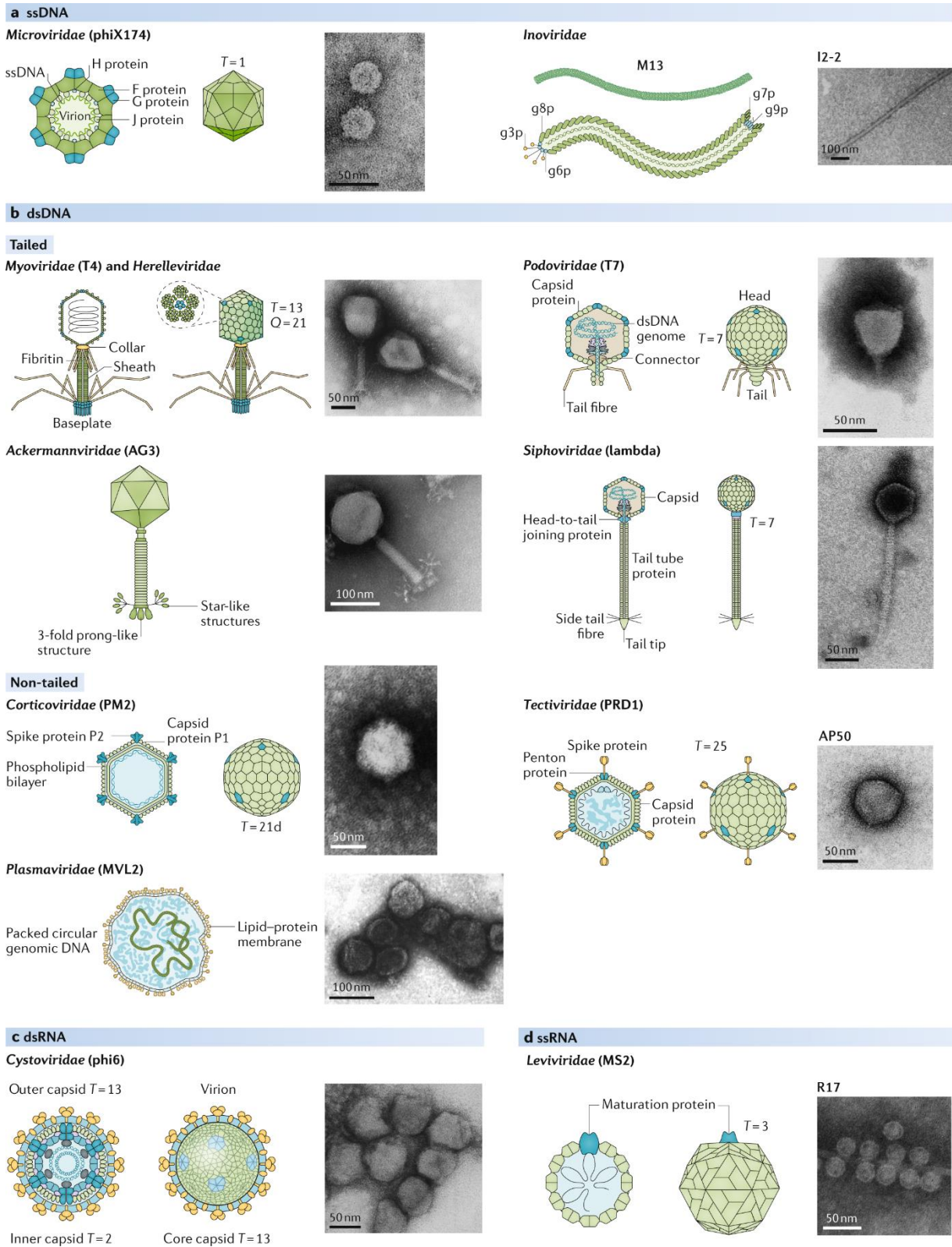


Figure 2. Classification des phages basée sur la morphologie de leurs virions et la nature de leur matériel génomique. Tiré de (Dion, Oechslin, and Moineau 2020).

Chaque type morphologique est représenté par un schéma et la photo (MET) d'un virion appartenant à cette catégorie.

1.3. Cycles de vie des phages

Dans l'intestin comme dans d'autres environnements, les phages peuvent être distingués sur la base de leur mode de vie, indépendamment de la taxonomie (**Figure 3**). Il existe différents types de phages : les virulents, les tempérés et les chroniques. Les phages virulents accomplissent des cycles lytiques, dans lesquels chaque infection est suivie de la production de virions et de la lyse de la cellule hôte. Cependant, dans certaines conditions, comme un stress nutritionnel, la multiplication des phages virulents peut être interrompue pendant une longue période. Ce phénomène, appelé pseudolysogénie, reste mal décrit mais est soupçonné d'exister dans l'environnement intestinal (Baess 1971, Los and Wegrzyn 2012, Ripp and Miller 1997). Les phages tempérés peuvent quant à eux réaliser deux types de cycles à travers le "choix de la lyse-lysogénie": l'infection est soit suivie d'un cycle lytique, comme pour les phages virulents, soit d'un cycle lysogénique au cours duquel le phage entre dans un état dormant et est appelé prophage. Dans cet état, des gènes sont exprimés pour maintenir la lysogénie et empêcher la multiplication active du phage via la lyse bactérienne. Le génome phagique se réplique passivement avec le génome bactérien. Suite à des signaux spécifiques déclenchés le plus souvent par une modification de l'environnement, comme une modification du pH ou de la température (Alves de Matos 2013, Choi, Kotay, and Goel 2010), une modification des nutriments disponibles dans le milieu (Lemaitre et al. 2015), une exposition à des rayonnements UV (Alves de Matos 2013) ou la présence d'antibiotiques (Humphrey, Stanton, and Jensen 1995, Masuda and Kawata 1979), les prophages peuvent être induits. Ce phénomène, appelé induction, entraîne le déclenchement du cycle lytique du phage et la lyse de la bactérie précédemment lysogène. Par conséquent, lorsque l'on considère les phages intestinaux, il faut à la fois tenir compte des phages libres et des prophages. L'inflammation, via la génération d'espèces réactives de l'oxygène dans l'intestin, pourrait être responsable de dégâts infligés à l'ADN bactérien et ainsi induire certains prophages (Diard et al. 2017, Kim and Bae 2016, Rigottier-Gois 2013).

Les phages filamenteux, quant à eux, réalisent un cycle chronique d'infection. Comme dans le cycle lysogénique, certains de ces phages sont également soumis au « choix lyse-lysogénie ». Contrairement aux phages virulents et tempérés, après la synthèse de nouveaux virions, les phages filamenteux s'échappent de leur hôte par des pores de petite taille, ce qui conduit rarement à la lyse bactérienne.

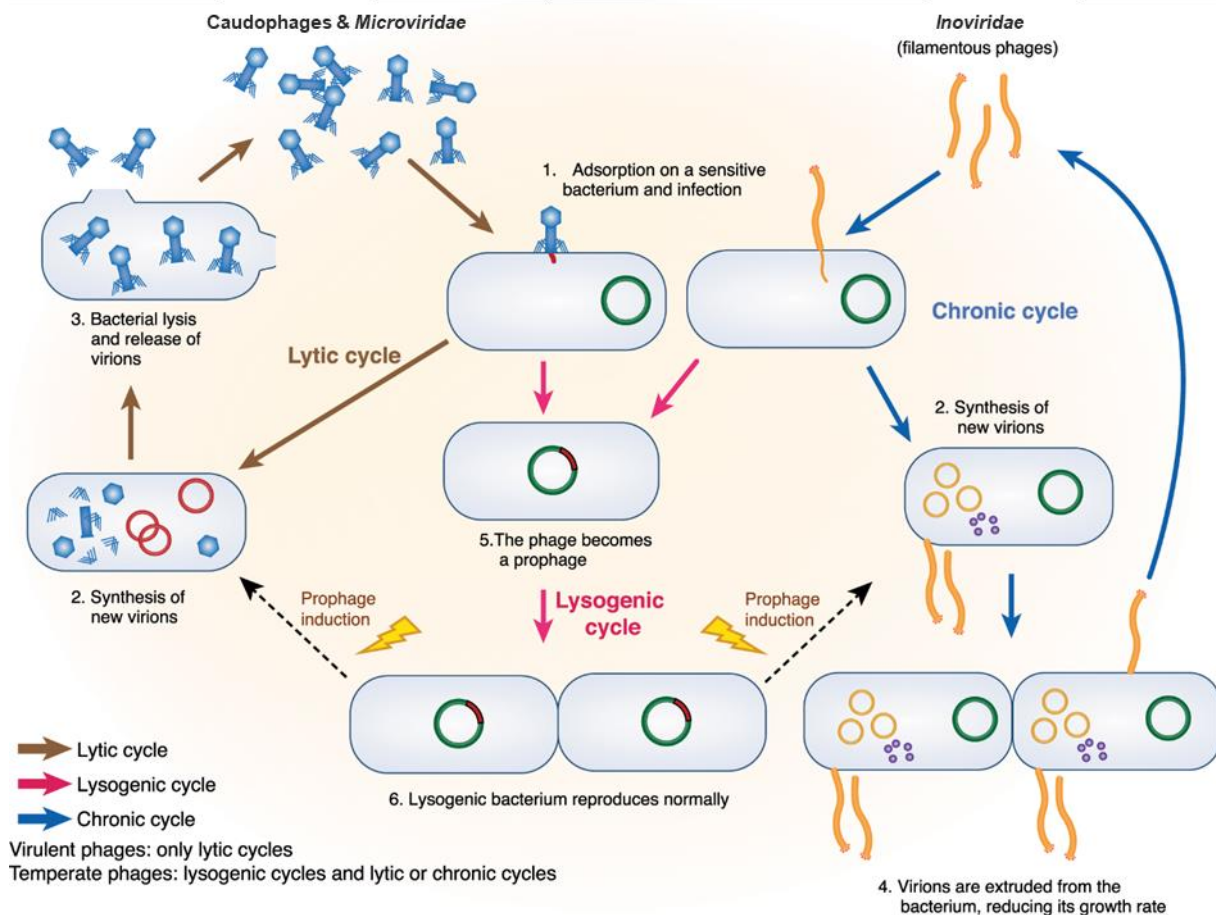


Figure 3. Cycles de vie des phages. Adapté de (Sausset et al. 2020).

La production de nouveaux virions est réalisée lors des cycles lytiques pour les phages caudophages et *Microviridae* (côté gauche de la figure, flèches brunes) ou par une infection chronique dans le cas des phages filamenteux, ou *Inoviridae* (flèches bleues). Tous deux débutent par l'adsorption d'un virion sur une bactérie hôte et son infection (1). L'ADN phagique est ensuite répliqué et de nouveaux virions sont synthétisés (2). Lors du cycle lytique, les nouveaux virions sont libérés au cours de la lyse bactérienne (3). Lors des cycles chroniques, les nouveaux virions des phages filamenteux sortent des bactéries sans les lyser via un appareil de sécrétion dédié (4). Les phages qui se reproduisent uniquement par des cycles lytiques sont dits virulents. Par opposition, les phages tempérés, réalisent des cycles lytiques ou chroniques et sont capables de réaliser des cycles lysogènes (flèches rose), à l'état de prophage (5). Ce prophage, intégré au génome bactérien ou à l'état épisomal, est répliqué avec le chromosome bactérien tant que les bactéries se divisent (6). Chez certaines bactéries, généralement lorsqu'elles sont soumises à un stress, le prophage est induit et le phage reprend un cycle lytique ou chronique.

1.4. Le dénombrement et la caractérisation des phages

Si la composition phagique de nombreux écosystèmes a été étudiée, le dénombrement absolu des phages au sein du microbiote intestinal humain est encore parcellaire et souffre de biais technologiques plus ou moins importants. La concentration en virions dans les selles d'individus sains

ou de patients atteints de la maladie de Crohn n'a à ce jour encore jamais été déterminée, notamment à cause du défi technologique que cela peut représenter.

Bien que nécessitant un appareillage onéreux, la microscopie électronique à transmission (MET) est l'une des technologies privilégiées pour la détection des virions de phages. Basée sur l'utilisation d'un faisceau d'électrons permettant l'obtention d'une résolution mille fois supérieure à celle des microscopes optiques, la MET permet de visualiser avec une grande résolution des objets nanométriques jusqu'à 0,2 nm (Acs, Gambino, and Brondsted 2020). Utilisée pour la première fois pour l'observation de phages en 1940 (Ruska 1940), elle est aujourd'hui couramment utilisée pour l'observation de virions, dont la concentration doit être suffisamment élevée ($\sim 10^9$ /mL). Si la MET permet de caractériser la morphologie des phages étudiés avec précision, elle reste néanmoins peu fiable concernant leur dénombrement (Ferris et al. 2002), du fait de la grande variabilité de l'efficacité de la coloration préalable des phages et du caractère aléatoire de leur adsorption et de leur répartition sur les grilles utilisées. La difficulté à différencier les phages non caudés d'autres objets biologiques comme les vésicules extracellulaires est aussi un facteur limitant de l'utilisation de la MET (Townsend et al. 2021).

D'autres méthodes de dénombrement des phages plus adaptées sont disponibles pour les étudier. Pour en saisir leurs avantages et leurs limites, il est nécessaire de comprendre quelles en sont leurs caractéristiques et fonctionnements.

1.4.1. Les méthodes classiques de dénombrement des phages

1.4.1.1. Les plages de lyse

La méthode la plus utilisée pour le dénombrement de phages est la détection de plages de lyse produites par l'action lytique de phages sur un tapis bactérien. Dans une boîte et sur une couche de milieu gélosé adapté à la croissance bactérienne, un mélange de phages, de leur bactérie hôte et éventuellement d'un milieu gélosé enrichi en nutriments et minéraux est déposé de façon uniforme au-dessus de la première couche de gélose. Après incubation de plusieurs heures à une température adaptée à la croissance des bactéries, un tapis bactérien plus ou moins lysé apparaît. Chaque plage de lyse correspondant à la présence initiale d'un phage et leur comptage permet de déterminer la concentration initiale de phages infectieux. Si cette méthode reste la plus répandue car simple à réaliser et peu onéreuse, elle ne permet que le dénombrement des phages infectieux et ne tient pas compte des potentiels agrégats phagiques ni des virions ayant perdu la capacité d'infecter leur hôte

(par exemple en s'adsorbant sur une vésicule ou une bactérie déjà infectée). De plus, cette méthode de dénombrement ne peut s'appliquer que pour les phages dont l'hôte bactérien est identifié et cultivable dans des conditions de laboratoire.

1.4.1.2. La réaction en chaîne par polymérase quantitative (qPCR)

Basée sur l'amplification spécifique de fragments d'ADN d'un phage ciblé, la qPCR permet une quantification précise et reproductible de phages connus (Acs, Gambino, and Brondsted 2020). Brièvement, son principe repose sur l'émission de fluorescence, proportionnelle à la quantité d'ADN mesurée (Mackay, Arden, and Nitsche 2002). Malgré sa rapidité, cette méthode de dénombrement ne peut s'appliquer qu'à des phages dont les génomes sont connus, puisqu'elle nécessite l'utilisation d'amorces spécifiques. À l'inverse des espèces bactériennes, il n'existe aucun gène marqueur spécifique des espèces virales, ne permettant ainsi pas une détection globale de ces derniers au sein d'écosystèmes par qPCR. Cependant, contrairement à la méthode des plages de lyse, la qPCR permet le dénombrement des phages indépendamment de leur infectiosité (Peng, Nguyen, and Ghosh 2018).

1.4.2. Les méthodes optiques de dénombrement des phages

1.4.2.1. La microscopie à épifluorescence

Également basée sur la fluorescence, la microscopie à épifluorescence se caractérise par sa capacité à marquer aspécifiquement toutes les particules virales d'un échantillon. Après déstabilisation des capsides phagiques par un choc thermique, celles-ci sont piégées par filtration sur des membranes Anodisc (0,02 μm) et marquées par un intercalant de l'ADN (le plus souvent un marqueur de la famille des SYBR) (Kaletta et al. 2020, Suttle 2010). Le filtre ainsi marqué est fixé sur une lame de microscopie et observé à l'aide d'un microscope à épifluorescence. Ce protocole, relativement peu onéreux et rapide à exécuter, permet la détection et le dénombrement de l'ensemble des particules virales au sein d'un échantillon complexe. Il présente néanmoins quelques désavantages. En effet, Patrick Forterre et co-auteurs (Forterre et al. 2013) ont rapporté le marquage de vésicules membranaires et de débris cellulaires contenant des acides nucléiques. Par ailleurs, la détection de phages aux génomes de petite taille et/ou à ARN ou à ADN simple brin, comme les *Microviridae*, peut s'avérer complexe. En effet, l'intensité de fluorescence détectée après marquage par des agents intercalant de l'ADN (qui ne s'insèrent donc dans les génomes à simple brin que grâce

à la conformation tridimensionnelle de celui-ci) est parfois difficile à différencier du bruit de fond de fluorescence (Holmfeldt et al. 2012).

1.4.2.2. La cytométrie en flux à haut débit, ou « flow virometry »

Initialement développée pour l'identification et la caractérisation de cellules entières, la cytométrie en flux repose sur la traversée de ces dernières à travers le faisceau d'un laser. Plusieurs signaux sont recueillis suite à ce passage, dont la lumière diffusée, qui donne une information sur la structure et la morphologie des cellules ; et la lumière absorbée, qui renseigne sur la taille des objets étudiés. Cette technologie a été adaptée comme système à haut débit (en faisant passer l'échantillon par un capillaire) pour la réalisation de mesures de quantification de la concentration en particules virales (Brussaard 2004, Brussaard, Marie, and Bratbak 2000). Comme pour la microscopie à épifluorescence, ce sont les acides nucléiques des particules virales, qui, après marquage par un agent fluorescent, peuvent être différenciés du bruit de fond présent dans l'échantillon (Lippe 2018, Rossi et al. 2015, Zamora and Aguilar 2018).

1.4.2.3. La technique de Tunable Resistive Pulse Sensing (TRPS)

La TRPS peut être utilisée pour obtenir la distribution des tailles et la concentration en nanoparticules d'un échantillon (Doyle and Wang 2019) et fonctionne selon le principe décrit dans la **Figure 4**. Cette technologie se heurte à des problèmes rendant parfois son utilisation peu pratique. En effet, elle repose sur le passage des nanoparticules à travers des pores dont les dimensions doivent être adaptées à celles des particules analysées, ce qui n'est pas toujours possible dans le cas d'échantillons inconnus ou hétérogènes en taille. Dans le cas contraire, les pores s'obstruent et, étant étirables, l'application d'une forte pression avec un piston peut modifier leur dimension initiale (Turkki et al. 2021, van der Pol et al. 2014).

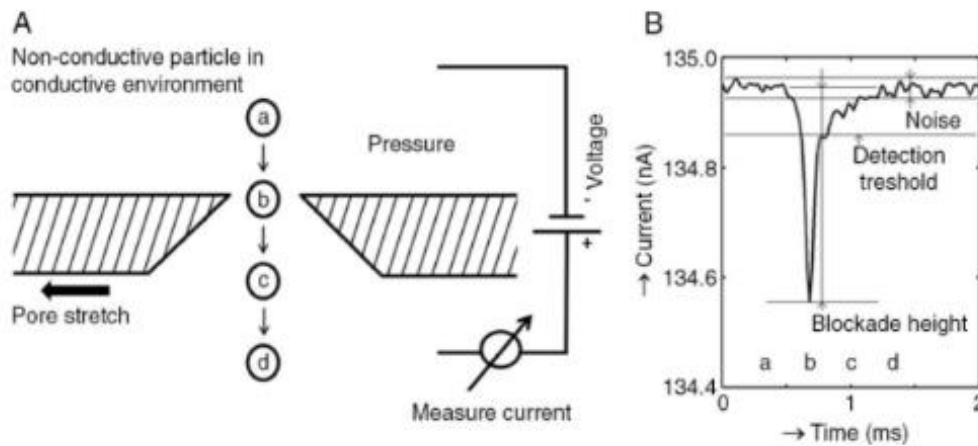


Figure 4. Principe de la TRPS. Tiré de (Coumans et al. 2014).

Dans un milieu conducteur, lorsqu'une particule non conductrice passe à travers un pore, il en résulte une brève augmentation de la résistance électrique du pore. Cette augmentation est mesurée en contrôlant le courant électrique à travers le pore. (A) montre une représentation schématique d'un pore avec une particule passant de la position *a* à la position *d*. (B) montre le courant mesuré suite au passage d'une particule se déplaçant à travers le pore, les lettres *a-d* correspondant aux positions *a-d* de la particule en (A). La différence de résistance du pore mesurée lors du passage de la particule est dépendante du rayon hydrodynamique de celle-ci.

1.4.2.4. La Dynamic Light Scattering (DLS)

La DLS repose sur l'illumination de l'échantillon par un laser et sur les variations d'intensité de la lumière diffusée par les particules. Ces fluctuations permettent de calculer le coefficient de diffusion des nanoparticules. Cette donnée permet de déterminer le mouvement brownien des particules, directement corrélé à leur rayon hydrodynamique, qui équivaut au rayon réel d'une sphère théorique ayant le même coefficient de diffusion que la particule observée et qui suivrait ainsi un mouvement brownien moyen identique (**Figure 5**) (Doyle and Wang 2019). Il est à noter que l'intensité de la lumière diffusée est proportionnelle à la puissance six du rayon des particules (**Figure 6**), ce qui rend la lumière diffusée des plus petites particules difficile à détecter, en particulier dans les mélanges où les tailles sont hétérogènes : les petites particules sont dissimulées par la lumière diffusée par les plus grosses particules. Ce phénomène s'appelle « l'effet d'écran », ou « effet de masquage » (Doyle and Wang 2019, Filipe, Hawe, and Jiskoot 2010, Palmieri et al. 2014).

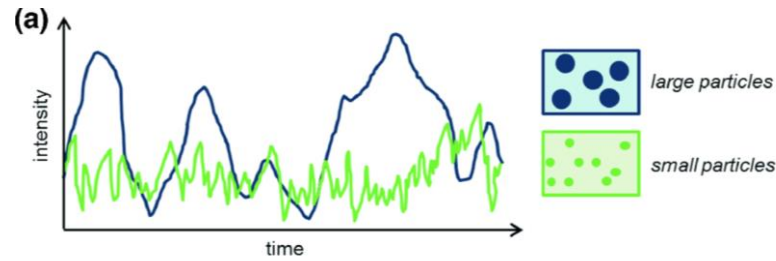


Figure 5. Variations d'intensités de particules de différentes tailles, utilisées par la DLS. Tiré de (Falke S. 2019).

$$I = I_0 \frac{16\pi^4}{\lambda^4} \frac{1 + \cos^2 \theta}{2s^2} \left(\frac{n^2 - 1}{n^2 + 2} \right)^2 r^6$$

Figure 6. Expression de l'intensité de la lumière diffusée I

Où I_0 = intensité du laser ; λ = longueur d'onde du laser; θ = angle de diffusion; s = distance avec la particule; n = indice de réfraction du milieu; r = rayon hydrodynamique de la particule

1.4.2.5. L'analyse de suivi des nanoparticules (NanoTracking Analysis ou NTA)

L'analyse NTA est, à ce jour, l'une des techniques les plus utilisées pour l'étude des nanoparticules et peut servir au dénombrement et à la détermination des tailles des phages (Kaletta et al. 2020). De même que pour la DLS, la trajectoire des nanoparticules dans un volume défini est suivie grâce à la lumière diffusée lors de l'illumination par un laser. L'application de la relation de Stokes-Einstein (**Figure 7**) permet la détermination des tailles des particules (Einstein 1905, William 1904). Le ZetaView (Particle Metrix, Allemagne) et le NanoSight (Malvern, Royaume-Uni) sont deux appareils reposant sur le NTA. Bien qu'ils soient basés sur le même principe, le ZetaView et le NanoSight présentent des différences importantes dans le dispositif optique et les logiciels d'analyse du signal (Bachurski et al. 2019). Par exemple, dans le cas du NanoSight, un flux est appliqué à l'échantillon par un système de pompe et de seringue, alors que l'échantillon est statique lors de l'acquisition par le ZetaView. Le ZetaView donne des mesures de concentrations plus précises pour les nanosphères de polystyrène et pour les vésicules extracellulaires isolées du sang ou de surnageants de cultures de cellules de Hodgkin (Bachurski et al. 2019). Le NanoSight fournit quant à lui des distributions de tailles plus proches des mesures réalisées en MET (Bachurski et al. 2019). Une étude de comparaison entre la DLS et le NTA a par ailleurs montré que le NTA est la plus précise des deux méthodes pour le dénombrement de billes de plastique calibrées (Filipe, Hawe, and Jiskoot 2010). Le NTA est cependant limité par l'effet d'écran dans les échantillons contenant des nanoparticules de tailles hétérogènes. Malgré leurs nombreux avantages, les techniques basées sur

le NTA sont relativement chronophages à cause du mode d'insertion des échantillons. Ces appareils sont sujets à la présence de bulles d'air dans leurs circuits. Difficiles à éviter lors de l'introduction des échantillons et parfois difficiles à déloger, ces bulles rendent toute analyse impossible tant qu'elles sont présentes. De plus, ces appareils se dérèglent facilement à cause de leur sensibilité aux vibrations, même faibles. Ces limitations les rendent peu pratiques pour une utilisation sur de nombreux échantillons.

$$D = \frac{kT}{6\pi r\mu}$$

Figure 7. Equation de Stokes-Einstein

Où D = coefficient de diffusion; k = constante de Boltzmann; T = température; r = rayon hydrodynamique de la particule; μ = viscosité du milieu

1.4.2.6. L'Interferometric Light Microscopy (ILM)

Le Videodrop, l'un des derniers appareils arrivés sur le marché du dénombrement et de la caractérisation de tailles des nanoparticules, se veut justement d'être rapide, fiable et simple d'utilisation. C'est la société Myriade qui le développe (**Figure 8B**) et qui finance ma thèse CIFRE. Cet appareil a été développé à partir d'un interféromètre (Boccaro et al. 2016, Roose-Amsaleg et al. 2017) utilisé pour dénombrer les nanoparticules dans différents fromages (Dugat-Bony et al. 2020) et dans des échantillons marins prélevés lors d'une mission de Tara Oceans (Boccaro et al. 2016). Comme les appareils basés sur le NTA, l'ILM utilise le mouvement brownien et l'application de la relation de Stokes-Einstein pour calculer les distributions de tailles des particules analysées. Cependant, ce n'est pas un laser mais une simple LED qui éclaire les échantillons. Un microscope à transmission à champ clair est utilisé comme interféromètre homodyme pour détecter les nanoparticules, sur la base des interférences créées par la lumière diffusée par celles-ci (**Figure 8A**). L'ILM est innovante par rapport à l'interférométrie classique dans la mesure où elle utilise une source lumineuse incohérente. Ainsi, la combinaison des ondes lumineuses résulte en l'addition des intensités de celles-ci. Par opposition, la combinaison d'ondes cohérentes crée des interférences, à l'origine d'un bruit de fond. L'ILM permet ainsi une soustraction simplifiée du bruit de fond des images acquises. L'un des avantages du Videodrop par rapport aux technologies basées sur le NTA est que la détection du signal diffusé n'est pas corrélée à l'intensité de celui-ci, mais à sa racine carrée, faisant ainsi passer la proportionnalité du signal d'une puissance 6 du rayon hydrodynamique à une puissance 3. Cette

différence notable permet, en théorie, la détection de nanoparticules de plus petits diamètres en ILM qu'en NTA et en DLS, ainsi qu'un effet d'écran moins important.

En fonction de la durée pendant laquelle elle est détectée, chaque particule est classée comme « suivie » ou uniquement « détectée » par le logiciel d'analyse du Videodrop, dénommé qvir. Que les particules soient uniquement détectées ou suivies, elles contribuent au dénombrement au sein des échantillons. Le diamètre hydrodynamique (corrélé à la taille des particules) n'est calculé qu'à partir des particules qui ont pu être suivies par le logiciel sur un nombre d'images imposé (**Figure 8C**). Un travail récemment publié par Vesa Turkki et co-auteurs a montré que le Videodrop permet le dénombrement relatif de vecteurs lentiviraux (~100 nm), d'adénovirus (d'un diamètre de 76 nm) et de baculovirus (300 x 50 nm) (Turkki et al. 2021). Cependant, le Videodrop a tendance à sous-estimer les concentrations de ces objets par rapport aux procédures classiques de titrage, en particulier pour les adénovirus (Turkki et al. 2021). Au cours de cette thèse, l'un de mes objectifs aura été de déterminer la performance du Videodrop pour le dénombrement de différents phages.

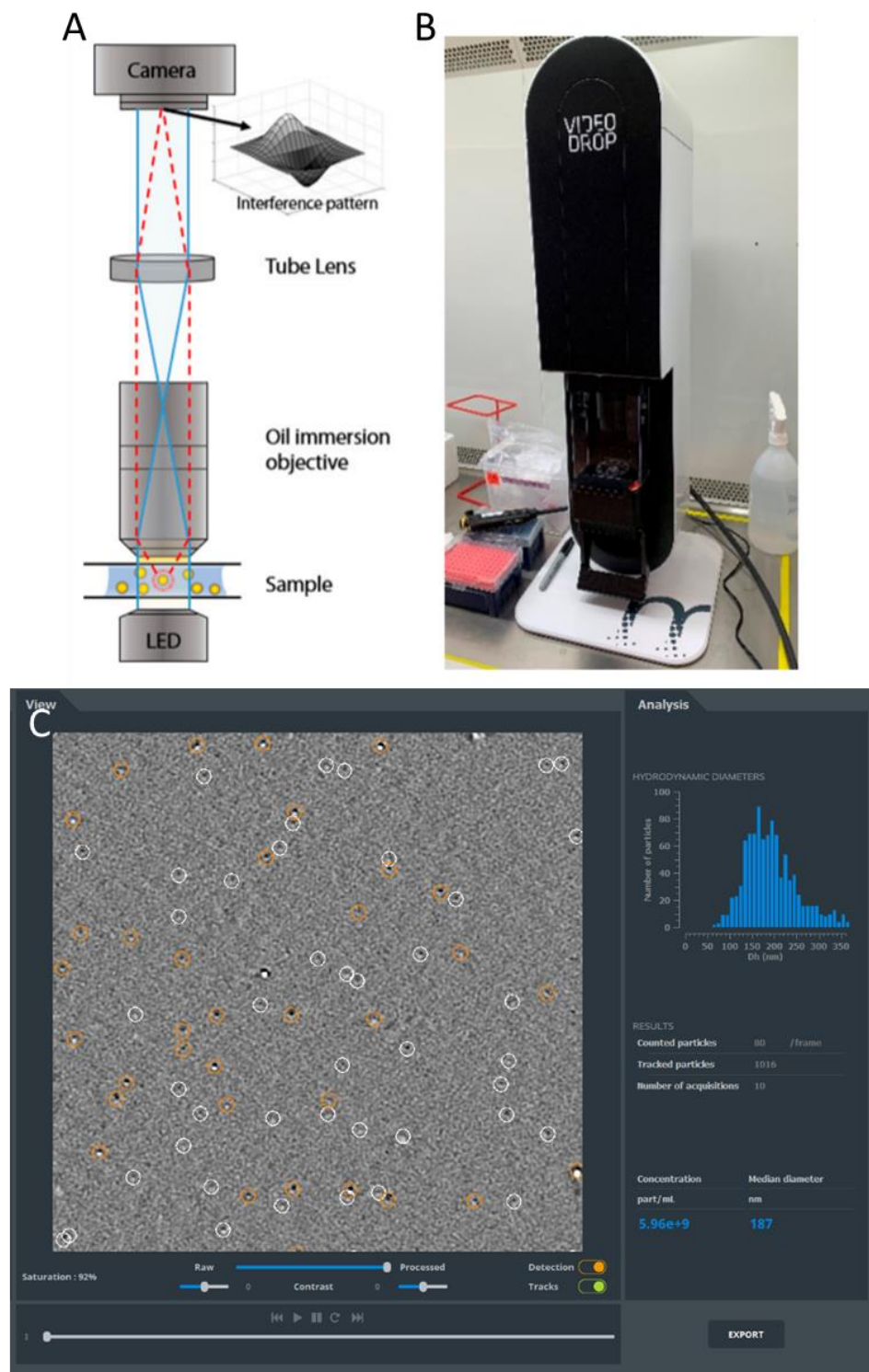


Figure 8. Présentation du Videodrop. Adapté de (Turkki et al. 2021).

(A) Schéma du principe de l'ILM utilisé dans le Videodrop. (B) Photographie du Videodrop. (C) Capture d'écran de qvir lors de l'analyse d'un échantillon. Les particules entourées en orange sont à la fois détectées et suivies sur plusieurs images (et participent ainsi au calcul de la concentration et de la distribution de tailles), tandis que les particules entourées en blanc sont détectées mais non suivies sur plusieurs images (et ne participent ainsi qu'au calcul de la concentration).

1.4.3. Les méthodes « -omiques » de caractérisation des phages

Bien qu'elles ne soient que très peu, voire pas quantitatives, d'autres méthodes de détection et d'identification des phages dans les écosystèmes tels que le microbiote intestinal existent. Basées sur des approches « -omiques », toutes ces techniques à haut débit se démarquent par leur capacité à délivrer de nombreuses informations concernant la diversité phagique au sein des microbiotes.

1.4.3.1. La métagénomique

Depuis le début des années 2000, les approches métagénomiques ont grandement contribué à la découverte d'une importante variété de microorganismes jusqu'alors inconnus car non-cultivables pour la plupart d'entre eux (Browne et al. 2016, Rajilic-Stojanovic and de Vos 2014, Shkoporov, Khokhlova, et al. 2018). Il existe deux principales approches en métagénomique. L'approche dite « shotgun » consiste à séquencer tous les acides nucléiques par complémentarité de bases. L'autre approche est basée sur l'amplification de gènes marqueurs des différentes espèces, comme le gène codant l'ARN ribosomique 16S, présent chez l'ensemble des espèces bactériennes (Hayes et al. 2017, Jovel et al. 2016). À cause de l'absence d'un tel gène marqueur des espèces virales, cette seconde approche n'est pas envisageable pour détecter les phages au sein d'écosystèmes. C'est pour cette raison que le séquençage de communautés entières (SCE) est à ce jour la seule possible pour l'exploration de la diversité phagique au sein d'écosystèmes microbiens (Hayes et al. 2017).

Pour que le séquençage soit le plus représentatif possible de l'échantillon, l'extraction du matériel génétique des virions présents dans les échantillons doit être optimale. Très schématiquement, les virions sont tout d'abord séparés du reste du microbiome par une filtration suivie par un traitement par une DNase. Leur ADN est ensuite purifié par des extractions successives par le phénol et le chloroforme. Les méthodes d'extraction ayant d'abord été développées en se focalisant sur les génomes à ADN, elles peuvent s'avérer limitantes pour les phages à ARN. Les étapes suivantes, comme la confection de bibliothèques et l'enrichissement des génomes viraux à séquencer, peuvent introduire des biais pour les génomes à ADN simple brin et négliger les phages à ARN. Les travaux en cours de l'équipe ont montré qu'au cours de cette étape, les génomes peuvent être plus ou moins amplifiés en fonction de leur nature nucléique et de la structure de leur brin d'ADN. Cela empêche par la suite la quantification fiable des génomes séquencés malgré l'ajout de séquences inoculées en quantités connues (Ilias Theodorou, résultats non publiés). Un autre facteur grandement limitant pour la détection des phages par métagénomique est le manque de richesse des banques de données actuelles concernant les phages présents dans l'environnement, surtout les phages à

ARN. Cependant, la communauté scientifique s'attèle de plus en plus à la résolution de cette problématique (Callanan et al. 2020, Shah et al. 2021).

Parallèlement à l'étude dédiées des viromes, des approches plus récentes consistent à rechercher l'ensemble des phages à ADN double brin dans les métagénomés issus du séquençage de microbiotes entiers, au sein desquels il est possible de détecter à la fois les phages fonctionnels et défectifs à l'état de prophage. Dans le génome de ces prophages défectifs, certaines régions contenant des gènes essentiels à l'accomplissement de leur cycle viral sont absentes ou mutées, empêchant alors la réalisation complète de celui-ci (Wang et al. 2010, Billaud et al. 2021).

Les populations phagiques détectées lors des analyses métagénomiques ne reflètent pas forcément leur réelle diversité, d'une part du fait du risque de perte des génomes de faible abondance au cours du processus d'extraction de ceux-ci, mais aussi à cause du risque d'introduire de nombreux biais lors de l'assemblage (Townsend et al. 2021). De façon à limiter le taux de mauvais assemblages, il est nécessaire de correctement définir les tailles minimales de contigs à conserver au cours de l'analyse, ainsi que la couverture des reads détectés. Si définir de tels seuils permet d'obtenir de faibles taux de faux assemblages, d'autres biais peuvent être introduits, comme l'exclusion des phages à ARN et à ADN simple brin dont les génomes sont bien plus petits que ceux des phages à ADN double brin. De tels seuils peuvent être également limitants pour la détection de prophages à cause de la nature modulaire de leur génome et de leur capacité à réaliser des processus de recombinaison (Lima-Mendez, Toussaint, and Leplae 2011).

Une fois le « catalogue » phagique établi pour un ensemble d'échantillons, une étape clé consiste ensuite à prédire quelles espèces bactériennes sont ciblées par ces derniers. De nouvelles approches pour la prédiction des hôtes bactériens des phages ont été mises au point au cours de cette dernière décennie. L'homologie entre les phages et les séquences spacer des CRISPR des hôtes bactériens permettent par exemple la prédiction de relations phages/hôtes (Stern et al. 2012). WISH, une approche *de novo* se basant sur les statistiques de présence de certains nucléotides, a également été développée à cet effet (Galiez et al. 2017). Une autre méthode *de novo*, baptisée meta3c (pour capture de la conformation des chromosomes) se base quant à elle sur le piégeage des molécules d'ADN dans leur configuration originale grâce aux propriétés de fixation du formaldéhyde et permet l'attribution des séquences phagiques à un ou plusieurs hôtes bactériens (Marbouty et al. 2017). Ainsi, après digestion de l'ADN par action enzymatique, des fragments d'ADN restent liés entre eux et leur probabilité de partager le même compartiment cellulaire n'en est que plus élevée.

1.4.3.2. La métatranscriptomique

La transcriptomique est une approche qui complète la métagénomique dans la mesure où elle permet de détecter les phages à ARN et d'étudier l'expression génique et par conséquent l'activité transcriptionnelle des phages à l'état de prophages. Cette approche d'alignement des reads de métatranscriptome sur les génomes phagiques permet ainsi de déterminer l'activité phagique et de différencier des phages actifs ou non. Ainsi, en analysant ces métatranscriptomes de façon longitudinale, il est possible d'étudier les dynamiques phages/bactéries au sein d'écosystèmes (Sieradzki et al. 2019).

1.4.3.3. La métaprotéomique

Derniers chaînons de l'expression des gènes, les protéines peuvent également être utilisées pour la détection des phages, dans la mesure où elles sont généralement spécifiques de leurs espèces phagiques productrices. Les capsides des phages étant des objets biologiques entièrement protéiques dans la plupart des cas (seuls quelques phages ont également une membrane lipidique), cette approche, bien que rarement citée dans la littérature pour l'étude des phages, semble prometteuse (Zhang et al. 2018). Brièvement, après purification puis digestion des protéines en peptides, ces derniers sont soumis à une spectrométrie de masse en tandem (MS/MS) de façon à déterminer leurs ratios masse sur charge (m/z) (Doyle and Wang 2019). La puissance de la MS/MS réside en sa capacité à déterminer des m/z spécifiques à chaque peptide, permettant ainsi de déterminer leur séquence, puis, après assemblage, de les associer à des groupes ou des sous-groupes de protéines. Ces groupes de protéines sont généralement spécifiques d'espèces phagiques, ce qui permet d'identifier les espèces présentes. Cette technique a par exemple permis la détection de protéines n'ayant pas été détectées par les approches de métagénomique utilisées sur des échantillons identiques (Ogilvie et al. 2013). La combinaison des approches métagénomiques et métaprotéomiques pourrait, à court terme, permettre d'augmenter la sensibilité de détection des phages au sein d'écosystèmes complexes. Cependant, actuellement, l'approche métaprotéomique ne permet pas de déterminer les concentrations des espèces phagiques et se heurte à certains biais partagés avec toutes les autres approches « -omiques », comme le faible référencement des phages à ARN et à ADN simple brin dans les bases de données.

2. Les Vésicules Extracellulaires (VE)

Les phages ne sont pas les seules nanoparticules présentes dans le microbiote intestinal et ayant potentiellement un lien avec les MICI. En effet, les VE, parfois difficilement différenciables des virions de phages en MET, pourraient, par le biais de leurs fonctions de communication intercellulaire et leur origine membranaire, être intimement liées à l'inflammation et l'état dysbiotique du microbiote des patients atteints de la maladie de Crohn.

2.1. Composition et classification des vésicules extracellulaires

Les vésicules extracellulaires sont des vésicules sécrétées par les cellules eucaryotes, les bactéries et les archées dans l'espace extracellulaire. Leur enveloppe peut être constituée d'une ou de deux couches lipidiques selon leur provenance (Liu et al. 2021, Yanez-Mo et al. 2015, Zaborowski et al. 2015). Leur présence dans tous les grands règnes du vivant met en avant leur importance (Woith, Fuhrmann, and Melzig 2019). Au regard de leur taille et de la biogénèse des VE eucaryotes, celles-ci sont classées en trois catégories (**Figure 9**) (van der Pol et al. 2012). Les exosomes, qui sont les plus petites, désignent des VE de 40 à 150 nm de diamètre, qui sont produites via les voies métaboliques lyso-endosomales de certaines cellules eucaryotes. Elles sont ensuite relarguées dans l'espace extracellulaire grâce à la fusion des corps multivésiculaires, dans lesquels elles sont produites, avec la membrane plasmique de la cellule productrice (Bebelman et al. 2018, Farooqi et al. 2018, Mathieu et al. 2019). La membrane lipidique des exosomes arbore quelques protéines marqueurs souvent présentes, comme les tétraspanines CD63, CD9 et CD81, ainsi que les protéines heat-shock HSP70 et HSP90 (Janockova et al. 2021). Le deuxième groupe de VE est composé des microvésicules, dont la taille varie entre 100 et 1000 nm. Elles sont produites à partir de la membrane plasmique des vésicules eucaryotes et, comme les exosomes, arborent des protéines de la famille des tétraspanines (CD9, CD63 et CD81) qui participent au bourgeonnement et à la fusion de la membrane (Simons and Raposo 2009). Enfin, les corps apoptotiques sont issus de la fragmentation cellulaire lorsque le cytosquelette est détruit au début de l'apoptose des cellules eucaryotes, leur taille pouvant varier de 1 à 10 μm . Contrairement aux deux autres familles de vésicules extracellulaires, les corps apoptotiques contiennent parfois des organelles intacts, de la chromatine et des petites quantités de protéines glycosylées (Escrevente et al. 2011, Kerr, Wyllie, and Currie 1972, Thery et al. 2001).

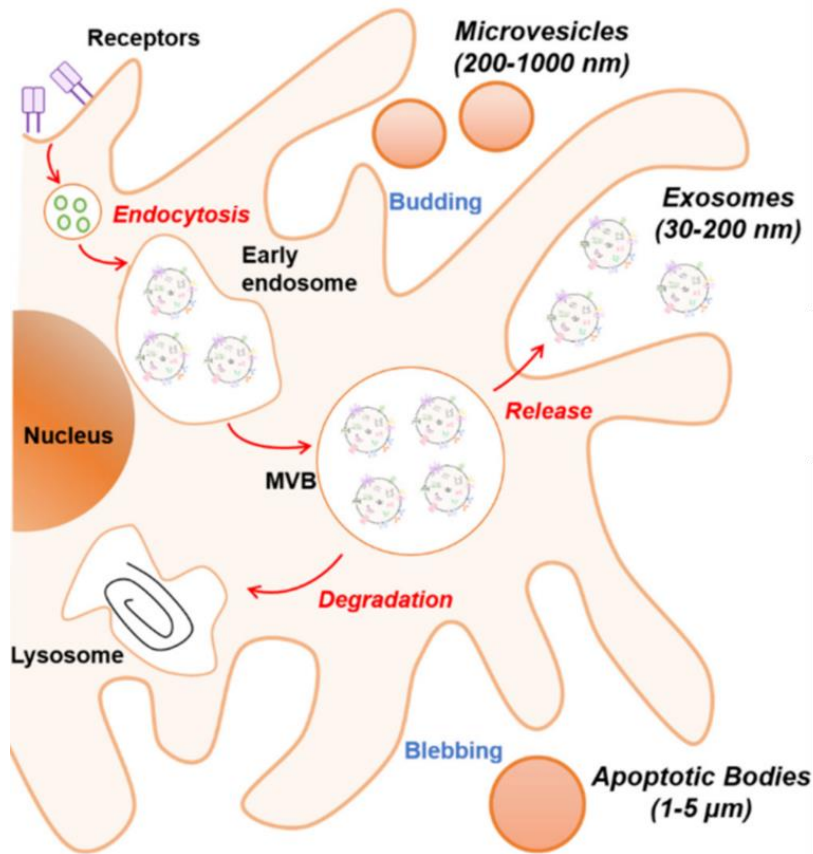


Figure 9. Schéma de la biogénèse des VE eucaryotes. Adapté de (Diaz-Garrido et al. 2021).

Bien que cette classification soit largement utilisée, elle ne s'applique qu'aux vésicules eucaryotes et exclut les vésicules d'origine bactérienne. Celles-ci sont parfois classées parmi le groupe des microvésicules, du fait de leur biogénèse. Leur diamètre peut cependant varier de 10 à 400 nm de diamètre (Alberti et al. 2021).

C'est en 1963 que la production de vésicules par des bactéries à Gram négatif a été mentionnée pour la première fois (Bladen and Waters 1963). Deux voies de genèse des vésicules extracellulaires par les bactéries à Gram négatif sont actuellement connues : par bourgeonnement de la membrane bactérienne externe ou par l'apoptose ou la lyse de la bactérie (**Figure 10B**) (Toyofuku, Nomura, and Eberl 2019, Yang et al. 2018). Contrairement aux VE eucaryotes, les vésicules extracellulaires produites par les bactéries à Gram négatif, souvent appelées OMVs pour Outer Membrane vesicles, sont caractérisées par la présence d'une couche externe de lipopolysaccharides et d'une couche interne phospholipidique (**Figure 10A**). Cette bicouche est également enrichie de peptidoglycanes et de protéines solubles, du fait de leur origine, i.e. la membrane externe bactérienne (Schwechheimer and Kuehn 2015).

Plus récemment, des vésicules produites par des bactéries à Gram positif ont également été observées, d'abord chez *Bacillus subtilis* et *Bacillus cereus*, puis dans d'autres espèces comme *Staphylococcus aureus* (Dorward and Garon 1990, Lee et al. 2009). Appélées MVs, pour Membrane vesicles, elles sont particulièrement enrichies en acide lipotéichoïque (Brown et al. 2015), comme les parois des bactéries à Gram positif. Contrairement aux OMVs, elles ne contiennent pas de lipopolysaccharides (**Figure 10A**) (Liu et al. 2018).

Qu'elles soient d'origine eucaryote ou procaryote, les vésicules peuvent contenir et transporter de l'ADN, de l'ARN, des lipides et des protéines, qu'elles protègent des enzymes présents dans le milieu extracellulaire (Bebelman et al. 2018, Yang et al. 2018, Zaborowski et al. 2015).

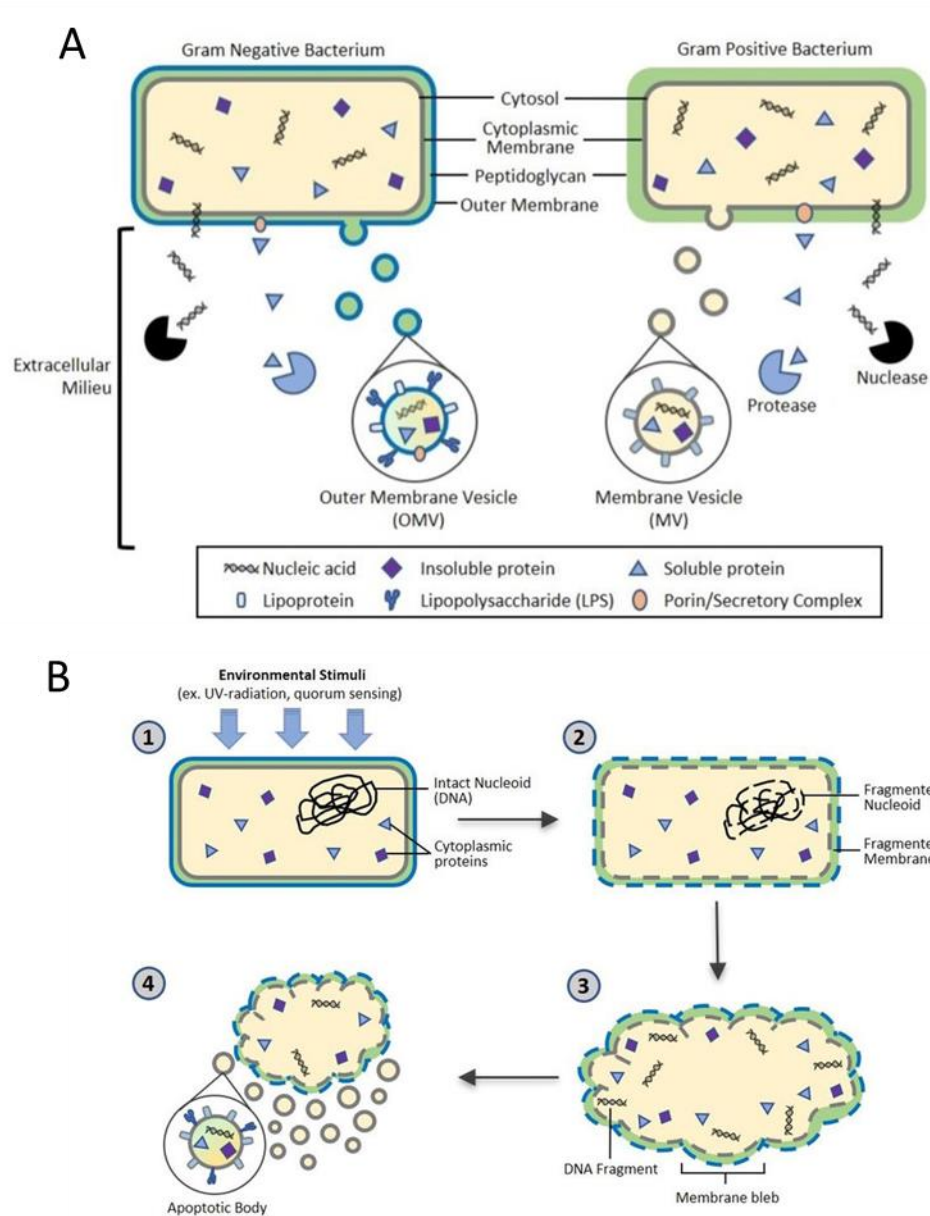


Figure 10. Adaptation des modèles de sécrétion des VE bactériennes proposé par (Yang et al. 2018).

(A) Sécrétion active d'VEs par les bactéries à Gram négatif ou positif. (B) Lors de l'apoptose induite par des stimuli environnementaux tels que le quorum sensing ou les dommages causés par les UV, l'ADN du bactérien est dégradé et fragmenté. S'ensuit alors la fragmentation de la membrane cellulaire, qui conduit finalement à l'éclatement de la membrane cellulaire, qui entraîne la libération de corps apoptotiques bactériens.

2.2. Rôles des VE

Bien qu'elles aient autrefois été considérées comme un simple moyen d'évacuation des déchets cellulaires (Johnstone et al. 1987), de nombreuses autres fonctions des VE ont été découvertes, comme leur importance dans la communication intercellulaire pour les VE eucaryotes

(Doyle and Wang 2019). Les VE sont présentes dans de nombreux fluides corporels humains (Slomka et al. 2018), où elles sont impliquées dans les processus homéostatiques tels que la coagulation et l'inflammation (Hessvik et al. 2016, Ratajczak et al. 2006, Simons and Raposo 2009, Thery, Ostrowski, and Segura 2009, Valadi et al. 2007, Vallabhaneni et al. 2015, van der Pol et al. 2012) et constituent des sources prometteuses de biomarqueurs pour de nombreuses pathologies, en particulier dans certains cancers (Lane et al. 2018). Elles sont également impliquées dans les maladies neurodégénératives et cardiovasculaires (Robbins, Dorronsoro, and Booker 2016, Thompson et al. 2016) et participent à la communication entre les neurones et les cellules gliales (Fruhbeis et al. 2013). Les VE bactériennes jouent également des rôles importants dans certaines pathologies, puisque certaines bactéries pathogènes délivrent des composés toxiques via les VE lors de l'infection des cellules (Bitto et al. 2017, Bomberger et al. 2009, Codemo et al. 2018, Furuta, Takeuchi, and Amano 2009). Enfin, les VE suscitent un énorme intérêt dans l'industrie car elles peuvent être utilisées comme véhicules pour délivrer des composés thérapeutiques actifs et servir de biomarqueurs (pour une revue du sujet, voir (Mateescu et al. 2017)).

2.3. Techniques de caractérisation des VE

Contrairement aux phages, il n'existe pas de méthode de dénombrement absolu des VE permettant de toutes les détecter au sein d'un échantillon, en particulier si la population étudiée présente des VE de tailles hétérogènes. La microscopie à épifluorescence, bien que théoriquement utilisable à ces fins, fait face à de nombreuses limitations découlant de la difficulté de marquer les vésicules. En effet, il n'existe pas de protéine marqueur commune à toutes les VE, ni même à une famille spécifique. Contrairement aux phages, la majorité des vésicules extracellulaires ne contiennent pas suffisamment d'acides nucléiques pour être visualisées avec des fluorophores. Puisqu'au cours de ma thèse j'ai été confronté à de nombreuses VE de nature inconnue, je me suis particulièrement intéressé aux approches physiques pour les dénombrer, ainsi qu'aux approches métaprotéomiques pour en déterminer leur nature.

2.3.1. Les approches physiques de détection des VE

La microscopie électronique à transmission peut être utilisée pour visualiser les VE, mais, comme pour les phages, elle ne peut pas être quantitative. De plus, la taille des VE par mesurée en

MET n'est pas précise, car un rétrécissement peut être observé en raison des étapes de déshydratation lors de la préparation de l'échantillon (Jensen, Prause, and Laursen 1981, Momen-Heravi et al. 2012, Szatanek et al. 2017, They et al. 2006, van Niel, D'Angelo, and Raposo 2018). Ce phénomène de rétrécissement peut être évité par l'utilisation de la Cryo-MET, mais elle est encore plus difficile d'accès et chronophage que la MET classique.

Utilisables pour la quantification et la détermination de toutes les nanoparticules, pourvu que leur taille soit suffisamment élevée pour être détectées, les appareils se basant sur le NTA comme le NanoSight et le ZetaView le sont donc également pour les VE, quelle que soit leur origine. Plusieurs équipes ont par ailleurs utilisé ces appareils pour la caractérisation de VE, comme Daniel Bachurski et co-auteurs (Bachurski et al. 2019), qui a comparé les deux appareils. Brièvement, leurs travaux montrent qu'aucun des deux appareils n'est capable de suivre les particules de moins de 60 nm (c'est-à-dire de les suivre sur suffisamment d'images successives pour en déduire leur taille selon le calcul de leur mouvement brownien). Cependant, ils concluent que le NanoSight est l'appareil le plus performant pour la détermination des tailles de VE, en particulier si leurs tailles sont homogènes. Cependant, le ZetaView fournit des résultats de dénombrement plus précis que le NanoSight. Bien que la DLS soit également utilisable pour l'analyse de VE, cette technologie souffre, comme exposé ci-avant, d'un fort effet d'écran au sein de populations de VE de tailles hétérogènes, la rendant moins sensible que les approches NTA pour les échantillons avec lesquels j'ai dû travailler au cours de ma thèse. L'utilisation de la TRPS n'a pas été envisagée non plus, pour ces mêmes raisons.

2.3.2. L'utilisation de la métaprotéomique pour l'étude des VE

Tout comme pour les phages, la métaprotéomique peut être envisagée comme une approche permettant la caractérisation des vésicules. Bien qu'elle ne soit pas utilisable pour le dénombrement de celles-ci, elle permet la détection de protéines spécifiques de certains types vésiculaires ou d'espèces bactériennes productrices. Cependant, cette approche très sensible ne permet pas de différencier les protéines extraites des VE des protéines libres présentes dans les échantillons à analyser. Il faut donc soigner la purification des VE avant de les soumettre à de telles analyses. Cette approche très puissante a par ailleurs déjà permis l'identification et le développement de plusieurs biomarqueurs de diverses pathologies dont des cancers (Doyle and Wang 2019). Dans un souci de clarification de la classification des VE, l'équipe de Joanna Kowal et co-auteurs (Kowal et al. 2016) a par ailleurs proposé de nombreux marqueurs protéiques spécifiques des VE eucaryotes, en fonction

de leur taille et de leur densité, en séparant des VE de cellules dendritiques humaines par ultracentrifugation différentielle: grandes VE denses, grandes VE peu denses, petites VE denses et petites VE peu denses. Ces quatre sous-groupes ont alors été analysés en protéomique et les résultats obtenus ont permis à Joanna Kowal et co-auteurs de suggérer l'utilisation de certains marqueurs protéiques spécifiques à chaque sous-groupe, tels que présentés dans la (**Figure 11**).

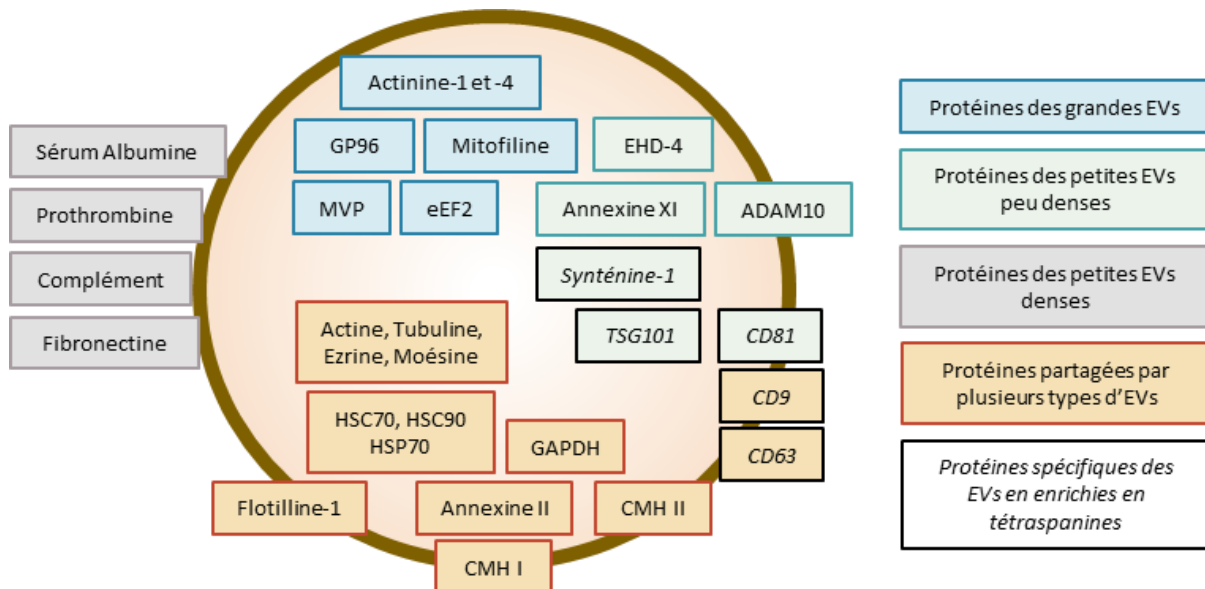


Figure 11. Adaptation du modèle de marqueurs protéiques de VE proposé par (Kowal et al. 2016).

Les VE sont schématisées par une membrane lipidique (cercle brun épais), entourant le cytosol (fond brun clair). Marron : protéines partagées par plusieurs types de VE. Vert : protéines spécifiquement enrichies dans petites VE peu denses. Gris : protéines spécifiquement enrichies dans les petites VE denses. Bleu : protéines spécifiquement enrichies dans les VE de grande et moyenne taille. En italique: protéines spécifiques des exosomes dérivés d'endosomes et enrichis en tétraspanines.

3. Les phages et les vésicules extracellulaires du microbiote intestinal

3.1. Composition du microbiote intestinal humain

Le tractus gastro-intestinal des humains et des autres espèces animales est un écosystème très complexe. Souvent considéré comme un organe à part entière et très variable d'un individu à l'autre (Flores et al. 2014, Mehta et al. 2018, Pasolli et al. 2019), le microbiote intestinal humain correspond à l'ensemble des micro-organismes commensaux retrouvés au sein de l'intestin. Celui-ci est composé de trillions de micro-organismes (Cresci and Bawden 2015), dont des bactéries

(Arumugam et al. 2011), des archées (Wampach et al. 2017), des micro-organismes eucaryotes (Laforest-Lapointe and Arrieta 2018) et des virus (Liang et al. 2020, Manrique et al. 2016, Shkoporov and Hill 2019). Les bactéries représentent la plus grande proportion de ces micro-organismes (Vandeputte et al. 2017), majoritairement représentées par des bactéries anaérobies strictes, cent à mille fois plus nombreuses que les bactéries anaérobies facultatives et les bactéries aérobies (Sekirov et al. 2010). Les bactéries anaérobies strictes étant difficile à cultiver, c'est l'avènement des méthodes de séquençage à haut-débit qui ont en premier lieu permis l'exploration du microbiote intestinal. Celui-ci est dominé par deux phyla principaux : les *Bacteroidetes* et les *Firmicutes* (**Figure 12**) (Arumugam et al. 2011, Eckburg et al. 2005). D'autres phyla, tels que les *Proteobacteria*, les *Verrucomicrobia*, les *Actinobacteria*, les *Fusobacteria* et les *Cyanobacteria* sont également présents, mais dans des proportions bien plus faibles.

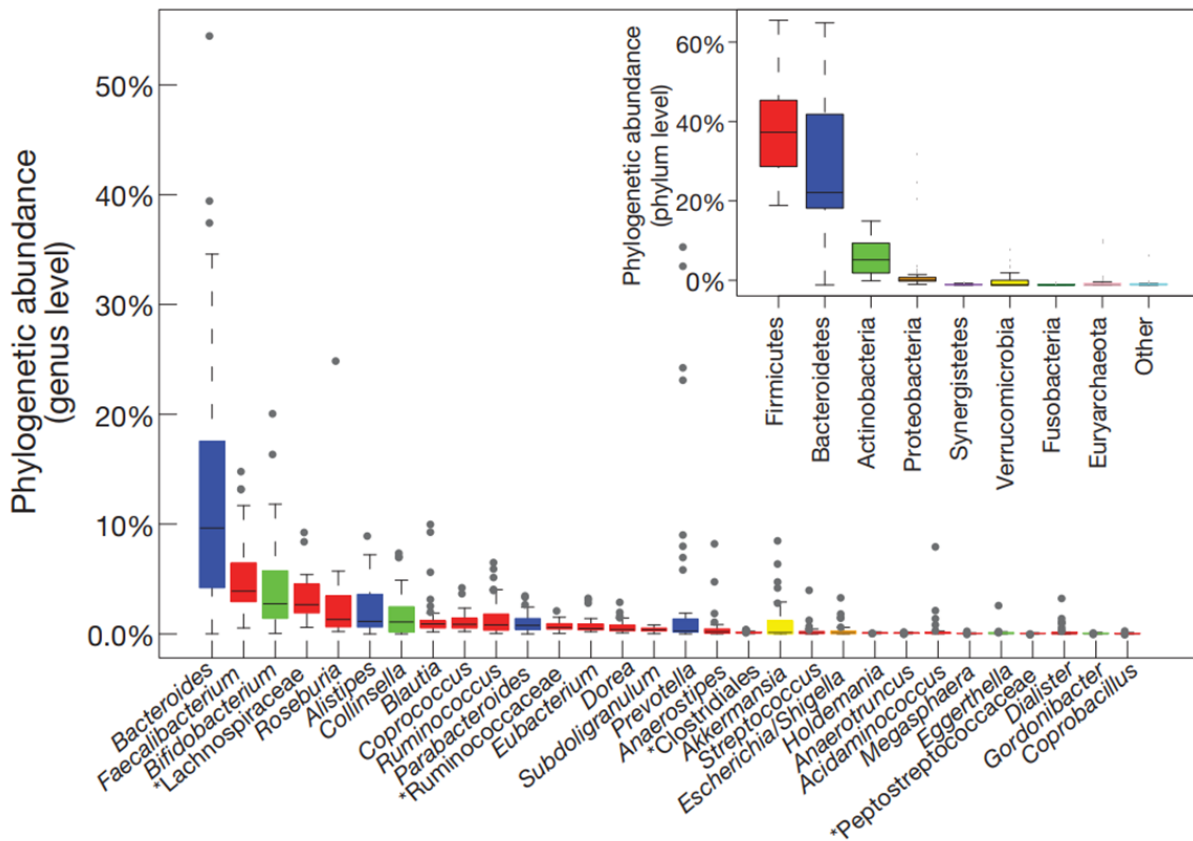


Figure 12. Composition du microbiote intestinal humain. Tiré de (Arumugam et al. 2011).

Proportion des trente genres les plus représentés du microbiote intestinal humain, par abondance de reads.

S'il est estimé que le microbiote intestinal humain collectif est composé de 1 000 à 1 150 espèces bactériennes différentes (Qin et al. 2010), il est admis que celui-ci ne contient qu'environ

100 à 200 espèces bactériennes différentes au sein d'un individu (Faith et al. 2013, Qin et al. 2010). Leur concentration, tout comme leur diversité, s'accroît du début à la fin du tractus digestif (**Figure 13A**). Ainsi, 10^1 à 10^3 bactéries par gramme de contenu sont retrouvées dans l'estomac et le duodénum, pour progresser graduellement jusqu'à 10^{11} - 10^{12} bactéries par gramme de contenu dans le colon (O'Hara and Shanahan 2006). De plus, la composition microbienne varie entre ces différentes portions du tractus gastro-intestinal. Ainsi, les bactéries appartenant aux *Bacteroidetes* et les *Lachnospiraceae* (*Firmicutes*) sont les plus abondantes dans le colon, tandis que l'intestin grêle est plutôt enrichi en *Firmicutes* et en *Actinobacteria* (Frank et al. 2007).

En plus de l'hétérogénéité longitudinale affichée par le microbiote intestinal, une grande variation latitudinale est également observée dans la composition du microbiote (**Figure 13B**). Ceci découle de la séparation de l'épithélium intestinal avec la lumière par une épaisse couche de mucus. Ainsi, nombreuses sont les espèces bactériennes présentes dans la lumière intestinale n'ayant pas accès au mucus et à la paroi épithéliale. Si les genres *Bacteroides*, *Bifidobacterium*, *Streptococcus*, *Enterobacteriaceae*, *Enterococcus*, *Clostridium*, *Lactobacillus* et *Ruminococcus* ont tous été trouvés dans les fèces, seulement des *Clostridium*, *Lactobacillus* et *Enterococcus* ont été détectés dans la couche externe de mucus et au niveau des parois épithéliales de l'intestin grêle (Swidsinski et al. 2005). Le microbiote associé au mucus du colon et du rectum est quant à lui majorité composé de *Bacteroides* et d'*Enterobacteriaceae*.

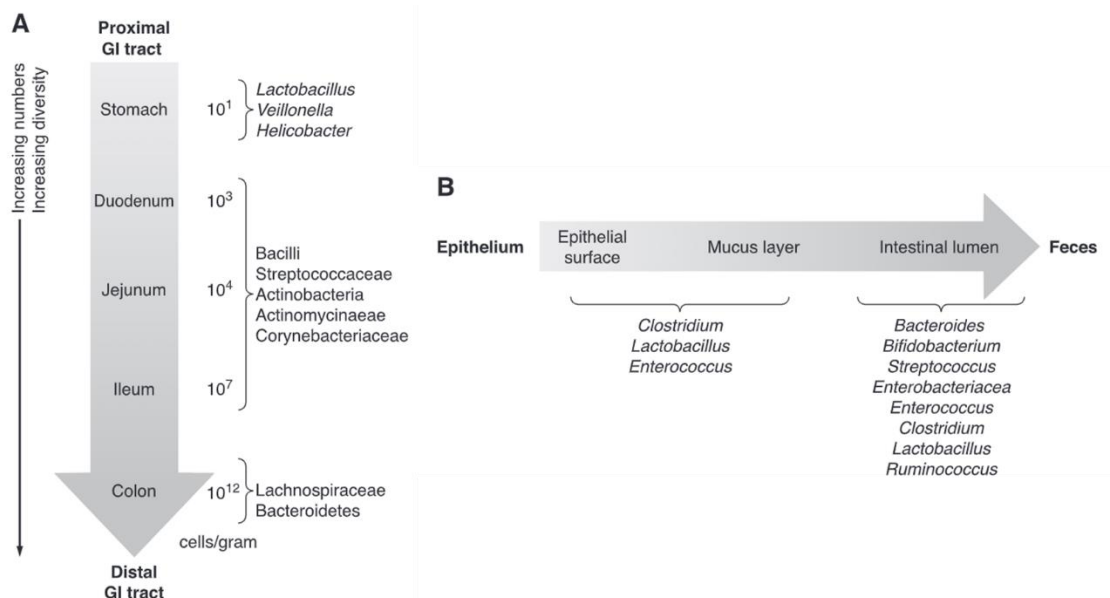


Figure 13. Diversité bactérienne au sein du tractus gastro intestinal. Adapté de (Sekirov et al. 2010).

(A) Diversité bactérienne tout au long du tractus gastro-intestinal. (B) Répartition transversale de la diversité bactérienne dans l'intestin.

3.2. Facteurs impactant la diversité du microbiote intestinal

In utero, le tractus gastro-intestinal est stérile. C'est au cours de l'accouchement que la première colonisation de celui-ci par les bactéries a lieu (Salazar et al. 2014). Le mode de délivrance de l'enfant lors de l'accouchement est l'un des premiers facteurs déterminant la composition du microbiote intestinal pour les premières semaines de vie (**Figure 14**). Si l'accouchement se déroule par voie basse, l'enfant acquerra un microbiote similaire au microbiote vaginal de la mère, tandis qu'en cas de césarienne, c'est d'abord un microbiote similaire au microbiote de la peau de la mère qui s'installera (Clemente et al. 2012, Jakobsson et al. 2014). Le régime alimentaire du nouveau-né impacte également sur la composition de son microbiote. Les bactéries aérobies sont ainsi plus communément retrouvées dans les selles d'enfants nourris au sein que dans celles de ceux qui ne le sont pas (Cresci and Bawden 2015, Le Huerou-Luron, Blat, and Boudry 2010, Lee et al. 2015, Stark and Lee 1982). Ces premières bactéries sont généralement aéro-tolérantes, dans la mesure où l'intestin contient initialement du dioxygène ; avant d'être progressivement remplacées par les bactéries anaérobies communes du microbiote intestinal adulte (Palmer et al. 2007). Le passage à l'alimentation solide marque également un tournant majeur de la composition du microbiote intestinal, au sein duquel les *Proteobactia* et les *Actinobacteria* laissent place aux *Bacteroidetes* et aux *Firmicutes* (Koenig et al. 2011, Ottman et al. 2012). À partir de l'âge de trois ans, le microbiote intestinal ressemble déjà à la composition d'un microbiote adulte et ne variera plus aussi rapidement au cours du vieillissement (Claesson et al. 2011). Plus tard, le déclin des fonctions salivaires, de la dentition, de la digestion et du transit intestinal moduleront également la composition du microbiote (Lovat 1996). Des différences notables existent entre le microbiote des personnes âgées et des jeunes adultes, comme une prédominance des *Bacteroidetes* chez les personnes âgées, contre une prédominance des *Firmicutes* chez les jeunes adultes. Une diminution de la proportion des *Bifidobacteria*, des *Bacteroides* et des *Clostridium* a également été relevée chez les personnes âgées (van Tongeren et al. 2005, Zwieler et al. 2009). D'autres facteurs, comme l'alimentation, la situation géographique, le stress, la pratique d'activités physiques régulières, la prise de médicaments et en particulier d'antibiotiques, la génétique de l'hôte et la charge pondérale influent également sur la composition du microbiote intestinal d'un individu (Pour une revue voir (Cresci and Bawden 2015)). De plus, le microbiote intestinal est intimement lié à la survenue et au développement de certaines pathologies, un sujet qui sera abordé plus loin dans ce manuscrit.

Une étude a cependant estimé que l'ensemble de ces facteurs n'explique cependant que 16,4% de la variation de la composition du microbiote au sein d'une population d'adultes en bonne santé vivant dans le même pays (Falony et al. 2016). Il existe donc d'autres facteurs mal compris qui

régissent la composition du microbiote. Par ailleurs, il est communément admis que le microbiote intestinal présente une grande résilience et que, sitôt l'élément perturbateur disparu, il retrouve son état initial en l'espace de quelques semaines (Claesson et al. 2012, De Filippo et al. 2010, Lambert et al. 2015, Qin et al. 2010, Reyes et al. 2010, Walker et al. 2011). Certains individus semblent cependant posséder une composition plus instable du microbiote (Brussow 2016).

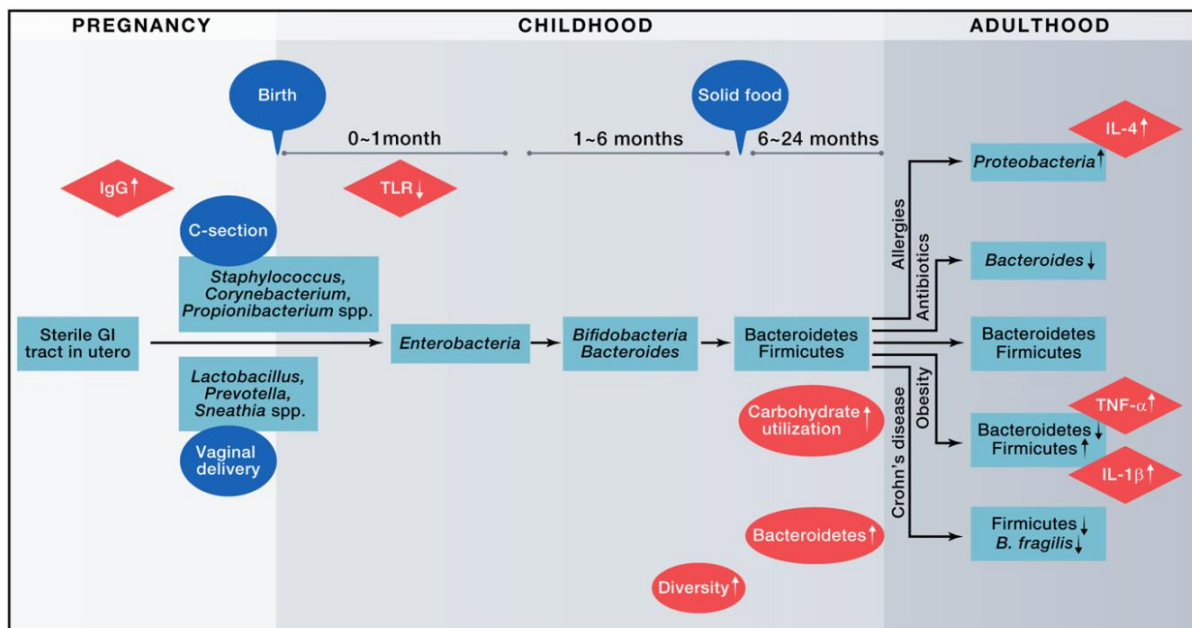


Figure 14. Développement du microbiote intestinal humain. Tiré de (Clemente et al. 2012).

3.3. Les fonctions du microbiote intestinal

Le microbiote intestinal joue un rôle important dans l'absorption de nutriments. Il permet notamment la fermentation de substrats non digérés, comme les fibres complexes qu'il transforme en composés assimilables par l'organisme (Wong et al. 2006). Le processus de fermentation permet également la production d'acides gras à chaînes courtes, comme l'acétate, le propionate et le butyrate, qui jouent de nombreux rôles au sein du corps humain. Le butyrate est par exemple la source majoritaire d'énergie pour les colonocytes, active la gluconéogenèse intestinale et induit l'apoptose des cellules cancéreuses du colon (De Vadder et al. 2014). Transporté vers le foie, le propionate régule également la gluconéogenèse et la satiété (De Vadder et al. 2014). L'acétate, qui est un métabolite essentiel à la croissance de certaines bactéries, est transporté vers les tissus périphériques, où il est utilisé dans le métabolisme du cholestérol et de la lipogenèse, ainsi que dans la régulation de l'appétit (Frost et al. 2014). Le microbiote intestinal intervient également dans l'une

des voies métaboliques du tryptophane, un acide aminé essentiel, en permettant la transformation de celui-ci en dérivés de l'indole, dont certains agissent en tant que ligands des récepteurs de l'aryle hydrocarboné (AhR). L'activation des AhR permet notamment la production de GLP-1 et IL-22, qui contribuent à l'intégrité de la barrière épithéliale intestinale (Agus, Clement, and Sokol 2021, Agus, Planchais, and Sokol 2018).

De nombreuses autres fonctions majeures du microbiote intestinal sont liées à la santé de l'hôte. Ainsi, la colonisation bactérienne par des bactéries commensales dans le mucus de l'intestin protège l'hôte de la colonisation par d'éventuels pathogènes, en les empêchant d'atteindre le tissu épithélial (Hirano et al. 2003, Kim et al. 2008, Lee et al. 2003, Yoon, Lee, and Yoon 2014). De plus, l'utilisation du butyrate par les colonocytes joue un rôle favorable dans la maturation et le développement de la barrière intestinale, augmentant ainsi sa capacité à empêcher les bactéries de pénétrer l'organisme de l'hôte (Soret et al. 2010). En consommant les nutriments présents dans l'intestin, les bactéries commensales empêchent également l'implantation de potentiels pathogènes au sein du tube digestif (Hooper et al. 1999). Certaines sont mêmes capables de synthétiser des composés antimicrobiens, tels que les bactériocines, pour conserver leur niche écologique et par conséquent réduire les risques d'implantations de nouvelles espèces potentiellement pathogènes dans l'intestin (Crost et al. 2011, Hatzioanou et al. 2013). D'autres espèces bactériennes commensales sont également importantes en santé humaine, comme *Faecalibacterium prausnitzii*, qui produit le peptide MAM, un peptide anti-inflammatoire jouant un rôle important dans l'équilibre entre le microbiote intestinal et la réponse immunitaire de l'hôte (Lenoir et al. 2020, Martín et al. 2014, Quevrain et al. 2016). *Faecalibacterium prausnitzii*, intervient également dans l'équilibre entre le microbiote intestinal et le système immunitaire de l'hôte par le biais de la synthèse de butyrate, qui intervient dans la nutrition des colonocytes. Ce composé est également produit par les bactéries du genre *Roseburia* (Duncan et al. 2002, Tamanai-Shacoori et al. 2017).

3.4. Les virus du microbiote intestinal

Le virome intestinal, bien qu'il contienne des virus eucaryotes, est majoritairement composé de phages (Shkoporov et al. 2019). En dépit d'une grande variabilité inter-individus concernant la diversité des phages intestinaux, les communautés phagiques, également appelés phageomes, sont très stables dans le temps au sein d'individus uniques (95% de la diversité virale est conservée dans le temps) (**Figure 15**) (Minot et al. 2013, Reyes et al. 2010, Shkoporov et al. 2019).

Comme pour les bactéries du microbiote intestinal, le phageome intestinal se développe après la naissance et connaît d'importantes modifications de sa diversité au cours des trois premières années de vie (Breitbart et al. 2008). L'origine du phageome pourrait à la fois être maternel, environnemental et alimentaire (Breitbart et al. 2008), et découler de l'induction de prophages des bactéries du microbiote (Breitbart et al. 2003). Jusqu'à l'âge de deux ans et demi, ce sont les caudophages qui dominent le phageome intestinal, avant d'être remplacés par les *Microviridae* (Lim et al. 2015). À l'âge adulte, les caudophages s'imposent à nouveau comme étant les phages dominants du phageome, en particulier par la présence des Crassphages, des podophages infectant plusieurs espèces de *Bacteroides*, récemment décrits (Shkoporov, Khokhlova, et al. 2018). Les *Microviridae* restent néanmoins très prévalents au sein du phageome de l'adulte (**Figure 15**) (Shkoporov 2019, Kim 2011, (Manrique et al. 2016). Des travaux récents suggèrent cependant une moindre prévalence des Crassphages chez les enfants âgés d'un an par rapport aux adultes (Shah et al. 2021).

À l'instar des bactéries, de nombreux facteurs déterminent la composition de la population phagique dans l'intestin, comme l'alimentation (Reyes et al. 2015, Singh et al. 2017), qui pourrait par ailleurs avoir un effet sur l'induction de prophages du microbiote intestinal (Boling et al. 2020, Oh et al. 2019).

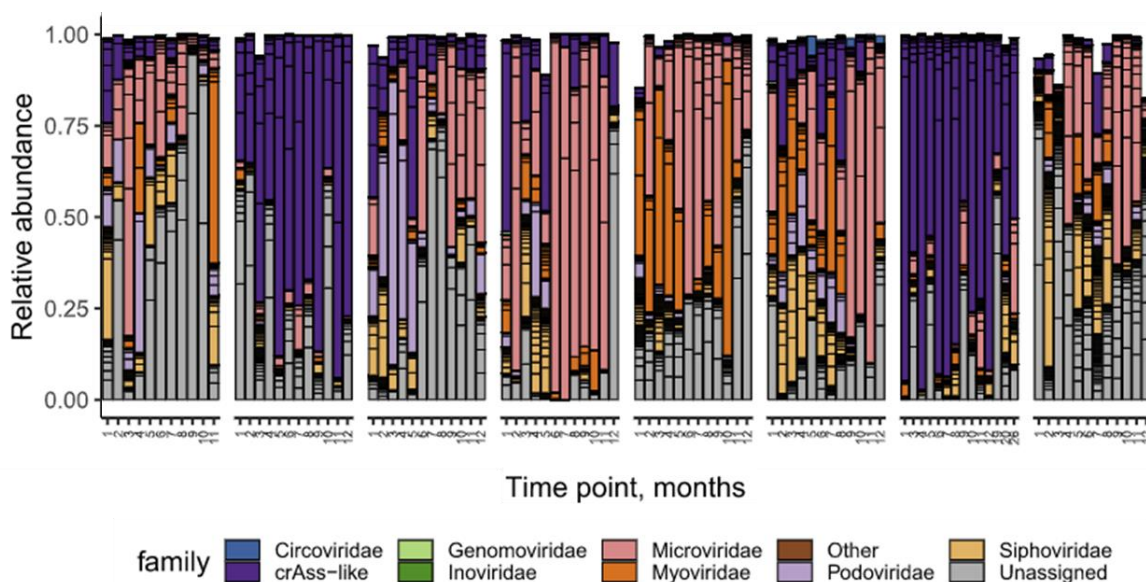


Figure 15. Composition des viromes de 7 individus sains. Adapté de (Shkoporov et al. 2019).

Composition taxonomique des viromes au niveau des familles chez sept sujets sains par points de temps mensuels, en abondance relative de contigs.

Chez l'adulte, le phageome semble dominé par des phages tempérés (Reyes et al. 2010, Silveira and Rohwer 2016). Cependant, les analyses réalisées sur le phageome ne permettent pas d'obtenir des valeurs précises et la diversité des phages tempérés pourrait masquer les phages virulents sous-dominants dans le microbiote (Sutton and Hill 2019).

Des travaux récents d'Andrey Shkoporov et co-auteurs (Shkoporov et al. 2021) ont par ailleurs caractérisé la diversité virale tout au long du tractus gastro-intestinal de primates (**Figure 16**). Leurs résultats suggèrent, en accord avec leurs précédents travaux chez l'humain (Shkoporov, Ryan, et al. 2018), l'importance de la quantité de *Microviridae*, de Crassphages et de caudophages au sein du microbiote intestinal de manière générale et permettent l'estimation, par l'ajout du phage Q33 en marqueur de concentration d'un total d'environ 1×10^9 copies de génomes viraux par gramme de selle. Cet ajout a également permis d'estimer un dénombrement absolu des phages, qui semblent être 10 à 100 fois plus abondants dans le lumen intestinal qu'au sein de la muqueuse intestinale, en accord avec de précédents travaux réalisés chez la souris (Lourenco et al. 2020). Ces résultats sont cependant opposés à l'hypothèse, non démontrée mais fréquemment retrouvée dans la littérature, selon laquelle l'affinité des phages avec les immunoglobulines favoriserait une concentration élevée de ces derniers dans le mucus intestinal (Barr et al. 2013, Silveira and Rohwer 2016). Enfin, le phageome intestinal s'avère être encore plus variable d'un individu à l'autre que le bactériome intestinal, bien que celui-ci soit très stable dans le temps (Garmaeva et al. 2021, Shkoporov et al. 2019).

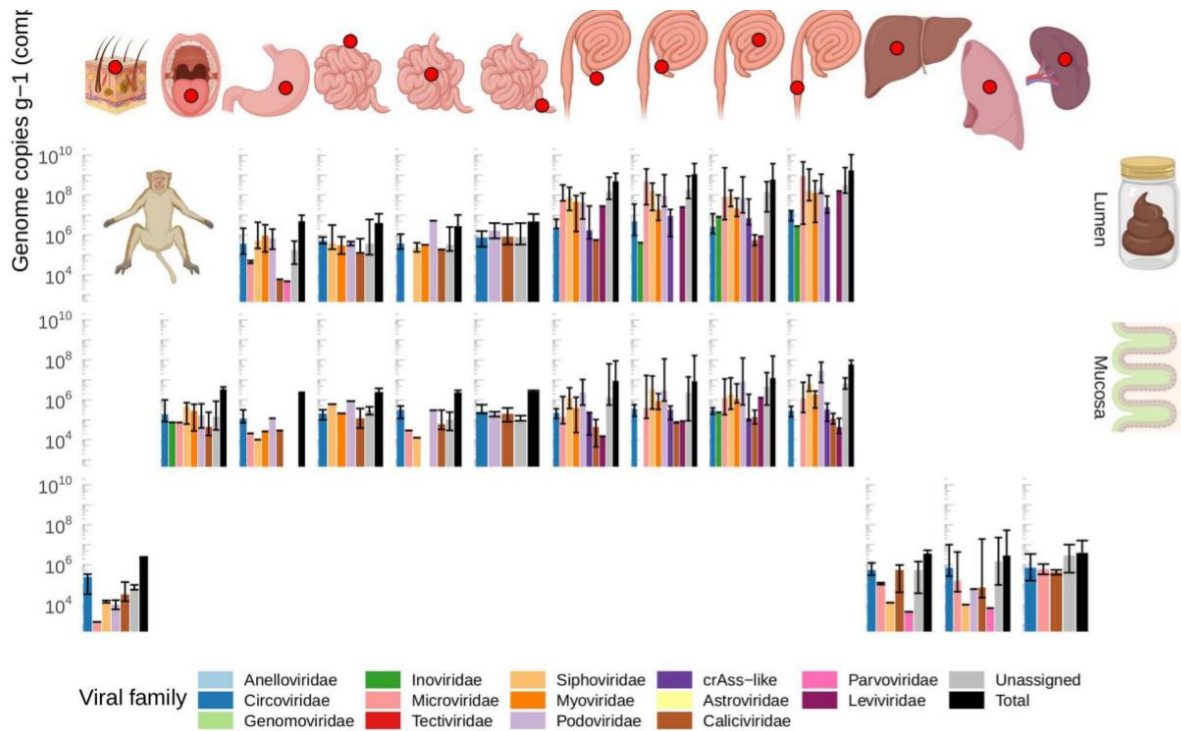


Figure 16. Quantifications absolues de génomes viraux par gramme, tout au long du tractus gastro-intestinal de macaques. Tiré de (Shkoporov et al. 2021).

La quantité absolue de génomes viraux a été calculée suite à l'inoculation en quantités connues du phage Q33. Seuls les génomes complétés à plus de 50% ont été pris en compte dans le calcul des quantités absolues.

De rares études font état de dénombrements absolus des VLP (Virus Like Particles, un terme prudent pour désigner des potentiels virus) au sein du microbiote intestinal, par microscopie par épifluorescence. Ainsi, Lesley Hoyles et co-auteurs (Hoyles et al. 2014) ont mesuré une moyenne de 3×10^9 VLP/g de selle chez des individus sains, alors qu'une étude de Patricia Lepage et co- auteurs (Lepage et al. 2008) a déterminé la présence de $1,2 \times 10^8$ VLP par biopsie muqueuse de côlons d'individus sains.

Enfin, le microbiote total de selles d'individus sains semble contenir une moyenne de 5,8% de reads viraux (Arumugam et al. 2011). Ceux-ci comprennent à la fois les virus eucaryotes et les phages, qu'ils soient sous forme de virions ou à l'état de prophages.

3.5. Les VE du microbiote intestinal

Les vésicules extracellulaires représentent également une population de nanoparticules du microbiote intestinal non négligeable. Une étude de Nader Kameli et co-auteurs (Kameli et al. 2021) les a dénombrées à l'aide du qNano (basé sur la TRPS), à une concentration d'environ 2×10^9 par gramme de selle, chez des individus sains, après purification par chromatographie à exclusion de taille. Étant donné la complexité de l'écosystème intestinal, des mécanismes de régulation impliquant des récepteurs immunitaires, des voies de signalisation, des protéines régulatrices et des miARN sont nécessaires pour assurer la symbiose et éviter des réponses inflammatoires exacerbées qui pourraient conduire à des états pathologiques. Dans ce scénario, la communication intercellulaire est cruciale pour orchestrer des réponses équilibrées pour préserver l'homéostasie intestinale. De nombreux travaux semblent supporter l'hypothèse que les VE du microbiote intestinal, qu'elles soient d'origine procaryote ou eucaryote, interviennent dans cette communication à grande échelle, en agissant par exemple sur l'intégrité de la barrière épithéliale, la régulation de l'immunité intestinale et la cicatrisation des plaies mucoales (pour une revue, voir (Diaz-Garrido et al. 2021)). De plus, plusieurs miARN de VE eucaryotes interagissent avec le microbiote intestinal en promouvant la croissance de certaines bactéries comme *Fusobacterium nucleatum* ou *Escherichia coli* (Liu et al. 2016). D'autres miARN ont par ailleurs été positivement corrélés à l'abondance de certains phyla bactériens (Moloney et al. 2019). Enfin, l'état inflammatoire de l'intestin impacte la nature des miRNA présents dans celui-ci (Viennois et al. 2019).

Dans les selles humaines, l'abondance relative des VE bactériennes a été estimée en ciblant, par qPCR, des gènes spécifiques dans la régions 16S des grands phyla (Kameli et al. 2021). Si cette méthode présente certaines limites, notamment du fait que seules les VE contenant des acides nucléiques entrent dans ces calculs d'abondance, leurs résultats, comparés aux abondances relatives des bactéries dans le microbiote intestinal, suggèrent néanmoins que certains phyla sont plus prompts à produire des VE que d'autres (**Figure 17**). C'est en particulier le cas des *Bacteroidetes* et des α -*Proteobacteria*. À l'inverse, les *Actinobacteria* semblent produire peu de VE.

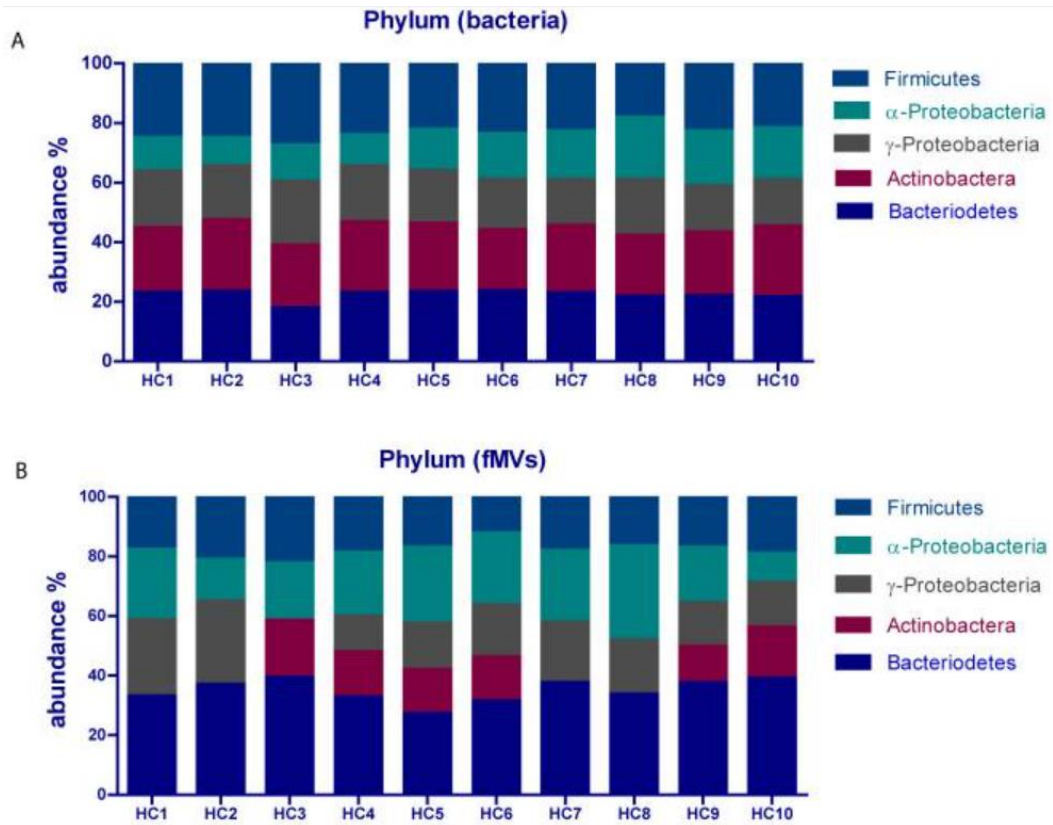


Figure 17. Détermination de l'origine bactérienne de vésicules fécales dans chez 10 individus sains basée sur la métagénomique des vésicules. Tiré de (Kameli et al. 2021).

(A) Abondance relative des bactéries appartenant aux phyla les plus communs. (B) Abondance relative des VE bactériennes en fonction de leur origine bactérienne.

4. Les dysbioses intestinales associées à la maladie de Crohn

4.1. Modifications du microbiote intestinal dans la maladie de Crohn

Bien que l'étiologie de la maladie de Crohn soit inconnue, la présence de dysbioses associées à une diminution de la diversité bactérienne a été largement documentée dans celle-ci (Dicksved et al. 2008, Spor, Koren, and Ley 2011). Ces dysbioses se caractérisent notamment par une importante diminution de l'abondance des *Firmicutes*, dont *Faecalibacterium* et *Roseburia*, ainsi qu'une augmentation des *Bacteroidetes* et plus particulièrement du genre *Bacteroides* (**Figure 18**). Une proportion accrue de bactéries anaérobies facultatives (surtout des *Enterobacteria*) est également observée chez les malades (Sokol et al. 2006).

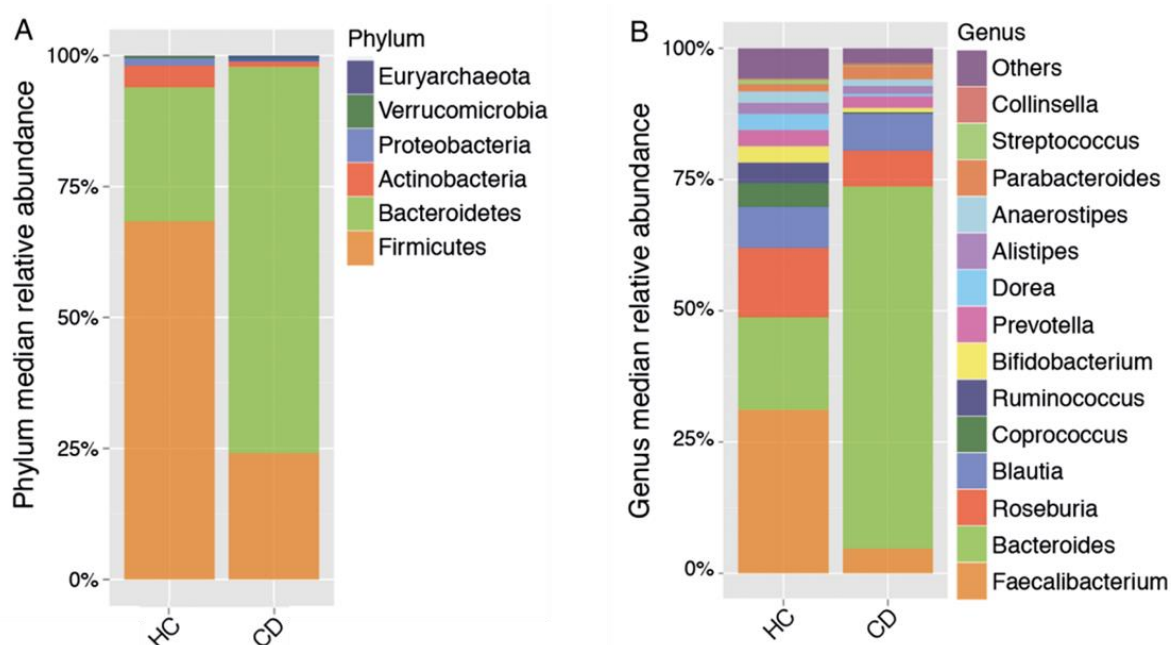


Figure 18. Composition du microbiote fécal par abondance relative des reads. Adapté de (Sabino et al. 2016).

Composition du microbiote fécal d'individus sains (HC) et de patients atteints de la maladie de Crohn (CD). (A) Abondance relative des reads au niveau des phyla bactériens. (B) Abondance relative des reads au niveau des genres bactériens. Les 14 genres les plus abondants sont représentés, les autres se trouvent dans la catégorie « others ».

De plus, la pathogénèse des MICI pourrait être liée à des altérations de la voie métabolique de la transformation du tryptophane en dérivés de l'indole. Une production diminuée de ligands aux récepteurs d'aryl hydrocarbonés (AhR) a été observée chez des patients atteints de la maladie de

Crohn, celle-ci étant due à des facteurs génétiques. Ainsi, il a été démontré que le microbiote de souris déficientes au niveau du gène CARD9 (pour caspase recruitment domain containing protein 9) est incapable de transformer le tryptophane en métabolites ligands des AhR. Cela entraîne ainsi une diminution de la libération d'IL-22, conduisant à une plus grande sensibilité de ces souris à l'inflammation intestinale. Cependant, après traitement par des agonistes des AhR ou après inoculation de souches capables de métaboliser le tryptophane en ligands des AhR, l'intensité de l'inflammation intestinale est réduite (Lamas et al. 2016).

Une récente preuve de concept réalisée chez le rat a par ailleurs mis en évidence l'existence d'états alternatifs du microbiote intestinal chez le rat. Dans certains cas où la composition du microbiote intestinal et l'inflammation du côlon diffèrent suffisamment de celui des individus sains, le modèle suggère qu'un retour à un état basal serait difficilement réalisable, à moins de traiter conjointement l'inflammation et de rétablir un microbiote plus proche de celui de l'état basal. Ce modèle met également en évidence des états de dysbioses intermédiaires, dont l'identification pourrait permettre aux individus concernés de bénéficier d'un traitement plus adapté pour espérer voir leur microbiote revenir à un état basal stable (Van de Guchte et al. 2020).

4.1.1. Espèces bactériennes dont l'abondance est modifiée

Dans les échantillons des patients atteints de la maladie de Crohn, les populations d'*E. coli* sont plus nombreuses et sont plus particulièrement représentées par des *E. coli* adhérents-invasifs (AIEC) (Martinez-Medina et al. 2009). En raison de leur forte prévalence parmi les *E. coli* des patients (51,9 % chez les patients atteints de Crohn, 16,7 % chez les individus sains), ces AIEC sont soupçonnées de participer à la maladie de Crohn en entretenant l'inflammation intestinale grâce à leurs capacités invasives et via l'expression de cytokines pro-inflammatoires (Boudeau et al. 1999, Carvalho et al. 2009, Glasser et al. 2001). Les populations d'*E. coli* ne sont cependant pas les seules à être drastiquement bouleversées chez les patients atteints de la maladie de Crohn, à l'instar de *Faecalibacterium prausnitzii*.

Le genre *Faecalibacterium* est l'un des plus répandus du microbiote intestinal humain (Arumugam et al. 2011, Qin et al. 2010, Sabino et al. 2016). Il est composé d'une unique espèce, *Faecalibacterium prausnitzii*, qui est extrêmement sensible à l'oxygène et qui produit différents acides gras à chaîne courte, dont le butyrate. Elle semble ainsi jouer un rôle important dans l'équilibre entre le microbiote intestinal et la réponse immunitaire, en raison de ses propriétés anti-inflammatoires. L'absence ou la diminution de la concentration de *F. prausnitzii* chez les patients atteints de MICI

pourrait expliquer une partie du cercle vicieux observé dans ces pathologies (Machiels et al. 2014, Sokol et al. 2009). En effet, la disparition ou la forte diminution des bactéries productrices de butyrate pourrait être à l'origine d'un contexte pro-inflammatoire dans l'intestin dans les MICI. De plus, cet environnement inflammatoire ne favorise pas le retour des bactéries productrices de butyrate, car elles sont très sensibles à l'oxygène et la production élevée de radicaux libres produits pendant l'inflammation ne permet pas leur croissance. L'utilisation de *F. prausnitzii* comme probiotique pourrait cependant constituer un potentiel thérapeutique, en améliorant les chances de rémission après une crise inflammatoire (Martín et al. 2014).

Comme *F. prausnitzii*, le genre *Roseburia* est très appauvri dans le microbiote intestinal des patients atteints de la maladie de Crohn (Machiels et al. 2014, Willing et al. 2010). Le genre *Roseburia* appartient à la famille des *Lachnospiraceae* (Reeves et al. 2012) et est le cinquième genre le plus représenté dans la flore intestinale humaine (Arumugam et al. 2011, Qin et al. 2010). C'est également un producteur d'acides gras à chaînes courtes et plus particulièrement de butyrate (Tamanai-Shacoori et al. 2017), en faisant un genre bactérien clé dans l'équilibre du microbiote intestinal, qui subit les mêmes effets de cercle vicieux de l'inflammation que *F. prausnitzii*. De plus, des études ont montré la capacité de *Roseburia faecis* à produire des composés antimicrobiens contre certaines espèces du genre *Bacillus*, participant probablement à la lutte du microbiote contre de potentiels pathogènes (Hatzioanou et al. 2013).

4.1.2. Le microbiote fongique intestinal dans les MICI

Le microbiote fongique a peu été exploré jusqu'à présent. Cependant, une étude de Harry Sokol et co-auteurs (Sokol et al. 2017) montre une augmentation du ratio Basidiomycota/Ascomycota chez les patients atteints de MICI, une diminution de la proportion de *Saccharomyces cerevisiae* et une augmentation de la proportion de *Candida albicans* par rapport aux individus sains. Il existe en effet une dysbiose du microbiote fongique dans les MICI, caractérisée par des altérations de sa biodiversité et de sa composition. Cette étude suggère que les champignons pourraient également jouer un rôle dans la pathogenèse des MICI (Sokol et al. 2017).

Pour résumer, certaines bactéries pathogènes comme *E. coli* se développent lors des inflammations intestinales ; tandis que d'autres populations bactériennes, comme *Faecalibacterium*

prausnitzii et *Roseburia spp*, sont drastiquement diminuées dans les dysbioses subies par les patients de la maladie de Crohn. L'ensemble de ces changements est favorable à l'installation ou à l'entretien d'un contexte dysbiotique pro-inflammatoire au sein du microbiote intestinal humain.

4.2. Altérations de la fraction virale du microbiote

Bien que l'on ne connaisse pas l'origine des dysbioses observées au sein du microbiote intestinal de patients atteints de la maladie de Crohn, plusieurs études ont montré que les populations phagiques sont elles aussi altérées au cours de la maladie. Le travail de Patricia Lepage et co-auteurs (Lepage et al. 2008) est l'un des premiers à ce sujet, en montrant une augmentation d'un facteur 20 du nombre de VLPs observées par épifluorescence dans des biopsies de zones inflammatoires de l'intestin de patients, en comparaison à des individus sains.

Contrairement au travail présenté ci-dessus, Adam Clooney et co-auteurs (Clooney et al. 2019) n'observent pas de différence d'abondance des phages entre les viromes de selles d'individus sains et de patients atteints de la maladie de Crohn (après quantification par inoculation d'un phage de concentration connue). En revanche, ils rapportent une plus forte proportion de phages tempérés chez les malades. Leurs résultats suggèrent une induction des prophages présents dans le microbiote des patients. Ce passage de la lysogénie vers la lyse pourrait expliquer l'augmentation de l'abondance et de la richesse relative des caudophages chez les patients et la diminution de l'abondance et de la richesse des *Microviridae* et des Crassphages, les phages tempérés étant plus souvent des caudophages (Clooney et al. 2019, Fernandes et al. 2019, Norman et al. 2015). Une tendance similaire a été observée par Guanxiang Liang et co-auteurs (Liang et al. 2020) dans des stades précoces de la pathologie et ce, sans différence de dénombrement des VLPs avec les contrôles sains. Une récente étude de métaprotéomique a par ailleurs révélé une corrélation positive entre la présence de protéines de phages dans le contenu intestinal et la maladie de Crohn. Au contraire, une anticorrélation entre les protéines de phages et les individus sains a été soulignée par ce travail, suggérant une nouvelle fois une augmentation du nombre de virions dans le microbiote intestinal au cours de la maladie (Zhang et al. 2018).

Au vu de ces résultats, il peut être supposé que ces altérations du virome sont en lien avec les modifications de populations bactériennes observées chez les patients. C'est dans ce sens que vont les résultats de Jeffrey Cornuault et co-auteurs (Cornuault et al. 2018), montrant une abondance

de certains phages tempérés de *Faecalibacterium prausnitzii* augmentée chez les patients. Leur induction chez ces bactéries productrices de butyrate pourrait entretenir le cercle vicieux de l'inflammation observé au cours de la maladie. De plus, une étude *in vitro* de Lasha Gogokhia et co-auteurs (Gogokhia et al. 2019) suggère que la reconnaissance d'ADN phagique par les cellules dendritiques pourrait stimuler la production d'interférons γ et ainsi exacerber le contexte inflammatoire intestinal.

4.3. Les rôles des VE dans la maladie de Crohn

En plus d'être produites par certaines bactéries et cellules eucaryotes à des fins de communication, lors de la mort bactérienne ou pour des raisons encore inconnues à ce jour, les VE peuvent également être sécrétées en cas de stress environnemental. L'inflammation étant un facteur de stress majeur caractérisant la maladie de Crohn, il est probable que la population totale de VE puisse avoir une importance non négligeable dans cette pathologie ou être affectée par la dysbiose ou l'inflammation. Dans des fractions purifiées de VE du colon, une augmentation de l'abondance relative de protéines d'origine eucaryote a ainsi été observée chez des patients pédiatriques atteints de MICI, par rapport aux contrôles sains (Zhang et al. 2018). De plus, il a été rapporté que les exosomes isolés de patients atteints de MICI contiennent des taux plus élevés d'IL-6, IL-8, IL-10 et TNF- α par rapport à ceux contenus chez des individus sains et que les concentrations de ces molécules pro-inflammatoires sont positivement corrélées à la sévérité de la maladie chez les patients étudiés (Mitsuhashi et al. 2016). Une autre étude récente réalisée *in vitro* suggère par ailleurs une libération augmentée de TNF- α inflammatoires par les cellules THP-1 en présence de VE isolées de fèces de patients ; et une corrélation positive entre le nombre de VE issues d'*Actinobacteria*/*γ-Proteobacteria* et la production de TNF- α . Ainsi, une fraction du microbiote intestinal serait capable de maintenir un environnement pro-inflammatoire via la production de VE (Kameli et al. 2021).

À l'inverse, certaines VE du microbiote intestinal possèdent des propriétés anti-inflammatoires. C'est notamment le cas des vésicules de *F. prausnitzii*, dont l'activité anti-inflammatoire a été observée *in vitro* sur des cellules épithéliales d'adénocarcinomes, à la fois par la répression de cytokines pro-inflammatoires dont l'IL-6, le TNF- α et le TNF- β , ainsi que par l'induction de cytokines anti-inflammatoires dont l'IL-10, l'IL-1Ra et le TGF- β 2 (Jafari et al. 2019). Des effets similaires ont été observés *in vivo*, suite à l'administration orale d'OMVs de *E. coli* Nissle 1917 chez des souris atteintes de rectocolite hémorragique. De plus, dans cette même étude, une diminution

de la cyclooxygénase-2 et de la synthase d'oxyde nitrique, qui sont des enzymes impliquées lors de l'inflammation, a été observée (Fabrega et al. 2017).

Il est aujourd'hui connu que les VE du microbiote intestinal peuvent traverser la barrière épithéliale, interagir avec le système immunitaire de l'hôte et atteindre des sites distants de l'organisme humain via les circulations systémique et lymphatique. Cependant, les mécanismes de passages des VE au travers de la barrière intestinale restent aujourd'hui encore peu élucidés (pour une revue, voir (Diaz-Garrido, Badia, and Baldoma 2021)). Différentes voies de passage des VE sont actuellement considérées. Tout d'abord, les VE du microbiote intestinal pourraient passer dans le compartiment sanguin grâce à des phénomènes d'endocytose. Ce phénomène a notamment été démontré *in vitro*, avec des OMVs de *E. coli* Nissle 1917 et de *Bacteroides thetaiotaomicron* (Canas et al. 2016, Jones et al. 2020) et observé *in vivo* avec des OMVs de *B. thetaiotaomicron* (Jones et al. 2020). Le passage des VE intestinales au travers de la barrière épithéliale pourrait également s'expliquer par des phénomènes de transcytose via les cellules épithéliales, bien que ce phénomène n'ait été observé que dans le cadre de passage d'exosomes eucaryotes au travers de la barrière hémato-cérébrale, *in vitro* (Chen et al. 2016, Morad et al. 2019). Les VE intestinales peuvent également franchir la barrière épithéliale par un simple passage entre les cellules qui la constituent, dont la perméabilité des jonctions est diminuée lors de l'inflammation. En effet, une étude *in vitro* a mis en avant la capacité des OMVs de *Campylobacter jejuni* à cliver, par le biais d'une activité protéolytique, la E-cadhérine et l'occludine, qui sont acteurs des jonctions cellule-cellule de la barrière épithéliale intestinale. Cette atteinte à l'intégrité épithéliale permet par ailleurs une meilleure implantation des populations bactériennes à proximité des tissus intestinaux (Elmi et al. 2016).

On peut ainsi supposer que la concentration en VE dans le microbiote intestinal soit augmentée dans la maladie de Crohn. Il est également connu que l'infection phagique provoque la production à la fois active et passive de VE par certaines bactéries, que ce soit par bourgeonnement ou par explosion lors de la lyse bactérienne (Mandal et al. 2021). Enfin, Yue Liu et co-auteurs (Liu et al. 2022) proposent un modèle selon lequel un phage de *Lactococcus lactis* serait capable de quitter son hôte en s'insérant dans des vésicules produites par celui-ci, phénomène qui pourrait contribuer à l'augmentation de la concentration en VE dans le microbiote des patients.

En conclusion, les liens entre les nanoparticules du microbiote intestinal et la maladie de Crohn restent peu connus à ce jour. Cette thèse propose un travail exploratoire de dénombrement

de ces objets au sein de selles de patients en phase de poussée ou de rémission et de selles d'individus sains.

Résultats

Avant-propos

C'est autour de la question de du dénombrement des phages et des VE dans les selles au cours de la maladie de Crohn que gravite l'entièreté de cette thèse.

Dans un premier temps, et c'est le sujet du premier chapitre de résultats présentés de ce manuscrit, il a fallu évaluer les avantages et les limites du Videodrop, développé par Myriade, pour la détection et la quantification de phages. Il en va de même pour les VE, qui se sont révélées, assez rapidement au cours de ma thèse, comme étant très abondantes dans les selles. Ce premier chapitre s'attache également à la recherche d'une méthode de discrimination de ces deux populations de nanoparticules au sein de mélanges biologiques.

Dans le second chapitre, nous nous intéressons au dénombrement conjoint de bactéries, de phages et de VE au sein de selles d'individus sains et de patients issus d'une cohorte MICI. Les adaptations et mises au point techniques auxquelles que j'ai dû réaliser tout au long de ma thèse y sont également développées. Les hypothèses concernant ces résultats ainsi que les pistes qui pourraient être suivies sont ensuite discutées.

Enfin, en annexe, je présente d'autres projets parallèles à ma thèse, auxquels j'ai eu la chance de participer. Il s'agit notamment de la partie bio-informatique d'une étude sur les gènes d'antibiorésistance dans les viromes fécaux de lapins d'élevage, de ma participation au projet d'annotation de viromes dans une cohorte de plus de 600 enfants et de l'écriture d'une revue sur les phages intestinaux.

1. Détermination des limites et des performances du Videodrop

Malgré leurs potentiels rôles non négligeables dans les MICI, les VLPs n'ont presque jamais été dénombrées dans des selles. Les rares travaux de dénombrement réalisés par microscopie par épifluorescence manquent de contrôles, notamment concernant les seuils de sensibilité de cette technique et la validation de la nature virale et de la non-agrégation des VLPs détectées. De plus, nous ne savons pas dans quelle mesure les virus présents dans les selles humaines, dont la taille des capsides ainsi que la taille et la nature de leurs génomes peuvent être variables, sont détectables via cette méthode. Nous avons ainsi souhaité mettre au point un protocole de quantification des VLPs dans les selles d'individus sains et de patients atteints de la maladie de Crohn. Plusieurs approches concernant la détection des virus dans les selles ont été testées et leurs limites ont été investiguées.

Les VE n'ont, quant à elles, jamais été quantifiées dans les selles humaines. La recherche d'un protocole de dénombrement de ces dernières a également été conduite au cours de cette thèse afin de comparer les quantités de VE détectées dans les selles de patients et de contrôles sains.

Afin d'énumérer les phages et les VE dans les selles humaines, l'un des premiers objectifs de ce projet a été de produire des échantillons purifiés de représentants caractéristiques des populations attendues dans le microbiote intestinal, pour qu'ils servent de contrôles à ce que l'on détecte dans les selles. Ces standards ont été produits selon les méthodes décrites ci-dessous.

1.1. Les gradients d'iodixanol

Tout au long de ma thèse, nous avons eu recours à des gradients d'iodixanol. Séparant les nanoparticules sur base de leur densité, ceux-ci ont pour avantage de permettre l'obtention de fractions purifiées de phages, comme de VE. L'ensemble des gradients d'iodixanol ont été réalisés selon le protocole décrit dans l'article 1.

Si la densité des virus ($\sim 1,20$ g/mL) permet de les concentrer en une seule fraction (la fraction 4, qui, selon les préparations, peut également contenir des VE et des flagelles), ce n'est pas le cas pour les VE dont les densités s'étalent sur plusieurs fractions du gradient (**Figure 19** Erreur ! Source du renvoi introuvable.). Ainsi, c'est dans la fraction 2 que l'on retrouve les VE les plus purifiées, mais celles-ci peuvent se retrouver jusqu'à la fraction 4, qui peut également contenir des virus et des flagelles.

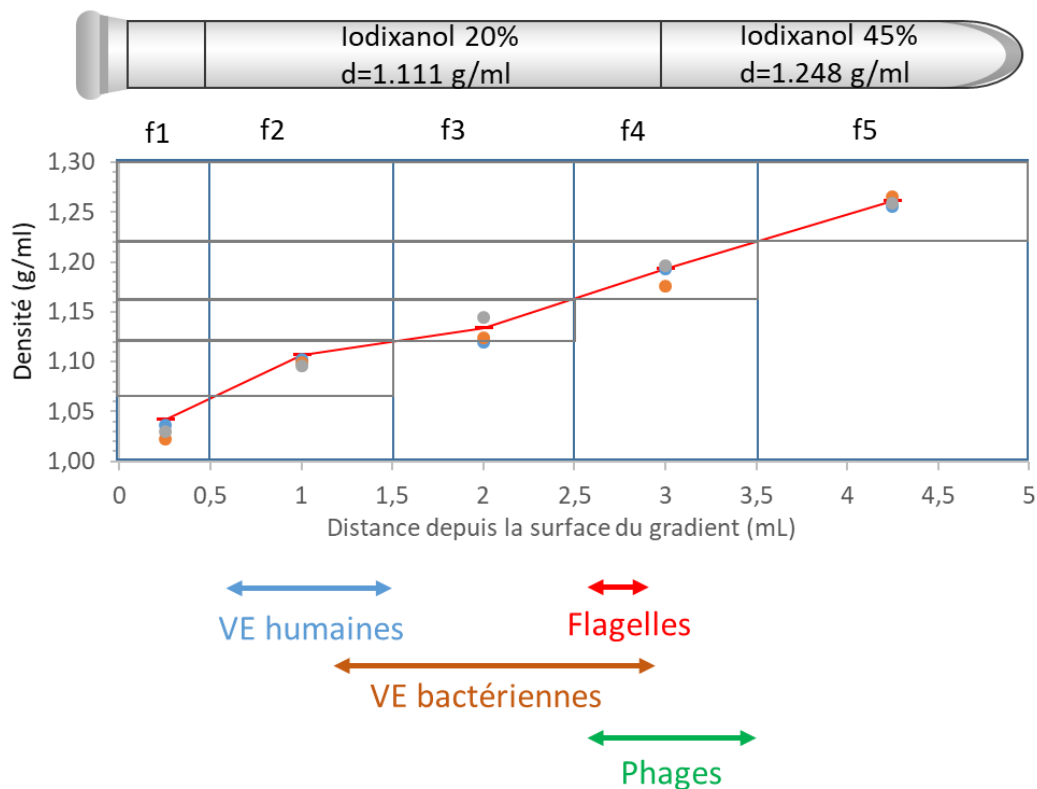


Figure 19. Schéma présentant la densité des fractions prélevées et les nanoparticules attendues dans les fractions

Les densités ont été calculées à partir des indices de réfraction mesurés dans les fractions issues de trois culots de filtrats fécaux (cf. 3.3. Préparation des filtrats fécaux) soumis à un gradient d'iodixanol. En iodixanol, les VE humaines sont trouvées à une densité comprise entre 1,08 et 1,14 g/mL (Kowal et al. 2016, Tulkens, De Wever, and Hendrix 2020), les VE bactériennes à une densité comprise entre 1,13 et 1,20 g/ml (Tulkens, De Wever, and Hendrix 2020), les flagelles à une densité de 1,16 g/ml (Kivela et al. 1999, Ravantti et al. 2003) et les phages à une densité comprise entre 1,16 et 1,22 g/ml (Chapuy-Regaud et al. 2017, Kivela et al. 1999).

Neuf préparations de phages différents et neuf lots de VE d'origines différentes ont été sélectionnés, produits et purifiés à l'aide de gradients d'iodixanol au cours de la thèse. Ils sont tous présentés, ainsi que l'intégralité de leurs protocoles de production et de purification, au sein de l'article 1.

1.2. Dénombrements de populations homogènes de phages et de vésicules

L'ensemble des résultats concernant la quantification des concentrations et l'obtention de distributions de tailles des phages et des VE sont décrits dans l'article 1. La numérotation des figures présentée dans ce résumé correspond à celle des figures présentées dans l'article.

1.2.1. Résumé de l'article 1

Le marché de la détection et de la quantification des nanoparticules, bien que déjà occupé par le NTA, le TRPS et la DLS, laisse encore de la marge à de nombreuses améliorations techniques. En effet, si les appareils basés sur le NTA sont les plus couramment utilisés pour la caractérisation de solutions de phages et de VE, l'utilisation de ceux-ci est chronophage et soumise à certaines limites importantes, comme leur sensibilité à la présence de bulles d'air dans leurs systèmes, le fait que les plus petites particules soient cachées par les plus grosses, ainsi que la relative fragilité de ces appareils. Le Videodrop, un nouvel appareil basé sur l'ILM et développé et commercialisé par Myriade est entré il y a peu sur ce marché afin de répondre à certaines de ces limites, tout en proposant un appareil à un prix plus attractif. Le Videodrop a été comparé à deux appareils basés sur le NTA, à savoir le NanoSight et le ZetaView, pour évaluer sa capacité à détecter et à quantifier les concentrations dans des populations homogènes de phages et de vésicules, avant de potentiellement l'utiliser ensuite pour déterminer les concentrations en phages et en VE dans les selles humaines.

1.2.1.1. Quantifications des phages purifiés

Neuf phages, aux morphologies et aux génomes variés et pour certains, représentatifs des communautés phagiques connues dans le microbiote intestinal humain (présentés dans le Tableau 2), ont été purifiés puis dénombrés. Ainsi, l'utilisation de la microscopie par épifluorescence s'est avérée comme la méthode la plus pertinente pour le dénombrement des phages (Figure 2). Donnant des résultats similaires à la qPCR, elle présente l'avantage de ne pas nécessiter de connaître les séquences génomiques des phages pour les détecter, et donc de permettre de quantifier tous les phages dans des mélanges de populations. À noter que les concentrations données par la microscopie par épifluorescence et la qPCR sont, pour certains phages comme T7, SPP1 et PRD1, supérieures à leur titre infectieux obtenu par plages de lyse, suggérant la présence de virions non infectieux dans ces préparations. Le Videodrop, le ZetaView et le NanoSight ont ensuite été utilisés pour mesurer les concentrations des différents stocks de phages purifiés, par rapport aux résultats de microscopie par épifluorescence. Parmi tous les phages quantifiés, le Videodrop n'a pu donner des concentrations similaires à la microscopie par épifluorescence que pour T4 et P1, deux myophages aux diamètres hydrodynamiques supérieurs à 100 nm (Figure 3). Le Videodrop a également pu détecter les phages T5 et Φ crAss001, dont les capsides mesurent environ 90 nm de

diamètre, mais à des concentrations égales à moins de 10% de celles obtenues en microscopie par épifluorescence. Le ZetaView et le NanoSight sont quant à eux capables de détecter des phages de plus petite taille. La limite de détection des phages pour le NanoSight se situe ainsi autour de 60 nm de diamètre (le NanoSight est capable de détecter tous les phages de la collection, excepté le plus petit de celle-ci, Φ X174, dont le diamètre est de 30 nm), tandis que celle du ZetaView se situe entre 60 et 75 nm de diamètre. De plus, le NanoSight apparaît comme étant l'appareil le plus performant pour les mesures de tailles des phages (Figure 3).

1.2.1.2. Quantifications de vésicules purifiées

Des vésicules purifiées issues de cultures bactériennes, de cultures cellulaires eucaryotes et de fèces de rongeurs ont également été purifiées pour être analysées par les trois appareils comparés dans l'article. Bien que la MET ne nous permette uniquement l'obtention de distributions de tailles approximatives en ce qui concerne les VE, c'est la méthode de mesure actuellement disponible qui se rapproche le plus de la réalité des échantillons utilisés pour la comparaison des trois appareils. Ainsi, nous avons décrit dans cet article que les seuils de détections propres à chaque appareil, déterminés à l'aide des phages, semblent s'appliquer également aux VE. Dans les populations où les VE présentent des diamètres majoritairement supérieurs à 100 nm, comme c'est le cas pour les VE des monocytes THP1, les trois appareils donnent des estimations de la concentration similaires (Figure 4). À l'inverse, dans les échantillons où la taille de la majorité des VE est inférieure au seuil de détection du Videodrop mais supérieure à celui des appareils basés sur le NTA, les concentrations données par le NanoSight et le ZetaView sont nettement supérieures à celles proposées par le Videodrop (Figure 4).

Cependant, bien que permettant la détection de phages et de VE de plus petite taille, le NanoSight et le ZetaView n'autorisent pas l'analyse d'échantillons non purifiés, comme les filtrats fécaux analysés au cours de ma thèse. La présence de flagelles et d'agrégats protéiques et lipidiques au sein de ces derniers empêche leur caractérisation par ces deux appareils, dans la mesure où les systèmes de fluide de ceux-ci sont difficile à laver et pourraient se retrouver bouchés par de tels objets. Ce n'est pas le cas du Videodrop, qui ne nécessite que le dépôt d'une goutte dans une cuvette très facile à nettoyer. De plus, ce travail a permis de mettre en évidence un gain de temps considérable lors de l'utilisation du Videodrop (permettant d'analyser 50 échantillons en 3 heures), plutôt que du ZetaView ou du NanoSight, appareil le plus chronophage de ce trio (qui ne permet que l'analyse de 5 à 15 échantillons lors d'une session de 3 heures).

Les appareils reposant sur le NTA semblent affectés par l'effet d'écran, comme cela a déjà été décrit. Ce phénomène se caractérise, dans les échantillons où les nanoparticules présentes des tailles hétérogènes, par le masquage des plus petites par les plus grosses. Grâce à ses différences de conception, cet effet est en théorie beaucoup moins important avec le Videodrop. D'autres expérimentations seront nécessaires pour l'affirmer.

En conclusion, le Videodrop ne détecte et quantifie correctement que les phages caudés dont la capsidie présente un diamètre supérieur à 90 nm. Concernant les VE, il est difficile de déterminer un seuil de taille au-dessus duquel elles sont détectées, bien qu'il soit probablement similaire à celui des phages. Il faut donc bien avoir conscience que l'utilisation du Videodrop pour le dénombrement absolu ou relatif de VE ne prend en compte que celles dont le diamètre leur permet d'être détectées par l'appareil. Enfin, le Videodrop est l'appareil le moins chronophage parmi les trois outils comparés.

TITLE : Comparison of Interferometric Light Microscopy with Nanoparticle Tracking Analysis for the study of extracellular vesicles and phages

Romain Sausset^{1,2}, Zuzana Krupova³, Eric Guédon⁴, Sandrine Peron⁴, Alice Grangier⁵, Marie-Agnès Petit¹, Luisa De Sordi⁶, Marianne De Paepe¹

¹Micalis Institute, INRAe, AgroParisTech, Université Paris-Saclay, 78350, Jouy-en-Josas, France

²Myriade, 68 boulevard de Port Royal, 75005, Paris, France

³Excilone, Département R&D, 6 rue Blaise Pascal, Parc Euclide, Bat. A, 78990 Elancourt, France

⁴INRAe, Institut Agro, STLO, Rennes, France

⁵Laboratoire MSC Matière et Systèmes Complexes, CNRS UMR 7057, Université de Paris, 75013 and 75006 Paris, France

⁶Sorbonne Université, Centre de Recherche Saint Antoine, UMRS_938, Paris, France

ABSTRACT

The research on extracellular vesicles (EVs) and bacteriophages (phages) has been constantly expanding during the last decades, as many of their roles in medicine, biology and ecosystems were unveiled. This interest brought a need of new tools to quantify and determine the sizes of these biological nanoparticles. A new device based on interferometric light microscopy (ILM), the Videodrop, was recently developed for such purposes. Here, we compared this new device to two NTA devices commonly used for the study of EVs, the NanoSight (Malvern, UK) and the ZetaView (Particle Metrix, Germany), both based on Nanoparticle Tracking Analysis (NTA). We show that the performance of the Videodrop compares favorably with those of the NTA devices for the study of most EVs. Concerning bacteriophages, results are more mitigated, as only larger ones are detected. We also show that while the NanoSight instrument gives the most precise size distributions and that generally the ZetaView provides the most accurate estimations of concentrations, the Videodrop is by far the most time-saving device, making it the most appropriate for studies conducted on a large number of samples.

INTRODUCTION

The discovery of extracellular vesicles (EVs) importance in basic research, medicine, and industrial applications have put them in the limelight in the last decades. EVs are produced and released in

the environment by eukaryotic and prokaryotic cells, for which they are important actors in cell-cell and cell-host communication. Human EVs are present in numerous body fluids (Slomka et al. 2018), where they are implied in homeostatic processes (Ratajczak et al. 2006, Simons and Raposo 2009, They, Ostrowski, and Segura 2009, Valadi et al. 2007, Vallabhaneni et al. 2015, van der Pol et al. 2012) and constitute promising sources of biomarkers in cancer (Lane et al. 2018). They are also implied in neurodegenerative and cardiovascular diseases (Robbins, Dorronsoro, and Booker 2016, Thompson et al. 2016). Bacterial EVs play important roles as well, notably in pathogeny, as some pathogenic bacteria deliver toxic compounds through EVs during infection (Bitto et al. 2017, Bomberger et al. 2009, Codemo et al. 2018, Furuta, Takeuchi, and Amano 2009). Bacterial EVs have also been found to be abundant in aquatic ecosystems (Biller et al. 2014). Finally, EVs have attracted enormous interest in medicine and pharmaceutical industry as they can be used as vehicles to deliver active therapeutical compounds, and serve as biomarkers, as reviewed in (Mateescu et al. 2017).

All these studies and applications have driven the development of tools for the characterization of EV preparations. However, there is still a lack for simple, fast and reliable tool for the determination of sizes and concentrations of entire populations of EVs, i.e. without labelling of specific molecules that could be absent from a subpopulation (Filip 2021, They et al. 2018). Single-particle measurements are required because of the important heterogeneity of EVs, notably concerning their size, that typically range from 20 to 200 nm. Traditional microscopy techniques are not appropriate for EVs enumeration and sizing. Indeed, enumeration of EVs by epifluorescence microscopy is difficult, as they are not easily stained with non-specific dyes (i.e. dyes that would stain all EVs). In addition, fluorescence microscopy does not give information on the size of objects. On the contrary, Transmission Electronic Microscopy (TEM) can be used to visualize EVs and determine their size, but not for enumeration.

Therefore more recent techniques are used for the enumeration and size determination of EVs, such as Tunable resistive pulse sensing (TRPS), Microfluidic resistive pulse sensing (MRPS), Nanoparticle Flow Cytometry (NFCM), Dynamic Light Scattering (DLS), and Nanoparticle Tracking Analysis (NTA), but all these techniques have some limitations. TRPS relies on changes of impedance of a nanopore caused by the passage of a nanoparticle in an electrolyte fluid. Though very valuable and accurate for the characterization of EVs relatively homogenous in size (Doyle and Wang 2019, Maas, De Vrij, and Broekman 2014), it faces impractical issues for highly polydisperse populations. Indeed, the dimensions of the nanopores have to be adapted to the dimension of the analyzed particles, which is not possible in the case of unknown or very polydisperse samples (van der Pol et al. 2014). By the

addition of microfluidics to RPS, MRPS is more adapted to the analysis of size-heterogeneous populations, its exposition to clogging issues is limited, but biases can still be induced by the choice of the pore's size (Cimorelli et al. 2021). In addition, this technology remains more expensive than the others. Nanoflow cytometry (NFCM) is a flow-based technique that detects light scattered by nanosized particles. Compared to traditional flow cytometry, a smaller flow channel reduces the background signal and reduces system pressure, increasing particle residence time and therefore better signal integration (Lippe 2018, Rossi et al. 2015, Zamora and Aguilar 2018). DLS is a bulk method that detects the temporal fluctuations of intensities of the light scattered by a population of nanoparticles following illumination by a laser to calculate their diffusion coefficient (Doyle and Wang 2019). As the diffusion coefficients of particles are inversely proportional to their hydrodynamic diameter, this information gives the size of particles (Dragovic et al. 2011, Kestens et al. 2017). Though very valuable for the study of very small particles, as all bulk methods, it is prone to biases arising from the heterogeneity of samples (Filipe, Hawe, and Jiskoot 2010). NTA is, to date, the most widely used technique for the study of EVs. Like DLS and NFCM, NTA exploits light scattered by nanoparticles in solution upon illumination with a laser, but the trajectories of single nanoparticles are followed, enabling to determine their hydrodynamic diameter through the analysis of their Brownian motion. It can therefore provide good size distributions and quantification of heterogeneous-sized samples of EVs. The ZetaView (Particle Metrix, Germany) and the NanoSight (Malvern, UK) are two devices relying on NTA commonly used by researchers working with EVs. Even though they are based on this same principle, the ZetaView and the NanoSight have important differences in the composition of their hardwares and softwares (Bachurski et al. 2019). For example, on the NanoSight instrument, a flux is generally applied to the sample during acquisition, whereas the acquisition is recorded on a static sample with the ZetaView. On polystyrene and silica nanospheres, the ZetaView was shown to provide better concentration measurements, while the NanoSight provided higher precision on size (Bachurski et al. 2019).

Despite the strengths of NTA, its main limitation result from the sixth-power dependence of the scattered light intensity on the particle size, resulting in masking of small particles by larger ones, which leads to inaccurate measurements in polydisperse samples (Dehghani et al. 2021, Gardiner et al. 2013). In addition, the use of NTA-based devices is relatively time-consuming (approximately 10 and 20 minutes per sample for the ZetaView and the NanoSight respectively), which is a source of practical limitations. A fast (i.e. whole acquisition in less than a few minutes), reliable and very easy-to-use tool to size and enumerate nanoparticles would facilitate studies including a large number of samples, or necessitating regular monitoring of the same samples over time. These are

the announced strengths of a new instrument based on Interferometric Light Microscopy (ILM), the Videodrop (Myriade, France). Like NTA, ILM uses the Brownian motion and the application of the Stokes-Einstein relation to calculate the size distributions of the analyzed particles. However, a simple LED illuminates the samples, and a transmission bright-field microscope is used as a homodyne interferometer to detect the particles, based on the interferences created by the superposition of the incoming light field and the light scattered by the nanoparticles. Being mostly in the Rayleigh scattering regime, the intensity of the scattered amplitude is proportional to the third-power of the particle diameter (Hulst 1957).

The Videodrop has been developed from an interferometer (Boccarda et al. 2016, Roose-Amsaleg et al. 2017), which has been used to enumerate nanoparticles in cheeses (Dugat-Bony et al. 2020) and aquatic environments (Boccarda et al. 2016, Roose-Amsaleg et al. 2017). A recent study showed that the Videodrop enables correct relative enumeration of lentiviral virions, an adenovirus, and a baculovirus, while underestimating viral concentrations as compared to classical titration procedures, especially for the adenovirus (Turkki et al. 2021). A Videodrop instrument was also used to estimate the concentrations of EVs in plasma, but no controls concerning the performance of the device were realized in this study (Sabbagh et al. 2021). **Table 1** summarizes the main characteristics of the Videodrop, ZetaView and NanoSight instruments.

Here, we evaluated the Videodrop for the determination of concentrations and sizes of purified EVs and bacteriophages (phages), and compared it with two NTA devices, the ZetaView and the NanoSight. Phages are interesting biological nanoparticles for this evaluation as they constitute monodisperse populations, the diameter of the capsid being very homogenous among virions of the same species. In addition, on the contrary to EVs, they can be reliably quantified using different techniques such as plaque counting, quantitative PCR and epifluorescence microscopy following labeling of the nucleic acid in virions. For this study, phages were chosen to cover a large range of capsid diameters, from 30 nm and 120 nm, and to have different morphotypes: we used Myophages and Siphophages that have long proteic tails, and Podophages and *Tectiviridae* that have no or very small tails (**Table 2**). In addition, phage enumeration is of great interest in itself, since phages, as predators of bacteria, are important actors in all microbial ecosystems, and notably in Human associated microbiota (Sausset et al. 2020). Phage composition has been shown, for example, to be associated with Intestinal Bowel Diseases (IBD), such as Crohn's disease or ulcerative colitis (Cornuault et al. 2018, Diard et al. 2017, Norman et al. 2015). To understand the dynamics of complex microbial ecosystems, but also for the development of phage-based applications in biotechnologies and medicine, a fast and reliable enumeration of phage particles is of considerable

interest. In this study, we then compared the performance of the ILM and NTA devices for the observation of nine types of EVs, originating either from milk, from cells (human and bacterial), or from rodent faeces (germ-free or raised conventionally). For each device, we choose to use constant settings for all experiments to enable comparisons. We compared the size distributions obtained with Videodrop and the NTA devices with size distributions obtained by transmission electron microscopy. We show that the Videodrop detection threshold is higher than that of the NTA devices, but that it provides concentrations similar to the ones obtained by NTA for most of mammal derived EVs. In addition, the Videodrop is by far the most time-saving device, making it the most appropriate for studies conducted on a large number of samples and comprising sufficiently large objects.

MATERIAL AND METHODS

Phages lysates. Phage bacterial hosts and conditions of culture are listed in **Sup Table 1**. *Escherichia coli* and *Bacillus subtilis* cultures were grown at 37°C in LB (Luria Broth) supplemented with 5 mM of MgSO₄, 5 mM of CaCl₂, and maltose for lambda phages. The cultures were infected during exponential growth (OD_{600 nm} of 0.2). *E. coli* phages were grown on the MG1655 *hsdRM* strain, and the multiplicity of infection (m.o.i.) was comprised between 0.1 and 0.25, except for PRD1 and ΦX174 which were respectively grown at 37°C on *E. coli* HMS174 pL2 (with kanamycin at a final concentration of 25 µg/mL) and at 30°C on *E. coli* CQ2, and for which an m.o.i. of 5 was used. SPP1 was grown on *B. subtilis* YB886, and a m.o.i. of 0.01 was used. ΦCrAss001 was grown as described by Shkoporov *et al.* (Shkoporov, Khokhlova, et al. 2018). Briefly, *Bacteroides intestinalis* (DSM 108646) was grown in Fastidious Anaerobic Broth (FAB) in anaerobic conditions and infected at a m.o.i. of 1. All infections were performed in 500 mL of exponentially growing bacterial cultures (0.2 at 600 nm); and incubated until a deceleration of the bacterial growth was observed. Bacteria and debris were then pelleted by centrifugation at 5,200 g for 30 minutes, and the supernatants were filtered through 0.22 µm syringe filters.

Extracellular vesicles preparations. Bacterial culture supernatants containing EVs were obtained as follows. *B. subtilis* YB886 was grown in 500 mL of LB during 24 hours at 37°C with agitation. *Faecalibacterium prausnitzii* was grown in an anaerobic chamber filled with 5% H₂, 5% CO₂ and 90% N₂ in 500 mL of sterile BHIS (Brain Heart Infusion Supplemented) medium supplemented with L-cysteine (0,5 g/mL), maltose (1 g/mL), and cellobiose (1 g/mL), during 24 hours, at 37°C. In both cases, bacteria and debris were pelleted by centrifugation at 5,200 g, for 30 minutes, and the

supernatants were filtered with 0.22 µm syringe filters. *Staphylococcus aureus* HG003 was pre-inoculated in BHI broth and grown overnight at 37°C under 150 rpm/min agitation, and then inoculated 0.1% in 500 mL of fresh BHI (125 rpm/min, at 37°C) on a 1 L Scott flask. Bacterial cultures were retrieved after 12 h, centrifugated at 6,000 g for 15 min and filtered through 0.22 µm. *E. coli* Nissle 1917 was pre-inoculated in LB broth and grown 8 h at 37°C under 200 rpm/min agitation, and then inoculated 1% in 500 ml of fresh LB (130 rpm/min, at 37°C) on a 1 L Scott flask. Bacterial cultures were retrieved after 18 h, centrifugated at 10,000 g for 20 minutes and filtered through 0.22 µm pore size filters. To recover EVs from murine faeces and caecal content, faeces and caecal contents were collected from mice. The mice were 7 weeks old C57BL/6NRj males (Janvier Labs) maintained in 12 h-light/12 h-dark cycle and fed on a chow diet (Ssniff), or germ-free C3H/HeN and C57BL/6J mice from the INRAE Anaxem core facility. Faeces from germ-free F344 Fisher rats grown in the Anaxem-Micalis (INRAE, Jouy-en-Josas, France) core facility were also obtained. In order to obtain fecal filtrates, fresh or frozen (at -80°C, immediately after sampling) materials were diluted 40-fold in cold 10 mM Tris (pH 7,5) and resuspended by soft agitation at 4°C during 15 minutes. After centrifugation at 5,200 g for 30 minutes, the supernatant was filtered with 0.22 µm pore size syringe filters.

Purification of bacterial EVs and phages. Phages lysates and *B. subtilis* and *F. praustnitzii* supernatants prepared as described above were submitted to a centrifugation of 20,000 g for 16 hours (rotor SS-34; Sorvall RC 5C PLUS). Pellets containing phages and EVs were resuspended in SM buffer (200 mM NaCl, 50 mM Tris pH7, 10 mM MgSO₄) or Tris 10 mM pH7, respectively. Resuspended pellets were then subjected to an iodixanol density gradient in 5 mL Ultra-Clear centrifuge tubes (Beckman Coulter). The tubes were first filled with a two-layer gradient: 1.875 mL of 45% iodixanol was softly injected under 2.500 mL of 20% iodixanol. 512 µL of the resuspended pellets were layered on the top of the density gradient. After ultracentrifugation at 32,000 rpm during 5 hours at 4°C (rotor SW55Ti; in a XL-90 Beckman Coulter centrifuge), fractions containing purified EVs or purified phages were extracted following the scheme in **Supp. Fig. 1**. These fractions of interest were dialyzed overnight at 4°C against Tris 10 mM or SM buffer for EVs and phages respectively, under soft agitation, using 25 kD Spectra/Por dialysis membranes. A second dialysis of 3 hours was performed the next day in the same conditions. All samples were stored at 4°C and analyzed in the following two weeks. TEM was performed to check the quality of all EV samples. Cell-free supernatants from *S. aureus* cultures were submitted to EVs isolation and purification by sucrose density gradient ultracentrifugation, as described previously (Luz et al. 2021). Cell-free

supernatants from *E. coli* cultures were submitted to EVs isolation and purification by size-exclusion chromatography, as described previously (Rodvalho et al. 2020).

Purification of THP1 EVs. THP1 cells (ATCC) were cultivated in RPMI medium supplemented with 10% FBS and 1% penicillin-streptomycin at 37°C in 5% CO₂. They were kept at a concentration between 3.0 x 10⁵ cells/mL and 1.0 x 10⁶ cells/mL. For EVs production, the cells were washed with RPMI + 1% PS without serum, and then resuspended in the same media in T175 flasks at a concentration of 2.5 x 10⁵ cells/mL in 50 mL. After 48h, the conditioned media was harvested and centrifuged during 10 minutes at 2,000 g. The supernatant was ultracentrifuged 2h at 150,000 g in an Optima XP (Beckman) with the rotor MLA-50. The EV pellet was resuspended in 0.8 mL of sterile PBS. The EVs were aliquoted and stored at -80°C until further analyses.

Isolation of bovine milk-derived EVs. Whole bovine milk samples were centrifuged at 3,000 g for 15 minutes at 4°C (Allegra X-15R, Beckman Coulter, France) to separate fat from skimmed milk. The whey was obtained after acid precipitation with 10% (v/v) acetic acid, incubation at 37 °C for 10 minutes and 10% (v/v) 1M sodium acetate for 10 minutes at RT. This was followed by centrifugation at 1500 g, 4°C for 15 minutes and by a filtration using vacuum-driven filtration system Millipore Steritop, 0.22 µm. The whey supernatants were concentrated by centrifugation at 4,000 g and 20°C using Amicon 100kDa centrifugal filter units (Merck Millipore) to the final volume ~6 mL. The aliquots of 500 µl of obtained retentate were loaded onto a qEVoriginal 70 nm SEC column (Izon Science, New Zealand) previously washed and equilibrated with PBS. Fraction collection (0.5 mL per fraction) was carried out immediately using PBS as elution buffer. The selected elution fractions were pooled and were subsequently concentrated using 100 kDa Amicon centrifugal filter units (Merck Millipore). Post-treatment processing with several washing steps with PBS was applied to obtain highly pure EV population. The EV standards were aliquoted and stored at -80°C until further analyses.

Transmission Electron Microscopy (TEM). 10 µL of purified EVs or phages samples were directly adsorbed onto a carbon film membrane on a 300-mesh copper grid, stained with 1% uranyl acetate dissolved in distilled water and dried at room temperature. Grids were examined with a Hitachi HT7700 electron microscope operated at 80 kV (Elexience) and images were acquired with a charge

coupled device camera (AMT). This work was carried out at and with the expertise of the MIMA2 platform, INRAE (Jouy-en-Josas, France).

Plaque Assay (PA). 10 μ L of the appropriate dilution of the purified phage preparations were mixed with 300 μ L of their bacterial hosts grown overnight and supplemented with appropriate nutrients and/or antibiotics (**Supp Table 1**). For all phages except Φ crAss001, the phage-bacteria suspensions were mixed with 5 mL of warm soft top agar (0.45% w/v agar, 0.25% w/v NaCl, 0.1% w/v Bacto Tryptone; in osmosed water) and immediately poured in Petri dishes already containing a solid LB agar layer (1.5% w/v agar and 2.5% w/v LB powder), in technical triplicates. In the case of the Φ crAss001, a published protocol was followed (Shkoporov 2018). Briefly, the anaerobically prepared phage-bacteria suspension was mixed in 5 mL of 0.4% Bacto agar, and immediately poured on three Petri dishes already filled with a solid layer of Fastidious Anaerobic Agar (FAA). After solidification, the Petri dishes were incubated overnight at 37°C, except for Φ X174 (30°C). The next morning, lysis plaques were manually counted and phage titers in plaque-forming units per mL (PFU/mL) were calculated.

Quantitative PCR (qPCR). DNA standards for calibration were prepared from phage genomic DNA preparations. Prior to genomic extraction, 500 μ L of high titer phage samples were treated with 0.25 μ L of Turbo DNase I (Ambion, 2 U/ μ L) and 1 μ L of RNase A (10 mg/250 mL) at 37 °C for 30 min. After adding EDTA (at a final concentration of 0.01 M) to inactive nucleases, phage DNA was extracted with two phenol-chloroform-isoamyl alcohol (25:24:1) extractions followed by three chloroform-isoamyl alcohol (24:1) purification steps. DNA was precipitated with two volumes of ethanol and potassium acetate (300 mM, pH4.8) and resuspended in Tris buffer pH8 at 10 mM. DNA was quantified with Qubit® and diluted in Tris pH8 10 mM to obtain standards with 1.0×10^6 to 5.0×10^6 genomes of phage in 6 μ L. Phage samples were quantified using the primers in **Supp Table 2**, used at a concentration of 10 μ M. qPCR was performed in a total volume of 15 μ L in MicroAmp Fast Optical 96-wells plates sealed by a MicroAmp Optical Adhesive Film, using the Takyon ROX SYBR Mastermix blue dTTP kit. Amplifications were run in duplicates on a StepOnePlus real-time PCR system with the following cycling conditions: 95 °C for 5 min, (95 °C for 15 sec, 58 °C for 45 sec, 72 °C for 30 sec) for 45 cycles, 72 °C for 5 min, 95 °C for 15 sec, 60 °C for 15 sec, 95 °C for 15 sec. All phage preparations were quantified three times independently. The analysis of the melting curves confirmed the specificity of the primers. Data analysis was performed using manufacturer's StepOne Software 2.3.

Epifluorescence Microscopy (EPI). Phage samples were diluted to a concentration of approximately 10^7 virions per mL. Glutaraldehyde was added to 1 ml of diluted sample at a concentration of 0.5% final and incubated at 4°C for 15 min. Samples were then submitted to flash freezing in liquid nitrogen. After defreezing, 4 mL of SM buffer was added to each sample, prior to filtration on 0.02 μm Anodisc filters. Each filter was then incubated on a 50 μL drop of SYBR Gold at 200X in the dark for 15 min, with the virus side up. After removal from the drop, filters were dried in the dark before being mounted on a glass slide with Fluoromount-G and a coverslip. Slides were kept at -20°C until observation. Microscopic observations were done using a Nikon Ti-E fitted with a 100 \times oil objective Apo TIRF (NA, 1.49; Nikon) with an iLas2 laser coupling system from Gataca Systems (150 mW, 488 nm). Ten images were shot per slide, with bright field and GFP fluorescence channels (with an excitation filter wavelength of 472/30 nm and emission filter wavelength of 520/35 nm). Emissions were collected using interference filters and imaged onto a pair of sCMOS cameras (Orca Flash 4.0 v2 sCMOS; Hamamatsu), with a gain defined at 300 attached to a $\times 2.5$ magnification lens, with a time exposure set up from 100 ms to 200 ms depending on the fluorescence intensity of each phage. The final pixel size was 64 nm. Metamorph v.7 software packages were used to control and process the image acquisition, further analyzed with ImageJ (v1.52a). Numbers of phages on the whole filters were calculated by multiplying the average counts with the quotient of the area of the filter by the area of the pictures.

ILM measurements with the Videodrop. All purified samples were diluted to appropriate concentrations (20-100 particles per frame) in their respective buffers and drops of 6 μL were used for the measurements. Triplicates of accumulations on 10 acquisitions of 100 frames were recorded per sample, in accumulation mode. In accordance with the manufacturer's specifications, a relative threshold of 3.8 was applied. Removal of macro particles was enabled, using a minimum radius of 10 and a minimum hot pixels of 80, as well as a drift compensation. Concerning the tracking settings, a maximum of 2 jumps was tolerated, for a minimal length track of 10 frames. The doublet detector (qvir software version 2.5.2.6196) was used for all the measurements. Size distributions were obtained after the application of a mobile mean with a period of 3 on histograms with classes of 5 nm. For linearity measurements polystyrene beads (Polystyrene Nanosphere Suspension Series 3000, ThermoFisher) of various sizes were diluted in water after short sonication. Beads concentration was calculated following Nanospheres' supplier indications provided in the technical guide.

NTA measurements with the PMX 220 ZetaView.

The Zetaview system (Particle Metrix, Germany) was equipped with a 488 nm laser. Ideal measurement concentrations were obtained by pre-diluting the samples to the ideal 50–200 particles/frame. Each experiment was performed in duplicate on 11 different positions within the sample cell with following specifications and analysis parameters: cell temperature 25°C, sensitivity 70, shutter 100, Max Area 1000, Min Area 10, Min Brightness 25. The results were validated while obtaining at least 1 000 valid tracks for each run. For data capture and analysis, the Nanoparticle Tracking Analysis Software (ZNTA) vs 8.05.04 was used. Size distributions were obtained after the application of a mobile mean with a period of 3 on histograms with classes of 5 nm.

NTA measurements with the NanoSight NS300. All purified samples were diluted in their appropriate sterile buffer to a volume of 1 mL and at an ideal concentration of 20-100 particles per frame, and injected at a speed of 50 $\mu\text{L/s}$ into the machine's specimen chamber with a 1 mL sterile syringe. For each measurement, five acquisitions of 1 minute were recorded at 25°C. Concerning the hardware, the device was equipped with a sCMOS camera and a laser module at 488 nm for most experiments, except for phages PRD1 and ΦX174 and milk EVs, for which a laser module at 405 nm was used. Observation of phages T4 and PRD1 with both laser modules (at 405 nm and 488 nm) revealed almost no differences in concentration and sizes. A camera level of 15, a flux of 25 ml/h were used in all experiments. For each measurement, five 1-min videos were captured. After capture, the videos were analyzed by the NanoSight software NTA 3.3 Dev Build 3.3.104 with a detection threshold of 4. Size distributions were obtained after the application of a mobile mean with a period of 3 on histograms with classes of 5 nm.

RESULTS

Evaluation of the size threshold and linearity range of the Videodrop on polystyrene beads

In a first step, we observed standardized polystyrene beads of known concentrations at different dilutions with the Videodrop. For beads with a diameter superior to 80 nm, good agreement between theoretical and measured concentrations values was observed between 5×10^8 and 4×10^9 beads/mL (**Fig. S2**). For 70 nm beads, concentrations were underestimated by 5 to 10-fold, indicating that 80 nm is the size threshold for enumeration of polystyrene beads, in agreement with

the manufacturer's indications (**Table 1**). By comparison, linearity and agreement between theoretical and measured concentrations for 100 nm diameter polystyrene beads were previously determined with the NanoSight between 2×10^7 and 3×10^9 particles/mL (Maguire et al. 2017).

Of note, particle size thresholds are expected to be higher for phages and EVs. Indeed, the signal intensity depends not only on the diameter of the particle but also on its refractive index, and at wavelengths around 450 nm, the refractive index of polystyrene beads (approximately 1.59) is higher than that reported for viruses (between 1.42 and 1.49) (Holzwarth et al. 1974) (Pang, Song, and Cheng 2016) and EVs (between 1.36 and 1.39) (Chandler, Yeung, and Tait 2011, Gardiner et al. 2013, van der Pol et al. 2012).

Enumeration of purified phages

We then evaluated the performances of the Videodrop and the two NTA devices for measuring concentrations and sizes of biological nanoparticles of known concentrations and uniform in size, i.e. phages. To this aim, we first purified phage virions of nine different species, belonging to different taxonomic families with different capsid morphologies and sizes (**Table 2**). TEM observations indicated that phage preparations were almost free of EVs or other large contaminants (**Fig. 1**). When non-phage objects were observed, as in P1 and T5 preparations, they were 10-fold less abundant than phage particles, and should therefore not significantly modify neither sizing nor enumeration of the phage stock. Concerning PRD1 and Φ X174 preparations, virions were aggregated on TEM images, but subsequent observation in NTA indicated that this aggregation occurred on the TEM grids. Capsid and tail dimensions measured for each phage were close to published values (**Tables 2 & 3**).

We then determined the concentrations of our phage preparations using three reference methods: Epifluorescence Microscopy (EPI), Plaque Assay (PA), and quantitative PCR (qPCR). EPI and qPCR provided comparable concentrations, giving confidence that they correspond to reliable values (**Fig. 2**). In the following, we use the concentrations obtained by EPI as references for further comparisons. Concentrations obtained by PA were slightly lower to those obtained with EPI and qPCR for most of the phage samples (2 to 3-fold lower), except for T7 and SPP1 for which the differences were more important. This is most probably related to the presence of non-infectious particles in the preparations, as already reported for viruses (Heider and Metzner 2014, Huang and Baltimore 1970), especially after purification as procedures can damage particles. Comforting this hypothesis, tailless phage particles were observed in SPP1 stocks (**Fig. 1**).

We next compared the performances of the Videodrop and the two NTA devices in determining phage concentrations. In a first step, we measured concentrations of T4 serial dilutions with the Videodrop, which comforted the appropriate concentration range determined previously with polystyrene beads, i.e. from 5×10^8 to 5×10^9 particles/mL (**Fig. 3A**).

The NanoSight detected all the tested phages except the smallest one, Φ X174, whose capsid diameter is only 30 nm. NanoSight detection threshold is therefore between 30 and 60 nm for phages. Concentrations obtained were in relatively good accordance with EPI counts, with generally less than 2-fold difference, except for lambda and Φ CrAss001, for which 3-fold lower concentrations were found (**Fig 3A**). The ZetaView also detected most phages, except the small Φ X174, as the NanoSight, but also SPP1. This could be surprising as the published diameter of SPP1 virion, 61 nm, is similar to those of λ and T7 phages (**Table 2**), which were correctly detected. Of note, SPP1 particles create a visible signal, but that is too low to be detected by the ZetaView software with the parameters used in this study, indicating that SPP1 virions are just below the detection threshold of the device. This could result either from a slightly smaller size of SPP1 virions, as suggested by our TEM observations indicating a diameter around 57 nm (**Table 3**), or from a lower refractive index of SPP1 virions compared to other phages. We therefore considered that the detection threshold of the ZetaView is around 60 nm for phages. For the largest phages (T4, P1, T5 and Φ CrAss001), ZetaView concentrations were in very good agreement with EPI counts.

For its part, Videodrop detected and correctly enumerated only the two biggest phages, T4 and P1. Interestingly, although T5, P1 and Φ CrAss001 have approximately the same capsid diameter, 90 nm (**Tables 2 and 3**), only P1 virions were detected. This probably results from the larger P1 tail, that might contribute to the diffusion of light and affect the Brownian motion of the virions, facilitating its detection and tracking. Like for SPP1 with ZetaView, Φ CrAss001 particles create visible spots, but the intensity of the signal was too low to be detected by the software, indicating that the particle signal is just below the detection threshold of the device. Φ CrAss001 capsid size being between 76 and 90 nm (**Tables 2 & 3**), this suggests that the Videodrop detection threshold is around 90 nm for phages.

In conclusion, NanoSight and ZetaView enumerated correctly phage particles larger than 50 nm and 60 nm respectively, with globally better determination of concentrations by ZetaView, in line with previous studies on synthetic nanoparticles (Bachurski et al. 2019, Dehghani et al. 2021). The Videodrop detected and provided correct concentrations only for tailed phages with a capsid larger than 90 nm.

Particle size distributions in phage samples

In addition to the determination of particle concentrations, the three devices provide particle size distribution, thanks to the analysis of particle's Brownian motion. The NanoSight provided size distributions in agreement with sample's characteristics, i.e. in most cases narrow unimodal distributions with a maximal value close to the diameter of the virions (**Fig. 3B**). Hydrodynamic diameters of tailed phages is not expected to be exactly equal to the diameter of the capsid, in particular because the tail might influence the Brownian motion of virions. In agreement with this hypothesis, for phages with no or very short tail (Φ crAss001, T7 and PRD1), sizes obtained by NanoSight were very close to the diameters measured by TEM (**Table 3 and Fig. 3**). For phages with a large tail, a bimodal distribution could be observed most of the time, with a first small peak corresponding to the diameter of the capsid, and hence to damaged tailless virions, and a second higher peak, most likely corresponding to intact virions (**Fig. 3**). The bimodal distribution is particularly apparent in the case of the SPP1 sample, which was shown by TEM to contain close to 50% of tailless capsids (**Fig. 1**).

The ZetaView measured correct modal sizes for T4, T5 and Φ crAss001, but with poor precision (**Table 3 and Fig. 3**). Similarly, the Videodrop measured a correct modal size for phage T4, with poor precision as well. Surprisingly, in other cases the Videodrop and the ZetaView provided particle size distributions very different from expected distributions, concerning both mode and precision. In conclusion, the NanoSight provides more precise size estimations on phage particles than the other devices.

Quantification and size distribution of EVs from different origins

We next compared the performance of the three instruments in measuring the concentrations and size distributions of EV preparations from various origins. To this aim, we prepared EVs-enriched preparations (named EVs in the rest of the article for simplicity) from 9 different origins: bovine milk; THP1 human monocytic cell cultures; *Faecalibacterium prausnitzii*, *Bacillus subtilis*, *Escherichia coli* and *Staphylococcus aureus* bacterial cultures; and finally EVs from germ-free or conventional rodent faeces or ceecal content. In germ-free animals, EVs can exclusively originate from intestinal cells, while in conventional animals they can also originate from the intestinal microbiota. EVs preparations observations by TEM confirmed that most of the nanoparticles were EVs, with important dispersion of particle sizes (**Fig. 4**).

EV preparations were then observed with Videodrop, ZetaView and NanoSight. In a first step, we determined the optimal concentration range for measurements by determining the concentrations of serial dilutions of EVs from bovine milk with the three devices (**Fig. 5**). For both NTA instruments, linearity ranges were much narrower than previously reported for monodisperse samples: they range from 2.5×10^7 to 2×10^8 particles/mL instead of 3×10^9 . This probably results from the high camera levels used, chosen to detect all particles, that lead to a masking of small particles by large ones at higher concentrations. On the contrary, the linearity range of the Videodrop was similar to that determined previously with polystyrene beads and phage particles.

All EVs preparations were then observed with the three devices. Contrarily to phage preparations, the concentrations of EV preparations are unknown, and therefore we can simply compare the values obtained with the three instruments to each other. Concerning EVs originating at least partly from mammal's cells (i.e. EV from milk, THP1 monocytes and mouse intestinal contents), the three devices measured relatively close concentrations, with only 2-fold difference between the lowest and the highest concentrations, always provided by the ZetaView (**Fig. 6A**). The good performance of the Videodrop is surprising considering the presence, at least in some samples, of EVs smaller than its detection threshold, as deduced from TEM images (**Fig. 6C**). This could result from a lower masking effect in ILM than in NTA, as predicted by theory, which could compensate the higher detection threshold of Videodrop. In contrast, the devices were discordant in quantifying EVs originating from bacterial cultures, except for *E. coli*, with up to 10-fold higher concentrations obtained by NanoSight compared to Videodrop (**Fig. 6B**). Since the level of dispersity in EV size suggested by TEM images is similar for both types of EVs, we have no explanation for this result. *E. coli* EVs were smaller than the detection threshold of the three devices, probably explaining the similar concentrations obtained.

With regards to size distributions, all devices provided size distributions larger than TEM, yet, like for phages, the NanoSight provided the best measures, i.e. the distributions the closest to the ones determined by TEM (**Fig. 6**).

DISCUSSION

This study is, to the best of our knowledge, the first evaluation of the Videodrop for the measurement of concentrations and sizes of bacterial viruses and EVs. As a first step towards the evaluation of the Videodrop for phage enumeration, we realized a thorough comparison of classical

methods for phage concentration measurement. We established that quantitative PCR and epifluorescence microscopy provide very similar concentrations. A previous study found significantly lower concentrations with epifluorescence microscopy when compared to qPCR (Kaletta et al. 2020). The better agreements between the two methods in our study could result from the use of a more concentrated SYBR Gold staining solution, the adding of a flash freezing step to improve the staining, and/or to the use of a better microscopic setup. We found that the sensitivity threshold of the Videodrop for phage enumeration is around 90 nm of diameter, as only 5 to 10% of the 90 nm diameter virions of T5 and Φ crAss001 are counted. This contrasts with the reported good performance of Videodrop for the quantification of an adenovirus of 75.5 nm of diameter (Turkki et al. 2021). In this study, the authors could detect approximately 25% of adenoviruses particles (Turkki et al. 2021). Several factors could explain this difference. First, different viruses can have different refractive indexes, and the refractive index of adenoviruses might be higher than that of the phages tested here, improving its detection. Second, adenovirus observations were done using a relative threshold of 3.2, while we used a relative threshold of 3.8, in line with manufacturer's recommendations.

For the large T4 phage virus (120 nm of diameter), concentrations provided by the Videodrop instrument were in very good agreement with those obtained by epifluorescence and quantitative PCR. Similar to polystyrene nanoparticles, a linear count consistent with the dilution was observed between 5×10^8 and 4×10^9 particles/mL, indicating the successful enumeration of individual virions. Concerning the NTA instruments, our results suggest a sensitivity threshold for phages below 60 nm of diameter for NanoSight, and between 60 and 75 nm for Zetaview under the selected experimental conditions, in agreement with previous studies (van der Pol et al. 2014). We found that the ZetaView instrument provided slightly better concentration measurements than the NanoSight, as already described with polystyrene nanospheres, silica nanospheres and EVs (Bachurski et al. 2019).

Concerning the determination of sizes, in monodisperse phage samples, the NanoSight provided the most precise measurements, in line with a previous study (Bachurski et al. 2019). In polydisperse samples, we confirm the overestimation of particle sizes by NTA (Bachurski et al. 2019). As already discussed, this difference is too important to be solely explained by the shrinking phenomenon sometimes observed with TEM (Jensen, Prause, and Laursen 1981, Momen-Heravi et al. 2012, Szatanek et al. 2017, They et al. 2006, van Niel, D'Angelo, and Raposo 2018), to which EVs are particularly subjected. Rather, this overestimation most probably results from two phenomenon. The first is that small particles are difficult to track, due to low intensity and fast movement, and

therefore often cannot be attributed a size. The second is the so called masking effect, described in NTA-based devices, by which small particles mask small ones (Dehghani et al. 2021, Ortega Arroyo, Cole, and Kukura 2016). In theory, ILM is less subjected to this phenomenon because the detected signal is proportional to the third power of the diameter of the nanoparticles, while in NTA, the detected signal is proportional to the sixth power of the particle's diameter. This might explain why, despite a lower sensitivity of the Videodrop instrument, it measures concentrations of EVs similar to that of the NTA instruments in milk and in mice intestinal samples, that are highly polydisperse in size.

As measurements by the Videodrop seems to be not very affected by the masking effect, integration of a camera with a better sensitivity in a future version of the Videodrop might result in better performance than NTA for measuring concentrations of EV in polydisperse samples.

Finally, a major advantage of the Videodrop instrument is its ease and speed of use. In particular, sample positioning is very fast (a 6 μ L drop of the sample has to be deposited on a dedicated microcavity) compared to the positioning of the sample through a syringe and tubes in NTA, in which a syringe has to be carefully related to and removed from tubes to avoid generating air bubbles, increasing the required time of manipulation (**Table 4**). The ZetaView is facing similar limitations than the NanoSight; however, even though it necessitates a calibration prior to use, this device offers a better efficiency in terms of time of analysis. The ultra-fast measurements possible with the Videodrop is of particular importance when a large number of samples have to be observed. In addition, and especially with the NanoSight, clogging of the tubing could occur, necessitating to use only samples free from large particles and, ideally, from soluble lipids or proteins. For example, in the context of this study, fecal filtrates could not be observed with the NTA instruments, as they contain flagella, lipids and proteins either soluble or in aggregates, but could be observed with the Videodrop, which does not face clogging issues (**Fig. S3**).

In conclusion, the better sensitivity of the NanoSight instrument makes it the most appropriate device for the determination of the concentration of objects whose size is comprised between 50 and 90 nm. In addition, of the three devices we tested, the NanoSight provided the most precise size distributions, and projects requiring precision on a limited number of samples might take better advantage of this device. However, the higher precision and sensitivity of the NanoSight is balanced by three disadvantages: its time-consuming manipulation, its high susceptibility to masking effect and its price. For its part, the most important advantage of the Videodrop is its ease of use and the

rapidity of sample observation (**Table 2**). In addition, ILM is less prone to small-particle masking by larger particles, and can therefore compare advantageously to NTA for concentration measurements of highly polydisperse populations. Finally, Videodrop is cheaper and requires less consumables than the other devices (a simple pipet tip as opposed to a syringe), which is directly related to a lower environmental footprint, which should be taken into consideration.

REFERENCES

- Bachurski, D., M. Schuldner, P. H. Nguyen, A. Malz, K. S. Reiners, P. C. Grenzi, F. Babatz, A. C. Schauss, H. P. Hansen, M. Hallek, and E. Pogge von Strandmann. 2019. "Extracellular vesicle measurements with nanoparticle tracking analysis - An accuracy and repeatability comparison between NanoSight NS300 and ZetaView." *J Extracell Vesicles* 8 (1):1596016. doi: 10.1080/20013078.2019.1596016.
- Biller, S. J., F. Schubotz, S. E. Roggensack, A. W. Thompson, R. E. Summons, and S. W. Chisholm. 2014. "Bacterial vesicles in marine ecosystems." *Science* 343 (6167):183-6. doi: 10.1126/science.1243457.
- Bitto, N. J., R. Chapman, S. Pidot, A. Costin, C. Lo, J. Choi, T. D'Cruze, E. C. Reynolds, S. G. Dashper, L. Turnbull, C. B. Whitchurch, T. P. Stinear, K. J. Stacey, and R. L. Ferrero. 2017. "Bacterial membrane vesicles transport their DNA cargo into host cells." *Sci Rep* 7 (1):7072. doi: 10.1038/s41598-017-07288-4.
- Boccaro, M., Y. Fedala, C. V. Bryan, M. Bailly-Bechet, C. Bowler, and A. C. Boccaro. 2016. "Full-field interferometry for counting and differentiating aquatic biotic nanoparticles: from laboratory to Tara Oceans." *Biomed Opt Express* 7 (9):3736-3746. doi: 10.1364/BOE.7.003736.
- Bomberger, J. M., D. P. Maceachran, B. A. Coutermarsh, S. Ye, G. A. O'Toole, and B. A. Stanton. 2009. "Long-distance delivery of bacterial virulence factors by *Pseudomonas aeruginosa* outer membrane vesicles." *PLoS Pathog* 5 (4):e1000382. doi: 10.1371/journal.ppat.1000382.
- Chandler, W. L., W. Yeung, and J. F. Tait. 2011. "A new microparticle size calibration standard for use in measuring smaller microparticles using a new flow cytometer." *J Thromb Haemost* 9 (6):1216-24. doi: 10.1111/j.1538-7836.2011.04283.x.
- Cimorelli, M., R. Nieuwland, Z. Varga, and E. van der Pol. 2021. "Standardized procedure to measure the size distribution of extracellular vesicles together with other particles in biofluids with microfluidic resistive pulse sensing." *PLoS One* 16 (4):e0249603. doi: 10.1371/journal.pone.0249603.
- Codemo, M., S. Muschiol, F. Iovino, P. Nannapaneni, L. Plant, S. N. Wai, and B. Henriques-Normark. 2018. "Immunomodulatory Effects of Pneumococcal Extracellular Vesicles on Cellular and Humoral Host Defenses." *mBio* 9 (2). doi: 10.1128/mBio.00559-18.
- Cornuault, J. K., M. A. Petit, M. Mariadassou, L. Benevides, E. Moncaut, P. Langella, H. Sokol, and M. De Paepe. 2018. "Phages infecting *Faecalibacterium prausnitzii* belong to novel viral genera that help to decipher intestinal viromes." *Microbiome* 6 (1):65. doi: 10.1186/s40168-018-0452-1.
- Dehghani, M., R. K. Montange, M. W. Olszowy, and D. Pollard. 2021. "An Emerging Fluorescence-Based Technique for Quantification and Protein Profiling of Extracellular Vesicles." *SLAS Technol* 26 (2):189-199. doi: 10.1177/2472630320970458.
- Diard, M., E. Bakkeren, J. K. Cornuault, K. Moor, A. Hausmann, M. E. Sellin, C. Loverdo, A. Aertsen, M. Ackermann, M. De Paepe, E. Slack, and W. D. Hardt. 2017. "Inflammation boosts bacteriophage transfer between *Salmonella* spp." *Science* 355 (6330):1211-1215. doi: 10.1126/science.aaf8451.
- Doyle, L. M., and M. Z. Wang. 2019. "Overview of Extracellular Vesicles, Their Origin, Composition, Purpose, and Methods for Exosome Isolation and Analysis." *Cells* 8 (7). doi: 10.3390/cells8070727.
- Dragovic, R. A., C. Gardiner, A. S. Brooks, D. S. Tannetta, D. J. Ferguson, P. Hole, B. Carr, C. W. Redman, A. L. Harris, P. J. Dobson, P. Harrison, and I. L. Sargent. 2011. "Sizing and phenotyping of cellular vesicles using Nanoparticle Tracking Analysis." *Nanomedicine* 7 (6):780-8. doi: 10.1016/j.nano.2011.04.003.
- Dugat-Bony, E., J. Lossouarn, M. De Paepe, A. S. Sarthou, Y. Fedala, M. A. Petit, and S. Chaillou. 2020. "Viral metagenomic analysis of the cheese surface: A comparative study of rapid

- procedures for extracting viral particles." *Food Microbiol* 85:103278. doi: 10.1016/j.fm.2019.103278.
- Filip, R. 2021. "An Update on the Role of Extracellular Vesicles in the Pathogenesis of Necrotizing Enterocolitis and Inflammatory Bowel Diseases." *Cells* 10 (11). doi: 10.3390/cells10113202.
- Filipe, V., A. Hawe, and W. Jiskoot. 2010. "Critical evaluation of Nanoparticle Tracking Analysis (NTA) by NanoSight for the measurement of nanoparticles and protein aggregates." *Pharm Res* 27 (5):796-810. doi: 10.1007/s11095-010-0073-2.
- Furuta, N., H. Takeuchi, and A. Amano. 2009. "Entry of Porphyromonas gingivalis outer membrane vesicles into epithelial cells causes cellular functional impairment." *Infect Immun* 77 (11):4761-70. doi: 10.1128/IAI.00841-09.
- Gardiner, C., Y. J. Ferreira, R. A. Dragovic, C. W. Redman, and I. L. Sargent. 2013. "Extracellular vesicle sizing and enumeration by nanoparticle tracking analysis." *J Extracell Vesicles* 2. doi: 10.3402/jev.v2i0.19671.
- Heider, S., and C. Metzner. 2014. "Quantitative real-time single particle analysis of virions." *Virology* 462-463:199-206. doi: 10.1016/j.virol.2014.06.005.
- Holzwarth, G., D. G. Gordon, J. E. McGinness, B. P. Dorman, and M. F. Maestre. 1974. "Mie scattering contributions to the optical density and circular dichroism of T2 bacteriophage." *Biochemistry* 13 (1):126-32. doi: 10.1021/bi00698a020.
- Huang, A. S., and D. Baltimore. 1970. "Defective viral particles and viral disease processes." *Nature* 226 (5243):325-7. doi: 10.1038/226325a0.
- Hulst, H. C. van de. 1957. *Light Scattering by Small Metal Particles*. NYC.
- Jensen, O. A., J. U. Prause, and H. Laursen. 1981. "Shrinkage in preparatory steps for SEM. A study on rabbit corneal endothelium." *Albrecht Von Graefes Arch Klin Exp Ophthalmol* 215 (4):233-42. doi: 10.1007/BF00407662.
- Kaletta, J., C. Pickl, C. Griebler, A. Klingl, R. Kurmayer, and L. Deng. 2020. "A rigorous assessment and comparison of enumeration methods for environmental viruses." *Sci Rep* 10 (1):18625. doi: 10.1038/s41598-020-75490-y.
- Kestens, V., V. Bozatzidis, P. J. De Temmerman, Y. Ramaye, and G. Roebben. 2017. "Validation of a particle tracking analysis method for the size determination of nano- and microparticles." *J Nanopart Res* 19 (8):271. doi: 10.1007/s11051-017-3966-8.
- Lane, R. E., D. Korbie, M. M. Hill, and M. Trau. 2018. "Extracellular vesicles as circulating cancer biomarkers: opportunities and challenges." *Clin Transl Med* 7 (1):14. doi: 10.1186/s40169-018-0192-7.
- Lippe, R. 2018. "Flow Virometry: a Powerful Tool To Functionally Characterize Viruses." *J Virol* 92 (3). doi: 10.1128/JVI.01765-17.
- Luz, Bsrđ, A. Nicolas, S. Chabelskaya, V. R. Rodovalho, Y. Le Loir, V. A. C. Azevedo, B. Felden, and E. Guedon. 2021. "Environmental Plasticity of the RNA Content of Staphylococcus aureus Extracellular Vesicles." *Front Microbiol* 12:634226. doi: 10.3389/fmicb.2021.634226.
- Maas, S. L., J. De Vrij, and M. L. Broekman. 2014. "Quantification and size-profiling of extracellular vesicles using tunable resistive pulse sensing." *J Vis Exp* (92):e51623. doi: 10.3791/51623.
- Maguire, Ciarán M., Katherine Sillence, Matthias Roesslein, Claire Hannell, Guillaume Suarez, Jean-Jacques Sauvain, Sonja Capracotta, Servane Contal, Sebastien Cambier, Naouale El Yamani, Maria Dusinska, Agnieszka Dybowska, Antje Vennemann, Laura Cooke, Andrea Haase, Andreas Luch, Martin Wiemann, Arno Gutleb, Rafi Korenstein, Michael Riediker, Peter Wick, Patrick Hole, and Adriele Prina-Mello. 2017. "Benchmark of Nanoparticle Tracking Analysis on Measuring Nanoparticle Sizing and Concentration." *Journal of Micro and Nano-Manufacturing* 5 (4). doi: 10.1115/1.4037124.
- Mateescu, B., E. J. Kowal, B. W. van Balkom, S. Bartel, S. N. Bhattacharyya, E. I. Buzas, A. H. Buck, P. de Candia, F. W. Chow, S. Das, T. A. Driedonks, L. Fernandez-Messina, F. Haderk, A. F. Hill, J. C. Jones, K. R. Van Keuren-Jensen, C. P. Lai, C. Lasser, I. D. Liegro, T. R. Lunavat, M. J. Lorenowicz, S. L. Maas, I. Mager, M. Mittelbrunn, S. Momma, K. Mukherjee, M. Nawaz, D. M. Pegtel, M. W. Pfaffl, R. M. Schiffelers, H. Tahara, C. Thery, J. P. Tosar,

- M. H. Wauben, K. W. Witwer, and E. N. Nolte-'t Hoen. 2017. "Obstacles and opportunities in the functional analysis of extracellular vesicle RNA - an ISEV position paper." *J Extracell Vesicles* 6 (1):1286095. doi: 10.1080/20013078.2017.1286095.
- Momen-Heravi, F., L. Balaj, S. Alian, J. Tigges, V. Toxavidis, M. Ericsson, R. J. Distel, A. R. Ivanov, J. Skog, and W. P. Kuo. 2012. "Alternative methods for characterization of extracellular vesicles." *Front Physiol* 3:354. doi: 10.3389/fphys.2012.00354.
- Norman, J. M., S. A. Handley, M. T. Baldrige, L. Droit, C. Y. Liu, B. C. Keller, A. Kambal, C. L. Monaco, G. Zhao, P. Fleshner, T. S. Stappenbeck, D. P. McGovern, A. Keshavarzian, E. A. Mutlu, J. Sauk, D. Gevers, R. J. Xavier, D. Wang, M. Parkes, and H. W. Virgin. 2015. "Disease-specific alterations in the enteric virome in inflammatory bowel disease." *Cell* 160 (3):447-60. doi: 10.1016/j.cell.2015.01.002.
- Ortega Arroyo, J., D. Cole, and P. Kukura. 2016. "Interferometric scattering microscopy and its combination with single-molecule fluorescence imaging." *Nat Protoc* 11 (4):617-33. doi: 10.1038/nprot.2016.022.
- Pang, Y., H. Song, and W. Cheng. 2016. "Using optical trap to measure the refractive index of a single animal virus in culture fluid with high precision." *Biomed Opt Express* 7 (5):1672-89. doi: 10.1364/BOE.7.001672.
- Ratajczak, J., M. Wysoczynski, F. Hayek, A. Janowska-Wieczorek, and M. Z. Ratajczak. 2006. "Membrane-derived microvesicles: important and underappreciated mediators of cell-to-cell communication." *Leukemia* 20 (9):1487-95. doi: 10.1038/sj.leu.2404296.
- Robbins, P. D., A. Dorronsoro, and C. N. Booker. 2016. "Regulation of chronic inflammatory and immune processes by extracellular vesicles." *J Clin Invest* 126 (4):1173-80. doi: 10.1172/JCI81131.
- Rodvalho, V. R., B. S. R. da Luz, H. Rabah, F. L. R. do Carmo, E. L. Folador, A. Nicolas, J. Jardin, V. Briard-Bion, H. Blottiere, N. Lapaque, G. Jan, Y. Le Loir, V. A. de Carvalho Azevedo, and E. Guedon. 2020. "Extracellular Vesicles Produced by the Probiotic *Propionibacterium freudenreichii* CIRM-BIA 129 Mitigate Inflammation by Modulating the NF-kappaB Pathway." *Front Microbiol* 11:1544. doi: 10.3389/fmicb.2020.01544.
- Roose-Amsaleg, C., Y. Fedala, C. Venien-Bryan, J. Garnier, A. C. Boccara, and M. Boccara. 2017. "Utilization of interferometric light microscopy for the rapid analysis of virus abundance in a river." *Res Microbiol* 168 (5):413-418. doi: 10.1016/j.resmic.2017.02.004.
- Rossi, C. A., B. J. Kearney, S. P. Olschner, P. L. Williams, C. G. Robinson, M. L. Heinrich, A. M. Zovanyi, M. F. Ingram, D. A. Norwood, and R. J. Schoepp. 2015. "Evaluation of ViroCyt(R) Virus Counter for rapid filovirus quantitation." *Viruses* 7 (3):857-72. doi: 10.3390/v7030857.
- Sabbagh, Q., G. Andre-Gregoire, C. Alves-Nicolau, A. Dupont, N. Bidere, E. Jouglar, L. Guevel, J. S. Frenel, and J. Gavard. 2021. "The von Willebrand factor stamps plasmatic extracellular vesicles from glioblastoma patients." *Sci Rep* 11 (1):22792. doi: 10.1038/s41598-021-02254-7.
- Sausset, R., M. A. Petit, V. Gaboriau-Routhiau, and M. De Paepe. 2020. "New insights into intestinal phages." *Mucosal Immunol* 13 (2):205-215. doi: 10.1038/s41385-019-0250-5.
- Shkoporov, A. N., E. V. Khokhlova, C. B. Fitzgerald, S. R. Stockdale, L. A. Draper, R. P. Ross, and C. Hill. 2018. "PhiCrAss001 represents the most abundant bacteriophage family in the human gut and infects *Bacteroides intestinalis*." *Nat Commun* 9 (1):4781. doi: 10.1038/s41467-018-07225-7.
- Simons, M., and G. Raposo. 2009. "Exosomes--vesicular carriers for intercellular communication." *Curr Opin Cell Biol* 21 (4):575-81. doi: 10.1016/j.ceb.2009.03.007.
- Slomka, A., S. K. Urban, V. Lukacs-Kornek, E. Zekanowska, and M. Kornek. 2018. "Large Extracellular Vesicles: Have We Found the Holy Grail of Inflammation?" *Front Immunol* 9:2723. doi: 10.3389/fimmu.2018.02723.
- Szatanek, R., M. Baj-Krzyworzeka, J. Zimoch, M. Lekka, M. Siedlar, and J. Baran. 2017. "The Methods of Choice for Extracellular Vesicles (EVs) Characterization." *Int J Mol Sci* 18 (6). doi: 10.3390/ijms18061153.

- Thery, C., S. Amigorena, G. Raposo, and A. Clayton. 2006. "Isolation and characterization of exosomes from cell culture supernatants and biological fluids." *Curr Protoc Cell Biol* Chapter 3:Unit 3 22. doi: 10.1002/0471143030.cb0322s30.
- Thery, C., M. Ostrowski, and E. Segura. 2009. "Membrane vesicles as conveyors of immune responses." *Nat Rev Immunol* 9 (8):581-93. doi: 10.1038/nri2567.
- Thery, C., K. W. Witwer, E. Aikawa, M. J. Alcaraz, J. D. Anderson, R. Andriantsitohaina, A. Antoniou, T. Arab, F. Archer, G. K. Atkin-Smith, D. C. Ayre, J. M. Bach, D. Bachurski, H. Baharvand, L. Balaj, S. Baldacchino, N. N. Bauer, A. A. Baxter, M. Bebawy, C. Beckham, A. Bedina Zavec, A. Benmoussa, A. C. Berardi, P. Bergese, E. Bielska, C. Blenkiron, S. Bobis-Wozowicz, E. Boilard, W. Boireau, A. Bongiovanni, F. E. Borrás, S. Bosch, C. M. Boulanger, X. Breakefield, A. M. Breglio, M. A. Brennan, D. R. Brigstock, A. Brisson, M. L. Broekman, J. F. Bromberg, P. Bryl-Gorecka, S. Buch, A. H. Buck, D. Burger, S. Busatto, D. Buschmann, B. Bussolati, E. I. Buzas, J. B. Byrd, G. Camussi, D. R. Carter, S. Caruso, L. W. Chamley, Y. T. Chang, C. Chen, S. Chen, L. Cheng, A. R. Chin, A. Clayton, S. P. Clerici, A. Cocks, E. Cocucci, R. J. Coffey, A. Cordeiro-da-Silva, Y. Couch, F. A. Coumans, B. Coyle, R. Crescitelli, M. F. Criado, C. D'Souza-Schorey, S. Das, A. Datta Chaudhuri, P. de Candia, E. F. De Santana, O. De Wever, H. A. Del Portillo, T. Demaret, S. Deville, A. Devitt, B. Dhondt, D. Di Vizio, L. C. Dieterich, V. Dolo, A. P. Dominguez Rubio, M. Dominici, M. R. Dourado, T. A. Driedonks, F. V. Duarte, H. M. Duncan, R. M. Eichenberger, K. Ekstrom, S. El Andaloussi, C. Elie-Caille, U. Erdbrugger, J. M. Falcon-Perez, F. Fatima, J. E. Fish, M. Flores-Bellver, A. Forsonits, A. Frelet-Barrand, F. Fricke, G. Fuhrmann, S. Gabrielsson, A. Gamez-Valero, C. Gardiner, K. Gartner, R. Gaudin, Y. S. Gho, B. Giebel, C. Gilbert, M. Gimona, I. Giusti, D. C. Goberdhan, A. Gorgens, S. M. Gorski, D. W. Greening, J. C. Gross, A. Gualerzi, G. N. Gupta, D. Gustafson, A. Handberg, R. A. Haraszti, P. Harrison, H. Hegyesi, A. Hendrix, A. F. Hill, F. H. Hochberg, K. F. Hoffmann, B. Holder, H. Holthofer, B. Hosseinkhani, G. Hu, Y. Huang, V. Huber, S. Hunt, A. G. Ibrahim, T. Ikezu, J. M. Inal, M. Isin, A. Ivanova, H. K. Jackson, S. Jacobsen, S. M. Jay, M. Jayachandran, G. Jenster, L. Jiang, S. M. Johnson, J. C. Jones, A. Jong, T. Jovanovic-Talisman, S. Jung, R. Kalluri, S. I. Kano, S. Kaur, Y. Kawamura, E. T. Keller, D. Khamari, E. Khomyakova, A. Khvorova, P. Kierulf, K. P. Kim, T. Kislinger, M. Klingeborn, D. J. Klinke, 2nd, M. Kornek, M. M. Kosanovic, A. F. Kovacs, E. M. Kramer-Albers, S. Krasemann, M. Krause, I. V. Kurochkin, G. D. Kusuma, S. Kuypers, S. Laitinen, S. M. Langevin, L. R. Languino, J. Lannigan, C. Lasser, L. C. Laurent, G. Lavieu, E. Lazaro-Ibanez, S. Le Lay, M. S. Lee, Y. X. F. Lee, D. S. Lemos, M. Lenassi, A. Leszczynska, I. T. Li, K. Liao, S. F. Libregts, E. Ligeti, R. Lim, S. K. Lim, A. Line, K. Linnemannstons, A. Llorente, C. A. Lombard, M. J. Lorenowicz, A. M. Lorincz, J. Lotvall, J. Lovett, M. C. Lowry, X. Loyer, Q. Lu, B. Lukomska, T. R. Lunavat, S. L. Maas, H. Malhi, A. Marcilla, J. Mariani, J. Mariscal, E. S. Martens-Uzunova, L. Martin-Jaular, M. C. Martinez, V. R. Martins, M. Mathieu, S. Mathivanan, M. Maugeri, L. K. McGinnis, M. J. McVey, D. G. Meckes, Jr., K. L. Meehan, I. Mertens, V. R. Minciocchi, A. Moller, M. Moller Jorgensen, A. Morales-Kastresana, J. Morhayim, F. Mullier, M. Muraca, L. Musante, V. Mussack, D. C. Muth, K. H. Myburgh, T. Najrana, M. Nawaz, I. Nazarenko, P. Nejsum, C. Neri, T. Neri, R. Nieuwland, L. Nimrichter, J. P. Nolan, E. N. Nolte-'t Hoen, N. Noren Hooten, L. O'Driscoll, T. O'Grady, A. O'Loughlen, T. Ochiya, M. Olivier, A. Ortiz, L. A. Ortiz, X. Osteikoetxea, O. Ostergaard, M. Ostrowski, J. Park, D. M. Pegtel, H. Peinado, F. Perut, M. W. Pfaffl, D. G. Phinney, B. C. Pieters, R. C. Pink, D. S. Pisetsky, E. Pogge von Strandmann, I. Polakovicova, I. K. Poon, B. H. Powell, I. Prada, L. Pulliam, P. Quesenberry, A. Radeghieri, R. L. Raffai, S. Raimondo, J. Rak, M. I. Ramirez, G. Raposo, M. S. Rayyan, N. Regev-Rudzki, F. L. Ricklefs, P. D. Robbins, D. D. Roberts, S. C. Rodrigues, E. Rohde, S. Rome, K. M. Rouschop, A. Rughetti, A. E. Russell, P. Saa, S. Sahoo, E. Salas-Huenuleo, C. Sanchez, J. A. Saugstad, M. J. Saul, R. M. Schifflers, R. Schneider, T. H. Schoyen, A. Scott, E. Shahaj, S. Sharma, O. Shatnyeva, F. Shekari, G. V. Shelke, A. K. Shetty, K. Shiba, P. R. Siljander, A. M. Silva, A. Skowronek, O. L. Snyder, 2nd, R. P. Soares, B. W. Sodar, C. Soekmadji, J. Sotillo, P. D. Stahl, W. Stoorvogel, S. L. Stott, E. F. Strasser, S. Swift, H. Tahara, M. Tewari,

- K. Timms, S. Tiwari, R. Tixeira, M. Tkach, W. S. Toh, R. Tomasini, A. C. Torrecilhas, J. P. Tosar, V. Toxavidis, L. Urbanelli, P. Vader, B. W. van Balkom, S. G. van der Grein, J. Van Deun, M. J. van Herwijnen, K. Van Keuren-Jensen, G. van Niel, M. E. van Royen, A. J. van Wijnen, M. H. Vasconcelos, I. J. Vechetti, Jr., T. D. Veit, L. J. Vella, E. Velot, F. J. Verweij, B. Vestad, J. L. Vinas, T. Visnovitz, K. V. Vukman, J. Wahlgren, D. C. Watson, M. H. Wauben, A. Weaver, J. P. Webber, V. Weber, A. M. Wehman, D. J. Weiss, J. A. Welsh, S. Wendt, A. M. Wheelock, Z. Wiener, L. Witte, J. Wolfram, A. Xagorari, P. Xander, J. Xu, X. Yan, M. Yanez-Mo, H. Yin, Y. Yuana, V. Zappulli, J. Zarubova, V. Zekas, J. Y. Zhang, Z. Zhao, L. Zheng, A. R. Zheutlin, A. M. Zickler, P. Zimmermann, A. M. Zivkovic, D. Zocco, and E. K. Zuba-Surma. 2018. "Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines." *J Extracell Vesicles* 7 (1):1535750. doi: 10.1080/20013078.2018.1535750.
- Thompson, A. G., E. Gray, S. M. Heman-Ackah, I. Mager, K. Talbot, S. E. Andaloussi, M. J. Wood, and M. R. Turner. 2016. "Extracellular vesicles in neurodegenerative disease - pathogenesis to biomarkers." *Nat Rev Neurol* 12 (6):346-57. doi: 10.1038/nrneurol.2016.68.
- Turkki, V., E. Alppila, S. Yla-Herttuala, and H. P. Lesch. 2021. "Experimental Evaluation of an Interferometric Light Microscopy Particle Counter for Titering and Characterization of Virus Preparations." *Viruses* 13 (5). doi: 10.3390/v13050939.
- Valadi, H., K. Ekstrom, A. Bossios, M. Sjostrand, J. J. Lee, and J. O. Lotvall. 2007. "Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells." *Nat Cell Biol* 9 (6):654-9. doi: 10.1038/ncb1596.
- Vallabhaneni, K. C., P. Penfornis, S. Dhule, F. Guillonneau, K. V. Adams, Y. Y. Mo, R. Xu, Y. Liu, K. Watabe, M. C. Vemuri, and R. Pochampally. 2015. "Extracellular vesicles from bone marrow mesenchymal stem/stromal cells transport tumor regulatory microRNA, proteins, and metabolites." *Oncotarget* 6 (7):4953-67. doi: 10.18632/oncotarget.3211.
- van der Pol, E., A. N. Boing, P. Harrison, A. Sturk, and R. Nieuwland. 2012. "Classification, functions, and clinical relevance of extracellular vesicles." *Pharmacol Rev* 64 (3):676-705. doi: 10.1124/pr.112.005983.
- van der Pol, E., F. A. Coumans, A. E. Grootemaat, C. Gardiner, I. L. Sargent, P. Harrison, A. Sturk, T. G. van Leeuwen, and R. Nieuwland. 2014. "Particle size distribution of exosomes and microvesicles determined by transmission electron microscopy, flow cytometry, nanoparticle tracking analysis, and resistive pulse sensing." *J Thromb Haemost* 12 (7):1182-92. doi: 10.1111/jth.12602.
- van Niel, G., G. D'Angelo, and G. Raposo. 2018. "Shedding light on the cell biology of extracellular vesicles." *Nat Rev Mol Cell Biol* 19 (4):213-228. doi: 10.1038/nrm.2017.125.
- Zamora, J. L. R., and H. C. Aguilar. 2018. "Flow virometry as a tool to study viruses." *Methods* 134-135:87-97. doi: 10.1016/j.ymeth.2017.12.011.

	Videodrop	ZetaView	NanoSight
Technology	ILM	NTA	NTA
Light source	LED	Laser	Laser
Sample volume	6 μ L	1mL	1mL
Concentration range (particles/mL)	10 ⁸ – 10 ¹⁰	10 ⁵ - 10 ⁹	10 ⁶ - 10 ⁹
Sample introduction	Droplet in a tank	Syringe, tubings	Syringe, tubings
Acquisition time	20 sec	1 min	1 min
Temperature control	No	Yes	Yes
Fluorescence	No	Yes	Yes
Cost of consumables	~ 0	\$	\$
Cost of equipment	\$	\$\$	\$\$

Table 1. General characteristics of Nanosight, ZetaView, and Videodrop devices given by the manufacturers. Information was collated using manufacturer websites.

	T4	P1	T5	λ	SPP1	Φ crAss001	T7	PRD1	Φ X174
Family /Morphology	Myophage	Myophage	Siphophage	Siphophage	Siphophage	Podophage	Podophage	<i>Tectiviridae</i>	<i>Microviridae</i>
Capsid diameter (nm)	120 x 86	85	90	60	61	76	60	66	30
Genome type	dsDNA	dsDNA	dsDNA	dsDNA	dsDNA	dsDNA	dsDNA	dsDNA	ssDNA
Genome size (kb)	169	100	121	48	44	103	40	15	5

Table 2. Characteristics of the phages used in this study. Capsid diameters are from Viral Zone, except for SPP1 (Ignatiou 2019).

		T4	P1	T5	λ	SPP1	Φ crAss001	T7	PRD1	Φ X174
TEM	Capsid diameter (mean \pm SD)	110 \pm 6 x 85 \pm 7	87 \pm 7	79 \pm 5	64 \pm 5	57.5 \pm 1.6	90 \pm 4	63.7 \pm 4.1	65.3 \pm 4.2	26 \pm 3
	Tail length (mean \pm SD)	119 \pm 9	235 \pm 14	195 \pm 18	153 \pm 13	181 \pm 21	25 \pm 5	Very small	No tail	No tail
	Tail width (mean \pm SD)	16 \pm 2	18 \pm 3	9 \pm 2	9 \pm 2	9 \pm 2	Not measurable	Very small	No tail	No tail
NanoSight	Dh (mode)	176.8	100.5	127.9	87.9	65/95	94.6	63.9	71.3	-
	SD (mean of the distribution)	43.7	67.5	29.2	2.6	0.4	26.3	0.4	87.7	-
ZetaView	Dh (mode)	135.0	189.9	118.8	232.3	-	94.0	155.7	157.5	-
	SD (mean of the distribution)	47.4	72.2	87.6	102.1	-	33.7	67.7	83.1	-

Table 3. Phage dimensions measured by TEM, NTA and ILM, in nm. Modes, means and standard deviations are indicated.

	Videodrop	ZetaView	NanoSight
Calibration of the device before use	No	Yes (15 min)	No
Tolerates the presence of lipidic or proteic contaminants in samples	Yes	No	No
Tolerates the presence of detergents in samples	Yes	No	No
Linearity range on EVs (part/mL)	5.0×10^8 - 5.0×10^9	2.5×10^7 - 2.0×10^8	2.5×10^7 - 2.0×10^8
Number of tracked particles per acquisition under selected experimental conditions	200-1,000	1,000-1,500	1,000-10,000
Phage size threshold for absolute counting	100 nm	Between 60 and 90 nm	Between 30 and 60 nm
Phage size threshold for sizing	100 nm	100 nm	Between 30 and 60 nm
Number of samples that can be analyzed in 3 hours	50	15-20	5-10

Table 4. Evaluation summary of the Videodrop, the ZetaView and the NanoSight.

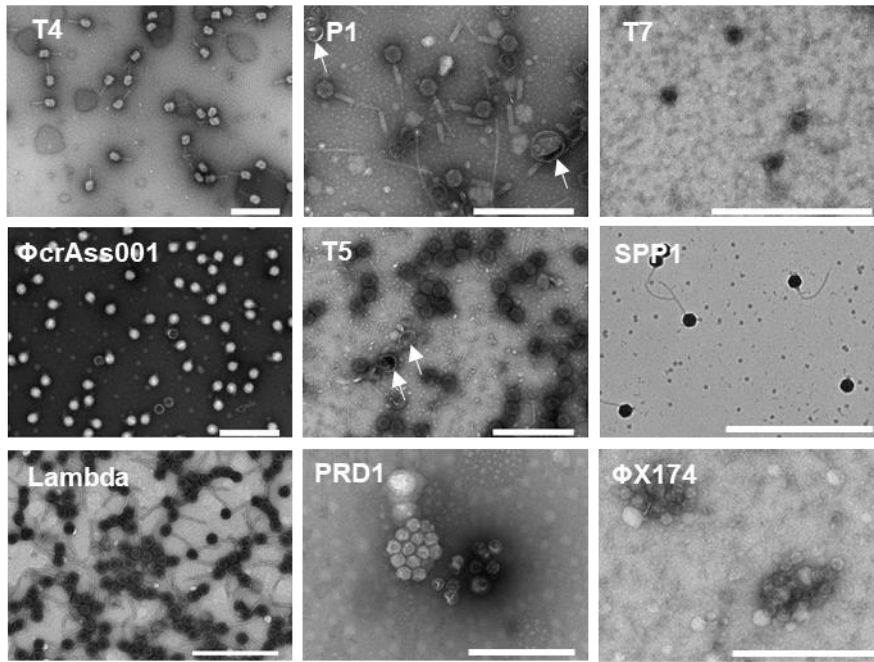


Figure 1. Quality control of phages used for the comparison of the Videodrop, the ZetaView and the NanoSight.

Representative TEM images of phage samples purified by an iodixanol gradient. White arrows point examples of EVs found in the P1 and T5 purified stocks. Scale bars are 500 nm.

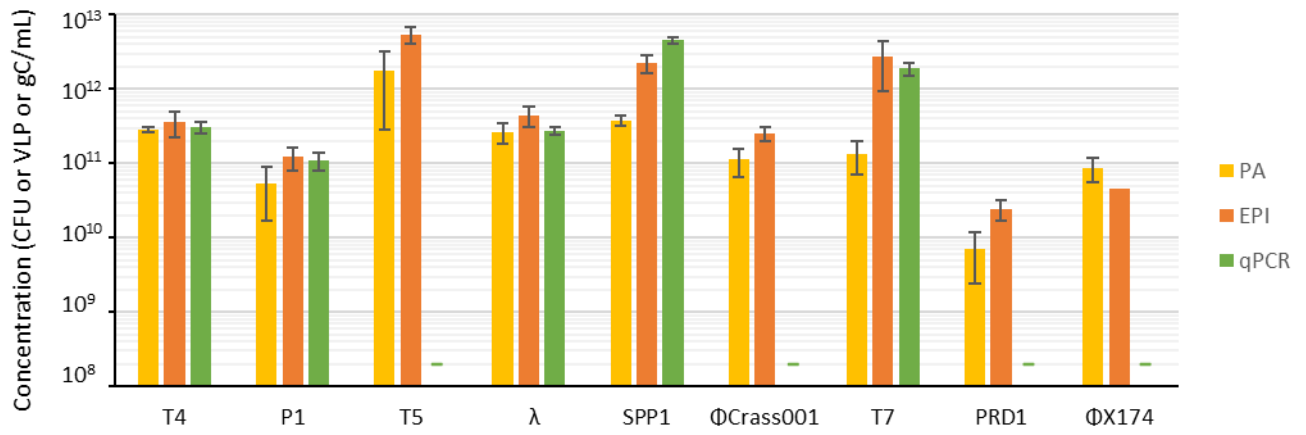


Figure 2. Quantification of phage preparations by standard techniques.

Phages concentrations obtained by Plaque Assay (PA, yellow), Epifluorescence Microscopy (EPI, orange), and quantitative PCR (qPCR, green). Error bars represent standard error of the mean on three independent experiments. Dashes (-) indicate that no measurement was done with this technique.

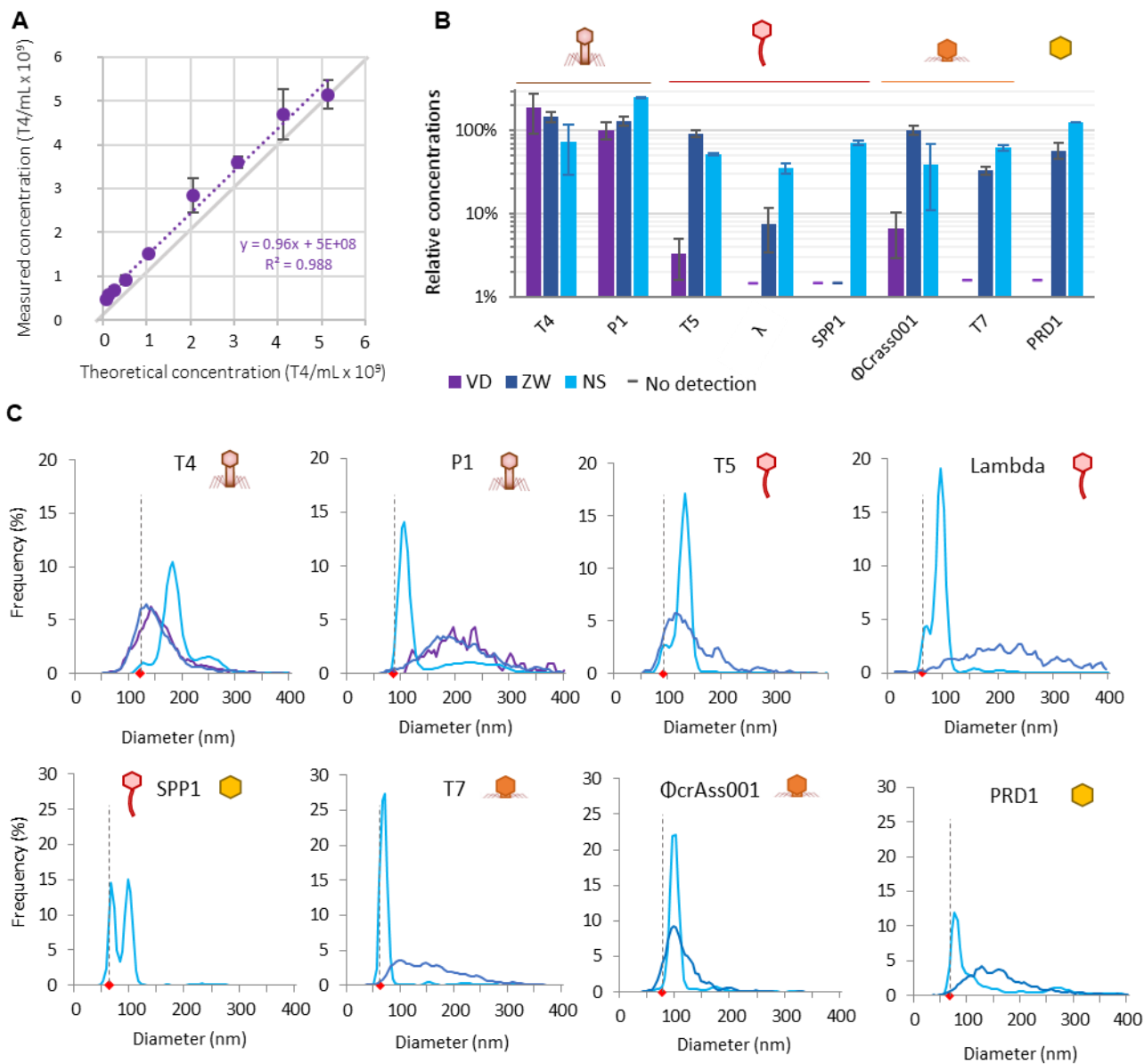


Figure 3. Comparison of ILM and NTA for the analysis of phages.

(A) Measurements on T4 serial dilutions to determine the appropriate range for concentration measurements. (B) Relative phage concentrations obtained with the Videodrop (purple), the ZetaView (dark blue) and the NanoSight (clear blue). Error bars represent standard error of the mean on three independent experiments. Dashes (-) indicate that the device provided no reliable measurement (<1% of EPI). (C) Virion size distributions in relative number of particles tracked for each technology. Colors are as in (B). Black dashed bars indicate the theoretical capsid diameter of phages, as presented in Table 2 .

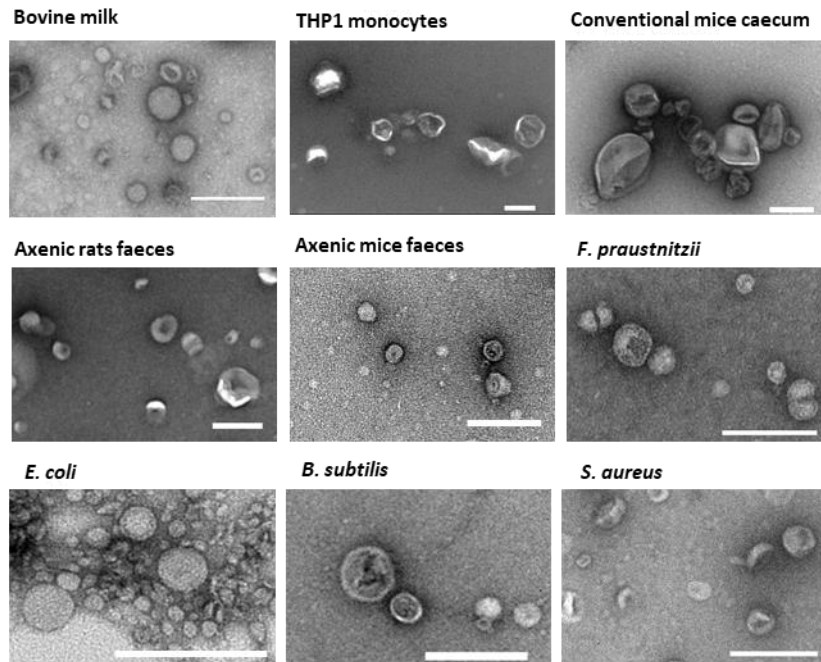


Figure 4. Quality control of EVs preparations used for the comparison of the Videodrop, the ZetaView and the NanoSight.

Representative TEM images of EVs preparations purified by an iodixanol gradient. Scale bars are 200 nm.

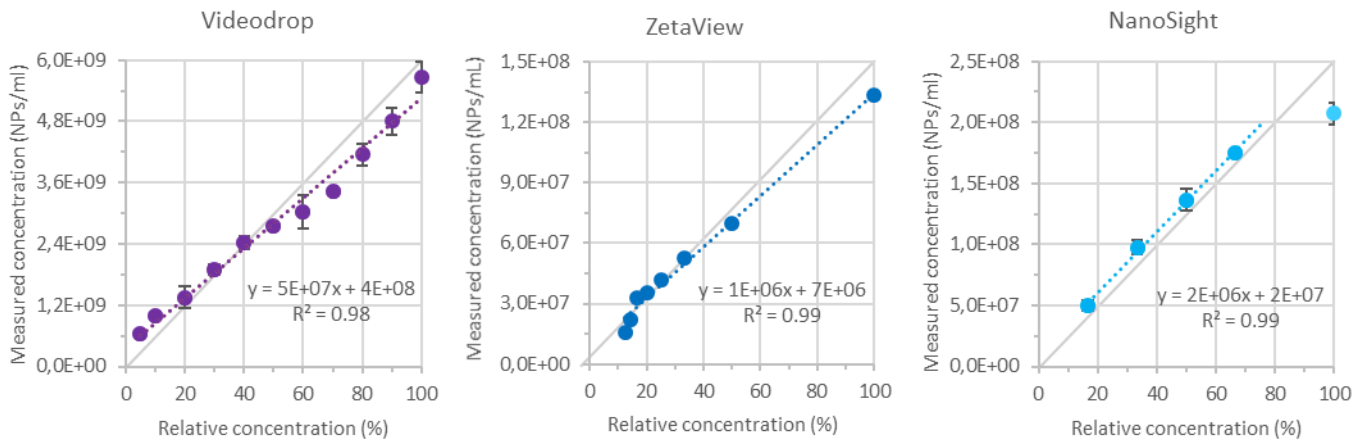


Figure 5. Linearity on concentration measurements on bovine milk-derived EVs.

Successive dilutions were realized in PBS, prior to measurement with Videodrop, NanoSight and ZetaView. Linearity range is slightly larger for the Videodrop (5×10^8 to 5×10^9 NPs/mL) than for the NanoSight and the ZetaView (3×10^7 to 1×10^8 NP/mL). Error bars represent standard error of the mean on three independent experiments.

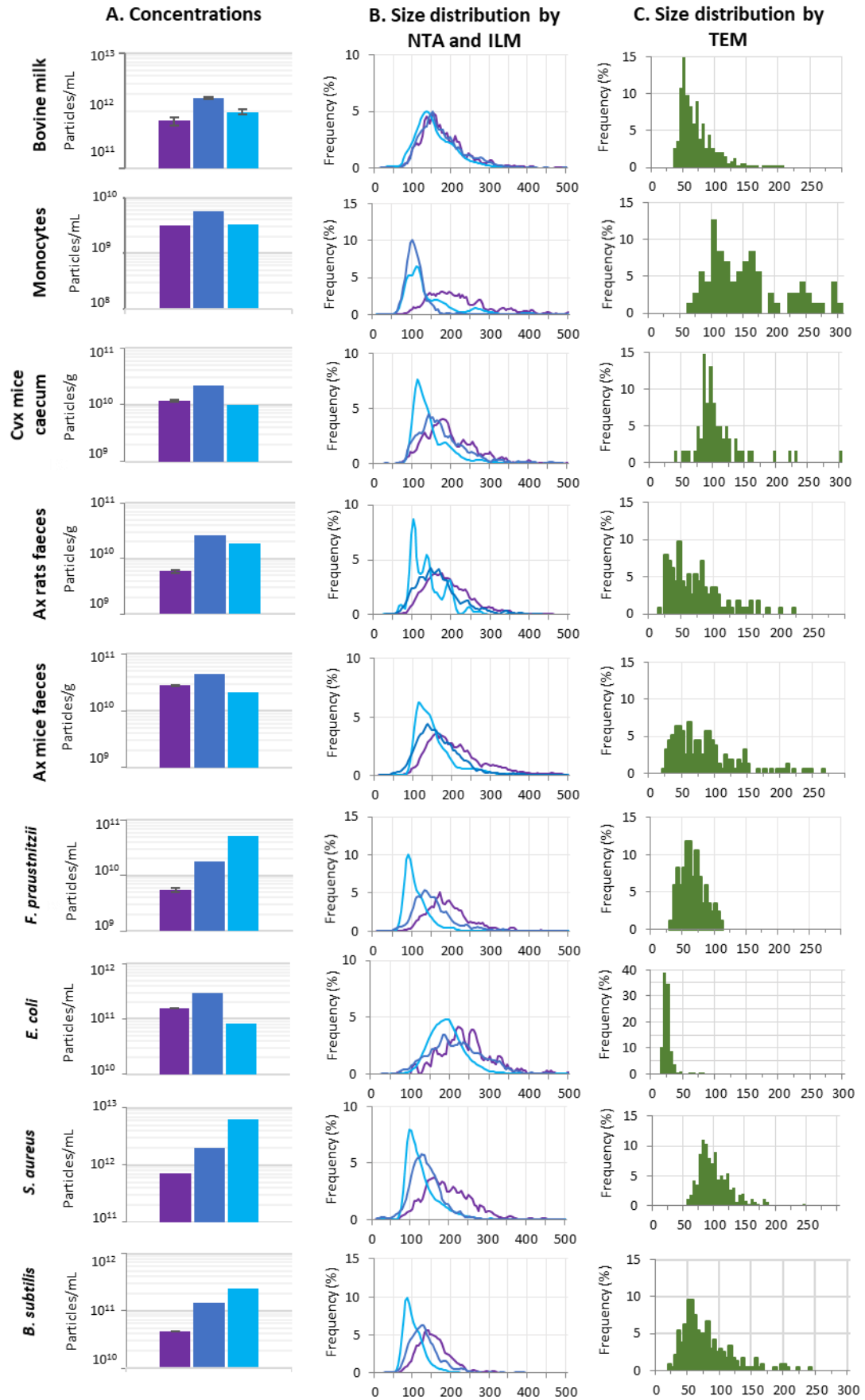


Figure 6. Determination of concentration and sizing of EVs. From the left to the right for each type of EVs: (A) Mean concentrations deduced from the Videodrop (purple), the ZetaView (dark blue) and the NanoSight (clear blue) measurements. (B) Size distributions from NTA and ILM. Colors are as in (A). (C) Size distributions of particles on from TEM image.

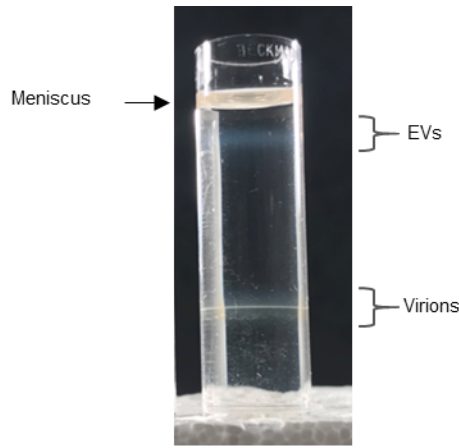


Figure S1. Picture of a representative iodixanol gradient

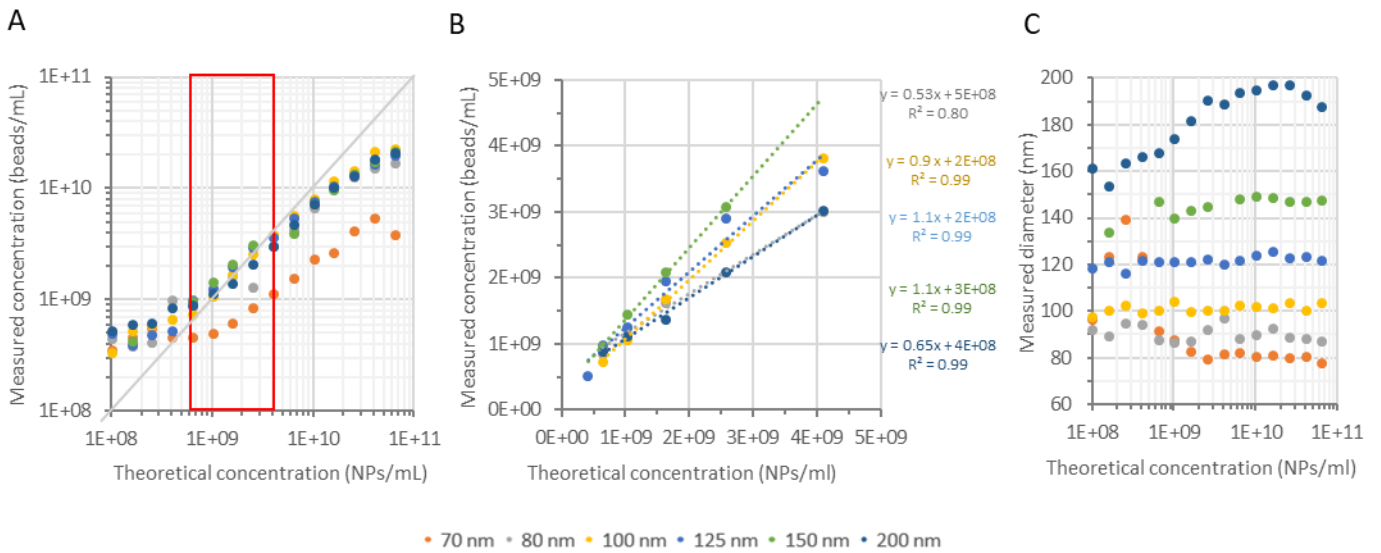


Figure S2. Concentrations and sizes of serial dilutions of polystyrene beads.

(A) Measured concentrations in function of theoretical concentrations from 10^{11} to 10^8 nanoparticles (NPs) per ml. The red rectangle delimits the concentration range in which a linear count consistent with the dilution was observed (5×10^8 to 4×10^9 beads/mL). (B) Measured concentrations in function of theoretical concentrations between 5×10^8 to 4×10^9 beads/mL. Equations and coefficient of determination of linear fits are indicated for each bead sizes. (C) Measured diameters of beads as a function of their concentration.

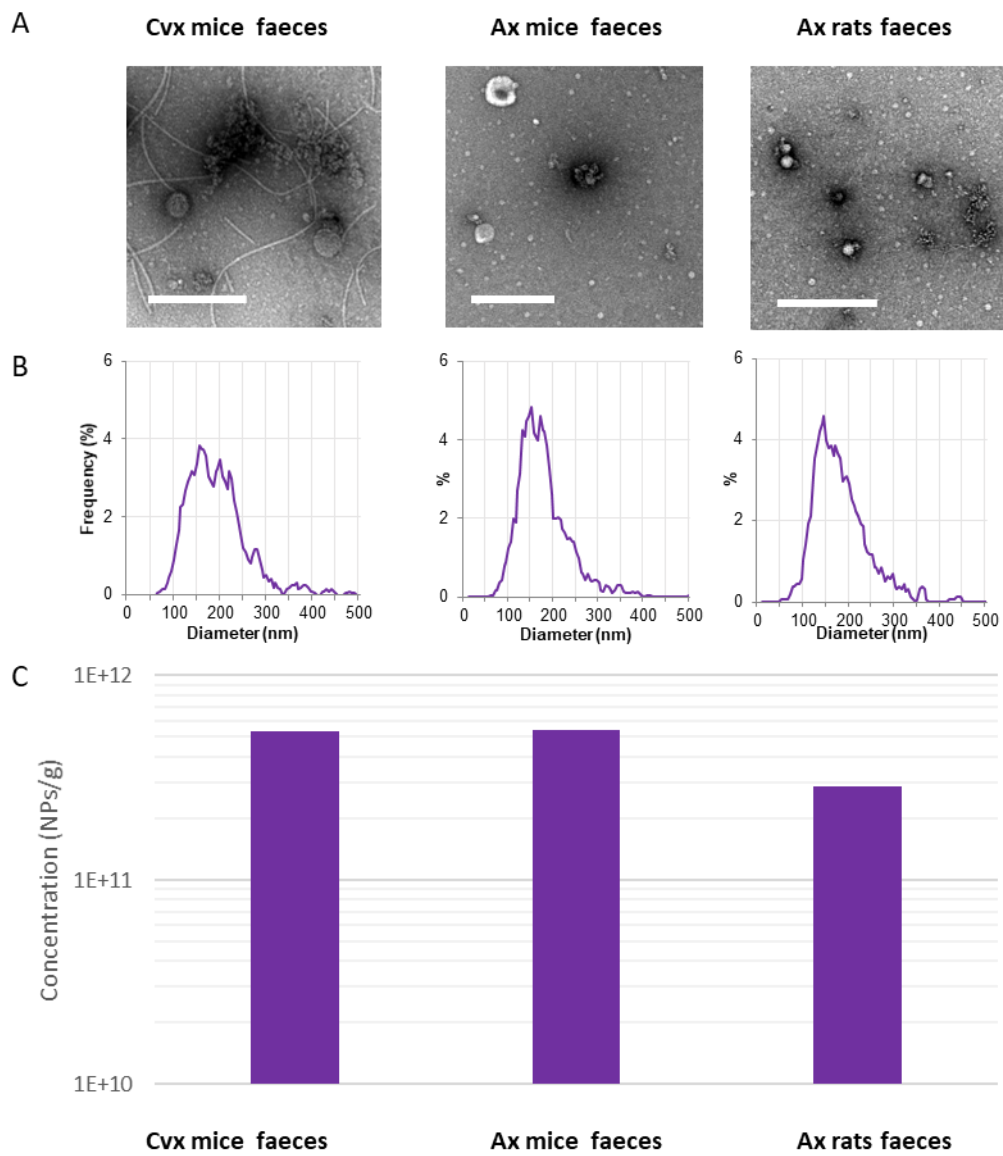


Figure S3. Determination of concentration and sizing of particles in fecal filtrates

(A) TEM fields of fecal filtrates. Scale bars are 500 nm. (B) Size distributions of particles from Videodrop. (C) Mean concentrations deduced from Videodrop measurements.

Phage	T4	P1	T5	λ	SPP1	φcrAss001	T7	PRD1	ΦX174
Bacterial host	<i>E. coli</i> MG1655 <i>hsdRM</i>	<i>E. coli</i> MG1655 <i>hsdRM</i>	<i>E. coli</i> MG1655 <i>hsdRM</i>	<i>E. coli</i> MG1655 <i>hsdRM</i>	<i>B. subtilis</i> YB886	<i>B. intestinalis</i> (DSM no: 108646)	<i>E. coli</i> MG1655 <i>hsdRM</i>	<i>E. coli</i> HMS174 pL2	<i>E. coli</i> CQ2
Bacterial host growth conditions	Aerobic, 37°C, LB + 5 mM MgSO ₄ + 5 mM CaCl ₂	Aerobic, 37°C, LB + 5 mM MgSO ₄ + 5 mM CaCl ₂	Aerobic, 37°C, LB + 5 mM MgSO ₄ + 5 mM CaCl ₂	Aerobic, 37°C, LB + 5 mM MgSO ₄ + 5 mM CaCl ₂ + 0,2% maltose	Aerobic, 37°C, LB + 5 mM MgSO ₄ + 5 mM CaCl ₂	Anaerobic, 37°C, FAB + 1 mM MgSO ₄ + 1 mM CaCl ₂	Aerobic, 37°C, LB + 5 mM MgSO ₄ + 5 mM CaCl ₂	Aerobic, 37°C, LB + 5 mM MgSO ₄ + 5 mM CaCl ₂ + kanamycin (25 µg/mL)	Aerobic, 30°C, LB + 5 mM MgSO ₄ + 5 mM CaCl ₂
m.o.i. used	0.25	0.25	0.1	0.1	0.01	1	0.25	5	5

Supp Table 1. Phages growth conditions used for this study

Phage	Primers	Reference
T4	F: ACTGGCCAGGTATTTCGCA R: ATGCTTCTTTAGCACCGGCA	(Klopot et al. 2017)
P1	F: AGGTTTCATCCGGTTGTGGTC R: TCATGCTGGGCTTGGATAGC	This article
λ	F: GAGTGC GGAAGATGCAAAGG R: TTAACAGTGC GTGACCAGG	(Fogg et al. 2010)
SPP1	F: CGGGCTGAAATACCTGTGGA R: TAGCCCCTCCTCCGATTGTT	(Labarde et al. 2021)
T7	F: CCCGAACTATAAGGCTAACC R: GCTCACGGATGCAATAGAAC	This article

Supp Table 2. qPCR primers used for this study

2. Dénombrement spécifique de phages et de VE

Bien que le Videodrop ne permette pas la détection de l'ensemble des phages présents dans le microbiote intestinal, il était prévu, au début de cette thèse, que le seuil de détection en taille soit suffisamment amélioré par Myriade pour permettre la détection et le dénombrement de tous les phages de plus de 30 nm (ce qui est théoriquement possible avec l'ILM).

Si le Videodrop permet la détection des VE (pourvu qu'elles soient suffisamment grandes pour être détectées), il ne permet pas de les discriminer des phages. Il a donc fallu réfléchir à des solutions permettant le dénombrement spécifique, avec pour objectif d'être par la suite capable de les énumérer dans des mélanges complexes comme les filtrats fécaux de la cohorte étudiée au cours de ce projet.

Deux approches différentes de distinction des phages et des VE ont été abordées : par microscopie par épifluorescence, après marquage spécifique des phages et des vésicules; et une seconde méthode consistant à détruire les VE à l'aide de traitements chimiques pour ne détecter que les objets restants, en théorie des phages.

2.1. Détection des VE et des phages par microscopie par épifluorescence

Afin d'énumérer spécifiquement les phages et les VE lorsqu'ils sont mélangés, l'intérêt de la microscopie par épifluorescence a été étudié. L'objectif était de trouver deux marqueurs, l'un spécifique des phages et l'autre des VE, utilisables conjointement pour la détection simultanée et la discrimination de ces deux populations.

2.1.1. Détection des phages

Pour le marquage et la détection des phages, des agents intercalants de l'ADN bien connus et communément utilisés pour la quantification absolue de phages ont été testés. Bien que certains phages soient à ADN simple brin ou à ARN, il a été rapporté que de tels marqueurs, malgré leur moindre efficacité sur les brins simples, permettent la détection de ces génomes en s'intercalant dans les structures quaternaires de ceux-ci (**Tableau I**). Plusieurs candidats ont été envisagés, à savoir le DAPI, le Sybr Gold, ainsi que le POPO-1. Ces trois intercalants de l'ADN diffèrent, notamment en termes de rendement quantique de fluorescence lorsqu'ils sont en présence d'ADN simple ou double brin, mais également en raison de leur spectre d'excitation et d'émission variés.

Agent intercalant de l'ADN	ADNdb	ADNsb
DAPI	0,34	0,19
SYBR Gold	0,66	0,55
POPO-1	0,60	ND

Tableau I. Rendements quantiques de fluorescence

Rendements quantiques de fluorescence (nombre de photons émis/nombre de photons absorbés) de différents agents intercalants de l'ADN lorsqu'ils sont complexés à de l'ADN double brin ou de l'ADN simple brin, à un ratio de une molécule de fluorophore pour 10 paires de bases, déterminés par G. Cosa et co-auteurs pour le DAPI et le Sybr Gold (Cosa et al. 2001) et par T. Netzl et co-auteurs pour le POPO-1 (Netzel et al. 1995). Les rendements quantiques de fluorescence sont calculés relativement à ceux de la quinine bisulfate pour le DAPI et de la fluorescéine pour le Sybr Gold et le POPO-1. ND = Non Déterminé.

L'avantage de tester différents marqueurs est d'avoir une marge de manœuvre plus importante concernant le choix des combinaisons de marqueurs des phages et des VE, ceux-ci devant posséder des longueurs d'onde d'émission et d'excitation différentes afin de pouvoir discriminer les deux populations (**Tableau II**). C'est pour cette raison que le DAPI a été testé. Malgré son rendement de fluorescence inférieur à celui du Sybr Gold, il présente l'avantage de posséder des longueurs d'onde d'excitation et d'émission différentes de celui-ci et de pouvoir être combiné à l'utilisation de la filipine III (**Tableau II**).

Objets ciblés	Marqueur	Longueur d'onde d'excitation (nm)	Longueur d'onde d'émission (nm)	Mécanisme d'action	Référence
Phages	DAPI	356	455	Intercalant de l'ADN	(Cosa et al. 2001)
	SYBR Gold	496	540	Intercalant de l'ADN	(Cosa et al. 2001)
	POPO-1	436	456	Intercalant de l'ADN	(Winter and Loeber 1997)
EVs	FM 4-64	558	734	Marqueur des lipides (membrane)	(Danielsen 2015)
	FM 1-43	488	580	Marqueur des lipides (membrane)	(Danielsen 2015)
	Filipine III	357	480	Marqueur du cholestérol	(Flesch et al. 2001)

Tableau II. Caractéristiques générales des fluorochromes utilisés

Puisque d'après certaines études près de 40% des phages du microbiote intestinal appartiennent à la famille des *Microviridae*, le marquage de Φ X174, un petit phage de 5 kpb à ADN simple brin, par le Sybr Gold, le DAPI et le POPO-1 a été testé. Le marquage par le Sybr Gold (Invitrogen) a été réalisé selon le protocole présenté dans l'article 1, à une concentration de 200X. Le même protocole a été utilisé pour le marquage de Φ X174 par le DAPI (Sigma-Aldrich), à une concentration de 1 μ g/mL, (comme réalisé par (Kenzaka, Tani, and Nasu 2010)). Il en va de même pour le POPO-1 (Invitrogen), utilisé à une concentration de 10 μ M (d'après (Hennes, Suttle, and Chan 1995)).

Si le Sybr Gold a permis la détection des virions de Φ X174 à des concentrations proches de celles attendues (comme présenté dans l'article 1 ; à savoir $4,55 \times 10^{10}$ VLPs/mL selon la microscopie par épifluorescence, contre $8,65 \times 10^{10}$ phages/mL selon le dénombrement par plages de lyse), ce ne fut pas le cas lors du marquage par le DAPI et le POPO-1, pour lesquels les tâches détectées sont trop diffuses pour être attribuées à des VLPs (**Figure 20**).

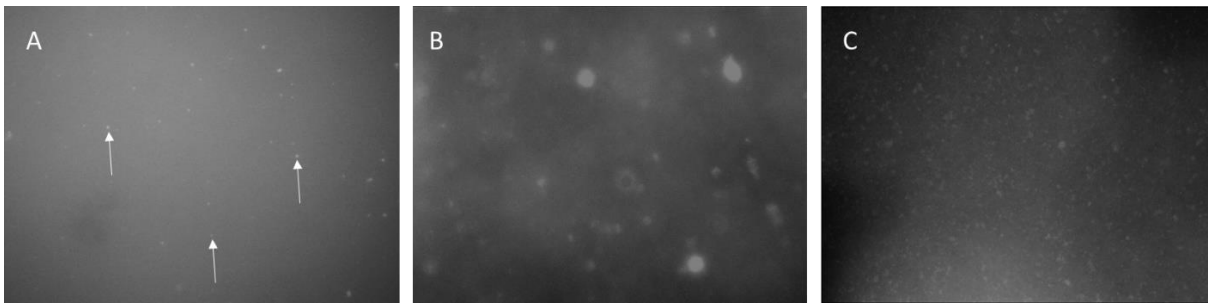


Figure 20. Φ X174 observés par microscopie par épifluorescence

A. Marquage par le Sybr Gold (200X). B. Marquage par le DAPI (1 μ g/mL). C. Marquage par le POPO-1 (10 μ M). Les flèches blanches indiquent la présence de VLPs.

Pour conclure, seul le marquage de Φ X174 par le Sybr Gold est possible, parmi les trois fluorochromes testés. Cela est probablement dû au meilleur rendement quantique de ce fluorochrome, combiné à une potentielle meilleure liaison aux doubles brins d'ADN. La proportion de *Microviridae* attendue étant importante au sein du microbiote intestinal, il fallait donc adapter le marquage des VE de façon à pouvoir le réaliser conjointement au marquage des phages par le Sybr Gold.

2.1.2. Détection des VE

Si le marquage et la détection de VE en épifluorescence suite à un marquage par des fluorochromes couplés à des anticorps spécifiques de certaines protéines est aujourd'hui répandu, peu d'articles traitent du marquage aspécifique de populations totales de vésicules après leur biogénèse. La recherche d'un tel marqueur a pourtant été menée, de façon à détecter les VE par la microscopie par épifluorescence et à les discriminer des phages marqués par le Sybr Gold.

Pour ce faire, des marqueurs des membranes cellulaires eucaryotes et bactériennes ont été testés. La majorité des VE étant produites à partir des membranes cellulaires internes ou externes, l'hypothèse était que ces marqueurs seraient capables d'également permettre la détection de VE, dont les compositions sont proches de celles de leurs cellules d'origine.

Deux marqueurs lipidiques de la membrane bactérienne couramment utilisés en microscopie, à savoir le FM 4-64 et le FM 1-43, ont été testés. J'ai également utilisé la filipine III, un marqueur du cholestérol (**Tableau II**). Le cholestérol étant présent dans les membranes eucaryotes mais absent des membranes bactériennes, la filipine III peut, en théorie, permettre la discrimination des VE selon leur origine eucaryote ou procaryote.

Ces différents marqueurs ont été testés sur les vésicules purifiées de fèces de rats axéniques. Ces dernières, uniquement d'origine eucaryote, sont susceptibles d'être marquées à la fois par les marqueurs de lipides et ceux du cholestérol.

Le marquage a été réalisé de la même manière que dans le protocole présenté dans l'article 1. Le FM 4-64 (Invitrogen) et le FM 1-43 (Thermo Fisher Scientific) ont été utilisés à une concentration de 5 µg/mL, selon les recommandations du fabricant, tandis que la filipine III (Sigma-Aldrich) a été utilisée à une concentration de 50 µg/mL (d'après (Wilhelm et al. 2019)).

Aucune vésicule n'a pu être détectée en épifluorescence par le biais de ces différents marquages par le FM 1-43, le FM 4-64 ou la filipine III. Une hypothèse concernant l'absence de signal de fluorescence après marquage par le FM 4-64 est que le filtre Anodisc utilisé pour piéger et observer les VE (et ainsi maîtriser la surface de dépôt pour permettre le calcul de leur concentration) autofluoresce légèrement dans les rouges, correspondant à la fenêtre d'émission FM 4-64, diminuant le ratio signal/bruit. Cependant, il est également probable que les interactions entre les fluorochromes et leurs cibles ne soient pas suffisantes au niveau des membranes vésiculaires, dont les surfaces sont bien inférieures à celles des membranes bactériennes, ce qui conduit à un très faible signal.

En conclusion, aucun des marqueurs testés n'a permis de détecter les VE, nécessitant de se tourner vers une autre approche permettant la détection différenciée des phages et des VE au sein d'un même échantillon.

2.2. Destruction spécifique des VE

Comme les phages et les VE ne peuvent pas être détectés simultanément grâce à la microscopie par épifluorescence, une autre approche basée sur le Videodrop et la destruction d'une des deux populations a été envisagée. Ainsi, bien que le Videodrop ne soit pas suffisamment sensible pour permettre la détection de toutes les espèces phagiques, nous espérons que sa sensibilité de détection soit améliorée de façon à pouvoir détecter la majorité des phages du microbiote intestinal. C'est dans cette optique que nous avons recherché un traitement permettant, dans des mélanges de phages et de VE, l'élimination des VE sans affecter l'intégrité de la capsidie des phages. Un tel traitement permettrait ainsi le dénombrement à la fois des phages et des VE, en réalisant une première acquisition par le Videodrop dans les mélanges, puis une seconde après l'application de ce traitement qui détruirait les VE uniquement. C'est dans ce but que les effets de six composés différents (deux solvants et quatre détergents) ont été testés : le chloroforme, le Vertrel

(1,1,1,2,3,4,4,5,5,5-decafluoropentane), le Sarkosyl (*N*-lauroylsarcosine), le Tween 20, le Triton X-100 et le Nonidet P-40 (NP-40).

2.2.1. Impact de la centrifugation

Deux préparations de phages purifiés (T4 et P1) et trois préparations de VE purifiées (issues de fèces de souris conventionnelles, de souris axéniques et de rats axéniques) (cf. 2.1.1. Détection des phages et 2.1.2. Détection des VE) ont été utilisées pour tester la sensibilité des phages et des VE aux traitements utilisés. Tous les protocoles testés et concentrations utilisées sont adaptés de publications (cf. Matériel et Méthodes).

Parfois utilisée dans des protocoles décrits dans la littérature, l'impact d'une étape de centrifugation à 12 000 g, pendant 4 minutes à 4°C a été investigué. L'avantage supposé d'une telle centrifugation serait de culotter les débris membranaires formés lors de la destruction des VE et de les séparer des phages restés en suspension dans le surnageant. En réalité, après récupération du surnageant et analyse de celui-ci par le Videodrop, la perte de 67% à 78% phages est observée au cours de cette étape (**Figure 21**), probablement par culottage. C'est pour cette raison que qu'aucune étape de centrifugation n'a été appliquée après traitement des phages ou des EVs par les produits testés.

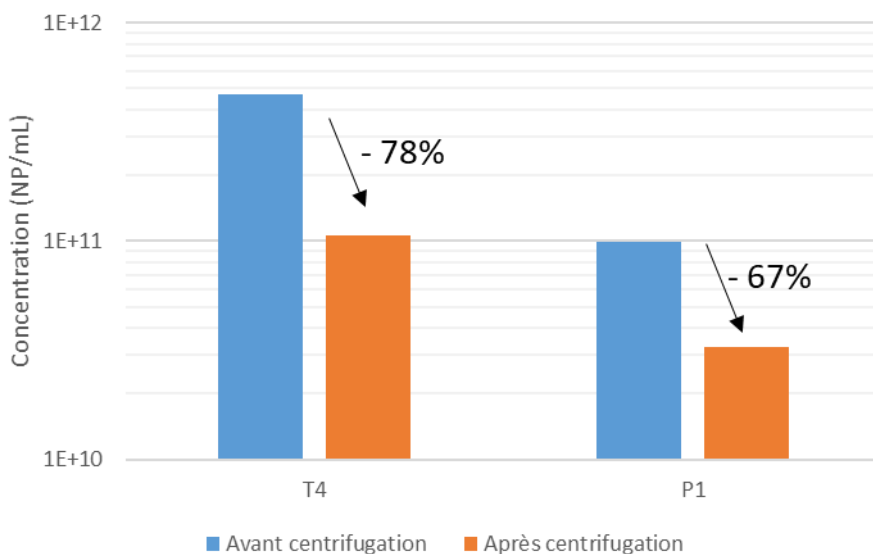


Figure 21. Impact de l'étape de centrifugation sur la quantité de phages dans le surnageant

Mesures de la concentration en nanoparticules détectées par le Videodrop (blocs de 5 acquisitions, de 100 images chacune, seuil relatif = 4.0) après centrifugation de 2 minutes, à 12 000 g réalisée sur les phages T4 et P1 préalablement purifiés, dans du SM buffer.

2.2.2. Impact des composés utilisés sur la concentration en phages et en VE

Les six composés décrits ci-dessus ont été utilisés en présence de phages purifiés et détectables par le Videodrop (T4 et P1) et de VE purifiées selon les protocoles décrits ci-avant (**Figure 22**).

Malgré l'utilisation d'une concentration élevée (50% v/v) de chloroforme, près de la moitié des VE détectées par le Videodrop ne semblent pas affectées par celui-ci. C'est une constatation assez étonnante dans la mesure où des équipes travaillant sur les phages utilisent du chloroforme, parfois à des concentrations beaucoup plus basses (allant jusqu'à 0,1% v/v) pour prévenir les contaminations bactériennes et éliminer les vésicules de leurs préparations (Biller et al. 2017, Dugat-Bony et al. 2020, Forterre et al. 2013). De plus, l'utilisation du chloroforme semble affecter les phages étudiés, puisque près d'un tiers de ceux-ci disparaissent après le traitement. T4 et P1 sont pourtant des phages de laboratoire « historiques », dont l'isolement a été rendu possible grâce à leur résistance au chloroforme. Agata Jurczak-Kurek et co-auteurs (Jurczak-Kurek et al. 2016) avaient déjà relevé cette résistance partielle des phages au chloroforme.

Quatre des six traitements testés n'ont pas permis un niveau de destruction intéressant des VE détectables par le Videodrop (**Figure 22**). En effet, 81% des VE ont résisté au traitement par le Tween 20, 72% ont résisté au traitement par le chloroforme, 71% ont résisté au traitement par le

Vertrel et 43% ont résisté au traitement par le Sarkosyl. Deux traitements permettant la destruction des VE détectables par le Videodrop sont cependant plus efficaces que les autres, à savoir ceux par le NP-40 et par le Triton X100, qui n'ont permis de préserver que 21% et 19% des VE, respectivement. Bien que le Triton X100 soit un peu moins délétère pour les phages que le NP-40, son utilisation rend particulièrement difficile l'observation des échantillons au Videodrop, puisque les gouttes déposées dans la cuve ont tendance à moins adhérer au support et à en sortir. C'est principalement pour cette raison que j'ai estimé que le NP-40 était le composé le plus approprié, parmi les six testés, pour la destruction de VE au sein de mélanges de phages et de VE avant observation au Videodrop.

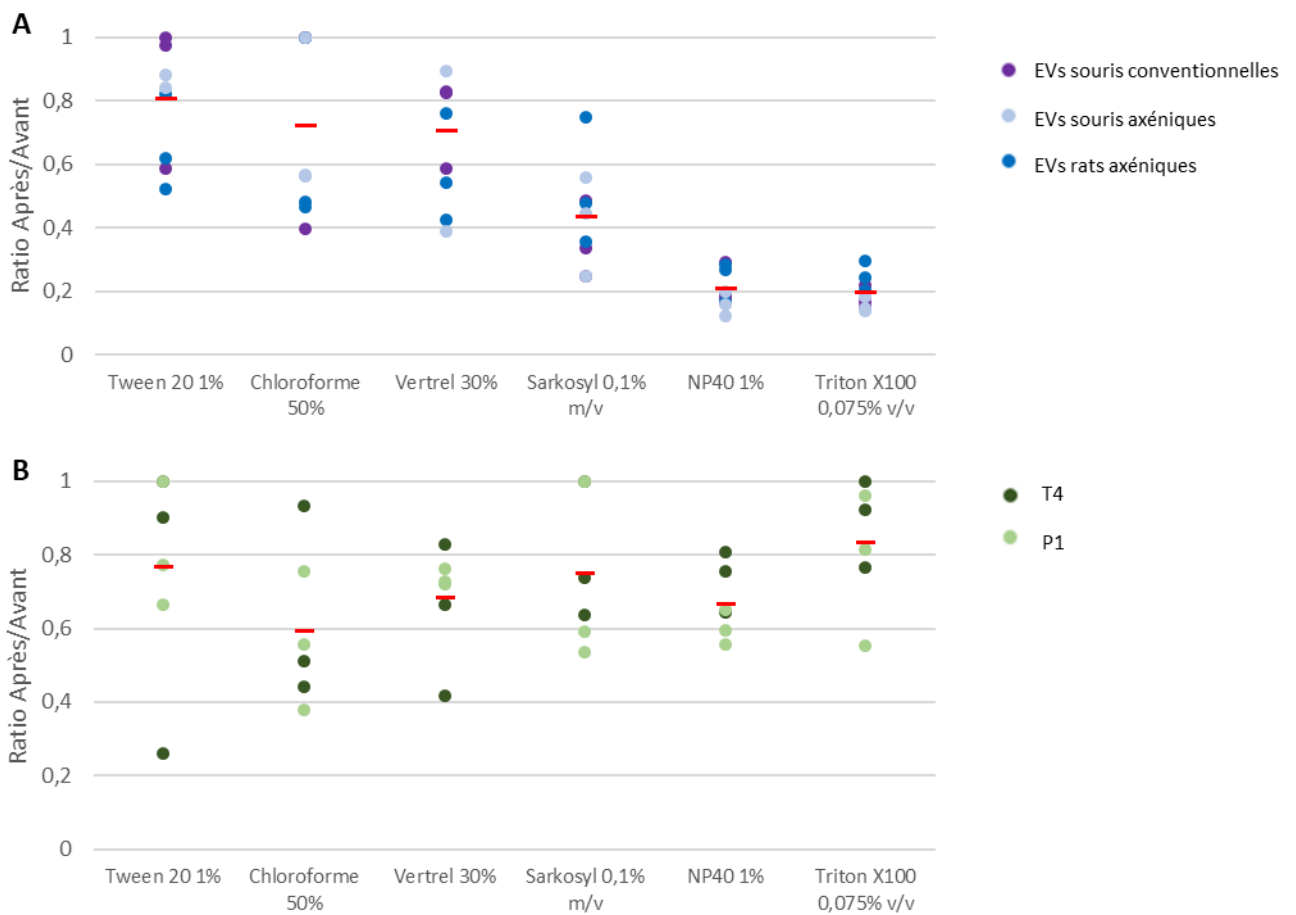


Figure 22. Effets des différents traitements sur les VE et les phages

Effets des différents traitements utilisés sur les VE (A) et sur les phages (B). L'ensemble des mesures ont été réalisées sous la forme de triplicats techniques. Les barres rouges correspondent à la valeur moyenne de chaque ensemble de répliquats.

Afin de s'assurer que les détections obtenues par le Videodrop après le traitement des phages au NP-40 correspondent bien à des phages T4 ou des P1, les distributions de tailles avant et après traitement ont été analysées (**Figure 23**). Bien que les courbes de tendance des distributions de tailles des particules trackées par le Videodrop soient moins lisses après utilisation du NP-40 du fait de la diminution du nombre de particules trackées, les distributions semblent similaires. C'est l'un des arguments permettant d'affirmer que les particules détectées et trackées après le traitement par le Videodrop sont les mêmes particules que celles détectées et trackées avant le traitement.

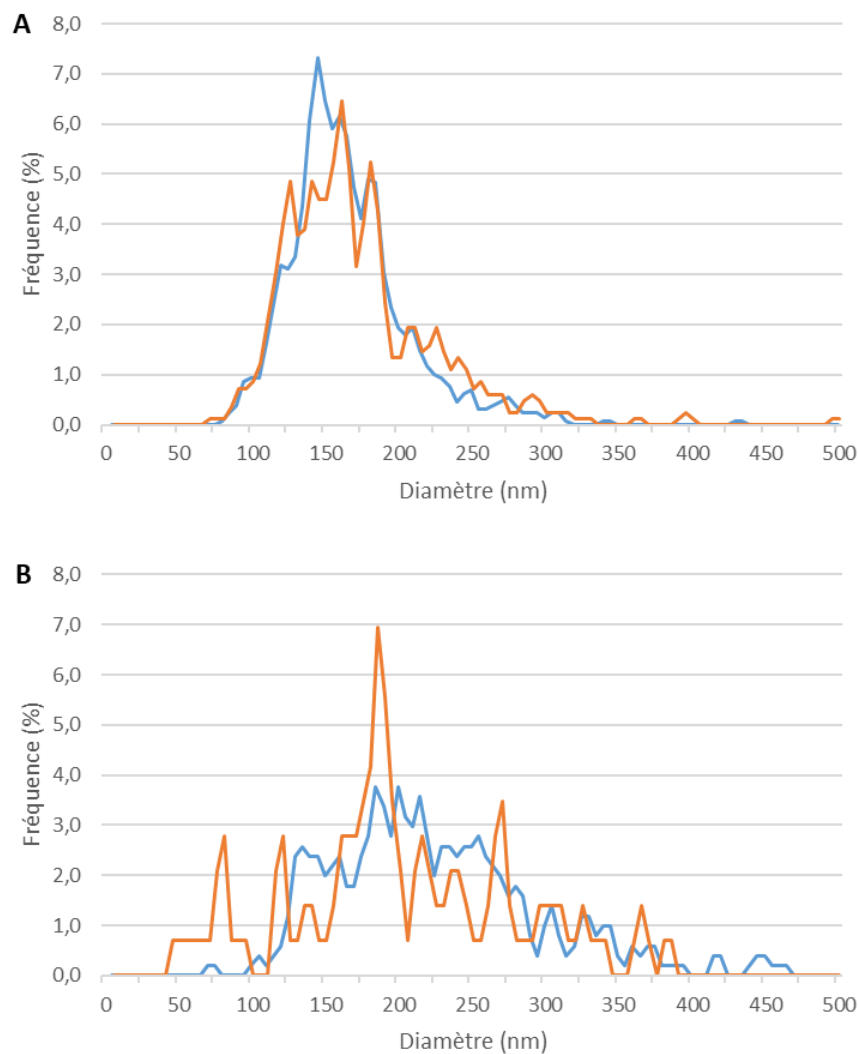


Figure 23. Impact du traitement par le NP-40 sur les distributions de tailles des phages

Impact de l'utilisation de NP-40 (1%) sur les distributions de tailles des phages T4 (A) et P1 (B) détectés et trackés par le Videodrop. Les courbes bleues correspondent aux distributions de tailles dans les solutions de phages purifiés, tandis que les courbes orange correspondent aux distributions de tailles dans les solutions de phages purifiés après traitement par le NP-40.

Afin de confirmer cette hypothèse, les concentrations de phages, avant et après traitement par le NP-40, ont également été déterminées par microscopie par épifluorescence, selon le protocole décrit dans l'article 1 (**Figure 24**). Cette expérience a été réalisée sur sept phages différents. Pour la majorité des phages (à savoir T4, P1, SPP1, λ et Φ X174), les concentrations obtenues par microscopie par épifluorescence ne diffèrent pas de plus d'un facteur deux, qu'il y ait eu un traitement par le NP-40 (1% v/v) ou non. Concernant les phages T7 et PRD1, leur concentration déterminée par microscopie par épifluorescence a été au moins doublée après le traitement par le NP-40. Toutefois, la microscopie par épifluorescence ne permet pas de faire la distinction entre des VLPs correspondant à des phages isolés et les VLPs correspondant à des paquets de phages agrégés. Il est ainsi probable que le NP-40 possède des propriétés désagrégantes pour les phages, ce qui expliquerait cette tendance à l'augmentation de la concentration en VLPs détectées après le traitement.

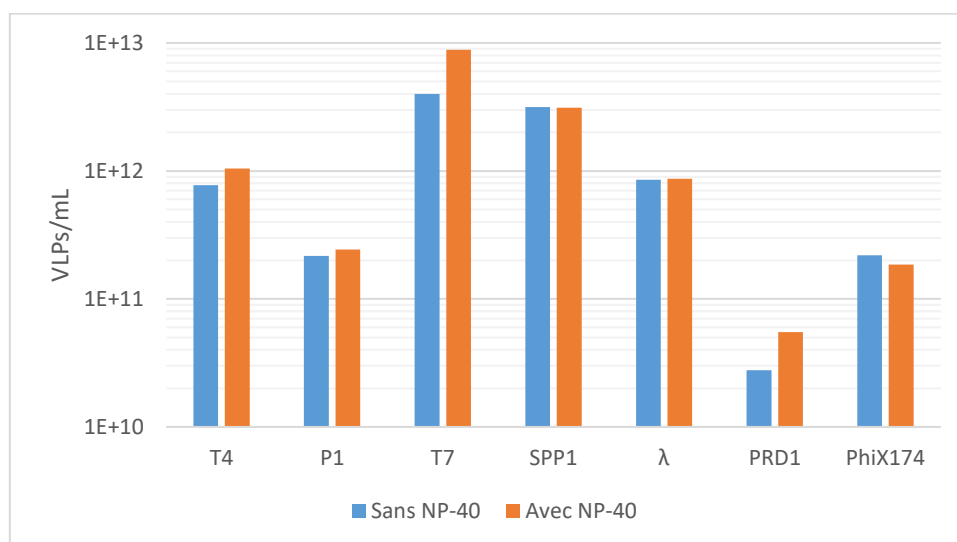


Figure 24. Concentrations de VLPs

Concentrations de VLPs obtenues par microscopie par épifluorescence, dans des stocks de phages purifiés (bleu) et dans des stocks de phages purifiés ayant été traités par le NP-40 (orange).

Bien que non répétés, ces résultats de concentrations de stocks de phages purifiés traités par le NP-40 obtenues par microscopie par épifluorescence sont contradictoires avec celles obtenues par le Videodrop, avec lequel une diminution de près d'un tiers du nombre de nanoparticules détectées est observée (**Figure 23**). Il est possible que le NP-40 induise une tension de surface qui repousse préférentiellement les phages vers les bords de la cuve du Videodrop, auxquels la caméra

permettant la détection des nanoparticules n'a pas accès. Ainsi, la concentration en phages dans la cuve, après traitement par le NP-40, ne serait pas homogène. Cette hypothèse n'a pas été vérifiée.

L'efficacité du traitement par le NP-40 sur les EVs purifiées issues de fèces de souris conventionnelles et de rats axéniques a également été vérifiée par MET. Bien que la MET ne soit pas une technique quantitative, le nombre de vésicules observées sur les grilles est réduit d'environ 90% par champs après utilisation du NP-40. Cela se remarque particulièrement dans le cas des VE de fèces de souris conventionnelles, pour lesquelles le nombre de vésicules, en particulier des plus petites, est fortement diminué. Toutefois, certaines VE, les plus grosses, résistent au traitement (**Figure 25A-B**). Il est également possible que ces VE se soient reformées à partir de lipides présents dans le milieu suite à la destruction de VE initiales lors du traitement. De plus, l'utilisation de NP-40 perturbe la coloration des grilles, en concentrant l'acétate d'uranyle dans la phase minoritaire de l'émulsion. Cela rend l'observation particulièrement difficile dans le cas des VE issues de fèces de rats axéniques (**Figure 25C-D**). Il est alors compliqué de déterminer si des VE sont présentes au sein de ces bulles de coloration.

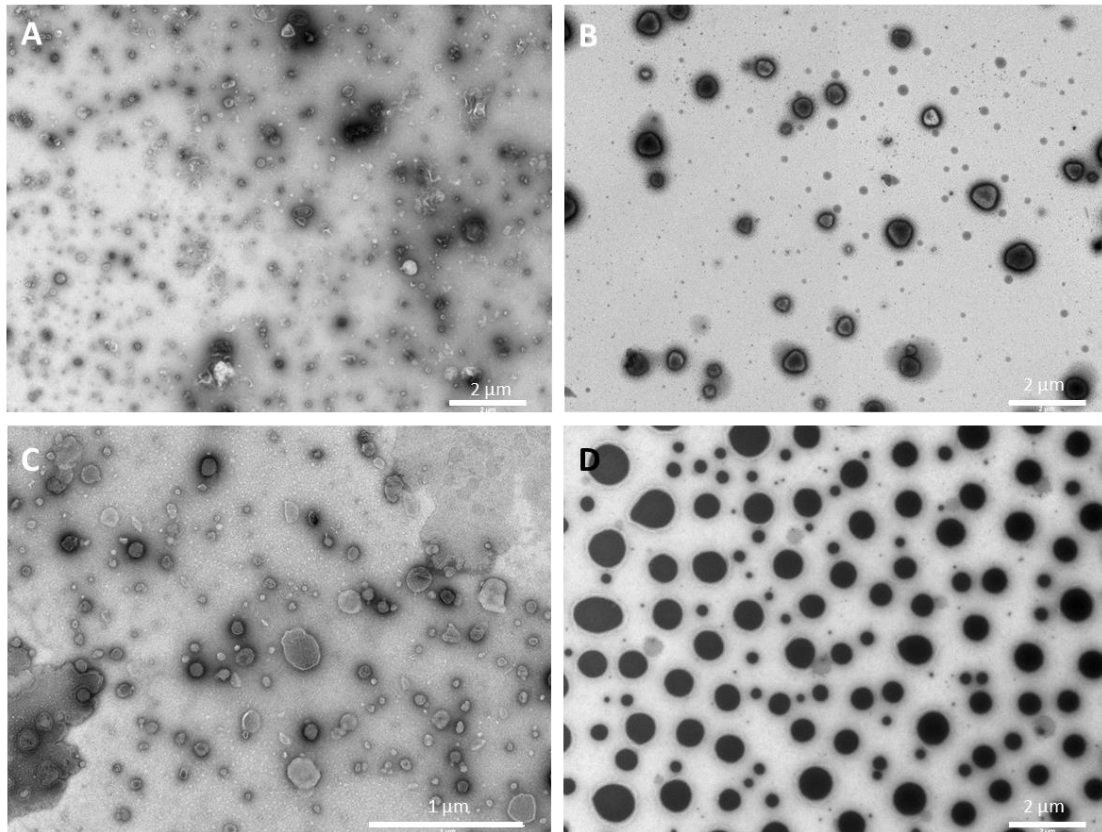


Figure 25. Effets du traitement par le NP-40 sur des VE purifiées, observées par MET

Photographies réalisées par MET sur des VE purifiées issus de fèces de souris conventionnelles (A) sans et (B) après traitement par le NP-40; et sur des VE purifiées issus de fèces de rats axéniques (C) sans et (D) après traitement par le NP-40.

2.3. Conclusion

Le marquage du matériel génétique des phages, couplé à l'utilisation de la microscopie par épifluorescence, est la méthode à la fois la plus fiable et la plus indépendante de la nature des phages pour le dénombrement de ceux-ci dans des mélanges contenant également des VE. Au contraire des phages, les VE présentes dans les filtrats fécaux humains analysés n'émettent pas de signal de fluorescence en présence de Sybr Gold (cf. 3.4.1.2. Contrôles de la spécificité de la détection des phages dans les filtrats fécaux), ce qui permet le dénombrement des phages au sein de mélanges. Le dénombrement des VE au sein de mélanges s'avère quant à elle plus délicate, dans la mesure où aucun marqueur général des VE n'est disponible à ce jour. Cependant, après application d'un traitement par le NP-40 et l'utilisation du Videodrop, il est possible d'en déduire, par soustraction, une approximation de la concentration en vésicules détectables par le Videodrop. Cette valeur n'est qu'un reflet de la concentration réelle en VE dans les échantillons, dans la mesure où le Videodrop ne détecte convenablement que les nanoparticules dont le diamètre hydrodynamique est supérieur

à 90 nm. D'autres moyens seront donc mis en œuvre pour tenter de dénombrer les VE dans les filtrats fécaux, qui contiennent, en plus de ces derniers et des phages, des flagelles et agrégats protéiques et lipidiques détectables par le Videodrop.

3. Quantification de phages et de VE dans les selles d'individus sains et de patients atteints de la maladie de Crohn

Bien qu'une augmentation de la quantité de phages au sein du microbiote intestinal de patients atteints de la maladie de Crohn soit suggérée par une partie des articles à ce sujet, ceux-ci n'ont à ce jour jamais été quantifiés au moyen de méthodes physiques. Seule Patricia Lepage et co-auteurs (Lepage et al. 2008) a quantifié des VLPs par microscopie par épifluorescence dans des biopsies intestinales de patients et a constaté une augmentation de leur nombre chez les malades. Cependant, l'acquisition de biopsies est un processus invasif et ne permet pas de réaliser de quantification rapide. Alors que le manque de biomarqueurs pèse toujours sur le diagnostic des MICI, l'observation d'une augmentation du nombre de VLPs dans les selles de patients pourrait être un premier pas vers un nouvel outil de diagnostic non invasif de ces pathologies, en plus de permettre une potentielle avancée de la compréhension de l'étiologie de la maladie.

Après avoir optimisé des protocoles de quantification des phages et des VE déjà existants, j'ai recueilli des selles provenant d'individus sains ou malades et ai procédé au dénombrement de ces objets nanométriques au sein de celles-ci afin de répondre à cette problématique.

3.1. Constitution des cohortes

La maladie de Crohn étant une pathologie caractérisée par l'occurrence de périodes de crises dites poussées et de phases plus stables dites de rémission déterminées cliniquement par l'intensité des symptômes et de l'inflammation intestinale. Trois différents groupes de selles ont été constitués en se basant sur ces données cliniques: des selles provenant de donneurs sains, de patients en période de poussée et de patients en période de rémission.

Vingt-trois échantillons de selles constituent le groupe des donneurs sains (S). L'ensemble de ces échantillons a été prélevé dans le cadre de la banque de selles Normathèque, à l'hôpital Saint-Antoine (Paris). Tous les échantillons ont été anonymisés et proviennent d'individus âgés de 18 à 65 ans, de sexe féminin ou masculin.

Les selles de patients atteints de la maladie de Crohn, dont la pathologie et le stade de celle-ci ont été attestés par un médecin, sont issues de la cohorte Suivithèque, également à l'hôpital Saint-Antoine. Le groupe de patients en phase de rémission (R) est constitué de vingt individus âgés de 23 à 84 ans au moment de leur don de selle et comprend dix femmes et dix hommes. Le groupe de patients en phase de poussée (P) est quant à lui composé de vingt individus, âgés de 20 à 57 ans et comprend 8 femmes et 12 hommes. Parmi les individus en poussée, deux hommes, de 24 ans et de 52 ans, avaient déjà subi une ablation partielle de l'intestin avant le recueil des selles.

L'ensemble des selles prélevées ont été aliquotées et conservées à -80°C avant d'être étudiées dans le cadre de cette thèse.

3.2. Dénombrement des bactéries dans les selles

Bien que les phages et les VE soient au cœur de ce projet de thèse, nous nous sommes également intéressés à la quantité de bactéries présentes au sein des selles des individus constituant les différents groupes, afin d'en comparer les ratios VLPs/bactérie et VE/bactérie. En plus d'être susceptibles de présenter des différences de concentrations en bactéries, toutes les selles analysées ne présentent pas les mêmes concentrations en eau. De plus, les individus malades sont plus sujets à la survenue de diarrhées et la faible quantité de matière récoltée (en particulier pour les selles de patients) n'a pas permis de mesurer la masse sèche des selles. Afin de pouvoir comparer les échantillons les uns aux autres, j'ai opté pour cette solution moins consommatrice en selles et

permettant l'obtention d'informations supplémentaires sur celles-ci, à savoir leur concentration en bactéries.

3.2.1. Détection des bactéries par cytométrie en flux

Pour quantifier les bactéries présentes dans les selles, le protocole de Doris Vandeputte et co-auteurs (Vandeputte et al. 2017), basé sur la cytométrie en flux, a été reproduit (cf. Matériel et Méthodes).

Afin de différencier les particules réellement détectées par l'appareil du bruit de fond, une stratégie de gating a été appliquée (**Figure 26**), en suivant le modèle suivant. Une solution de contrôle négatif, c'est-à-dire la solution physiologique utilisée pour l'homogénéisation et la dilution des selles, a été soumise aux mêmes filtrations et marquage que le reste des échantillons, puis analysée par FCM. En comparant les mesures de fluorescence obtenues entre le contrôle négatif et un premier échantillon issu de selles (**Figure 26A** et **Figure 26C**), la présence d'une population partagée entre les deux solutions est remarquée. Il s'agit de particules présentes dans la solution physiologique (dues au relargage de particules par les filtres, à l'autofluorescence du SYBR Green, etc). Un premier gate nommé P est établi de façon à exclure la population de débris de l'analyse. Afin d'éviter la prise en compte de débris non éliminés par ce premier gate dans les calculs de concentrations dans les selles, la lumière diffusée par les particules est également analysée. Un second gate (R) est ainsi déterminé pour écarter de l'analyse toutes les particules ayant une taille ou une granulométrie proche de celles détectées dans le contrôle négatif (**Figure 26B**). Ainsi, au sein d'une solution de selle homogénéisée, la quantité de bactéries détectées peut être déduite. Il s'agit du nombre d'évènements enregistrés dans le gate P, duquel est soustrait le nombre d'évènements enregistrés dans le gate R (souvent proche de 0) (**Figure 26D**).

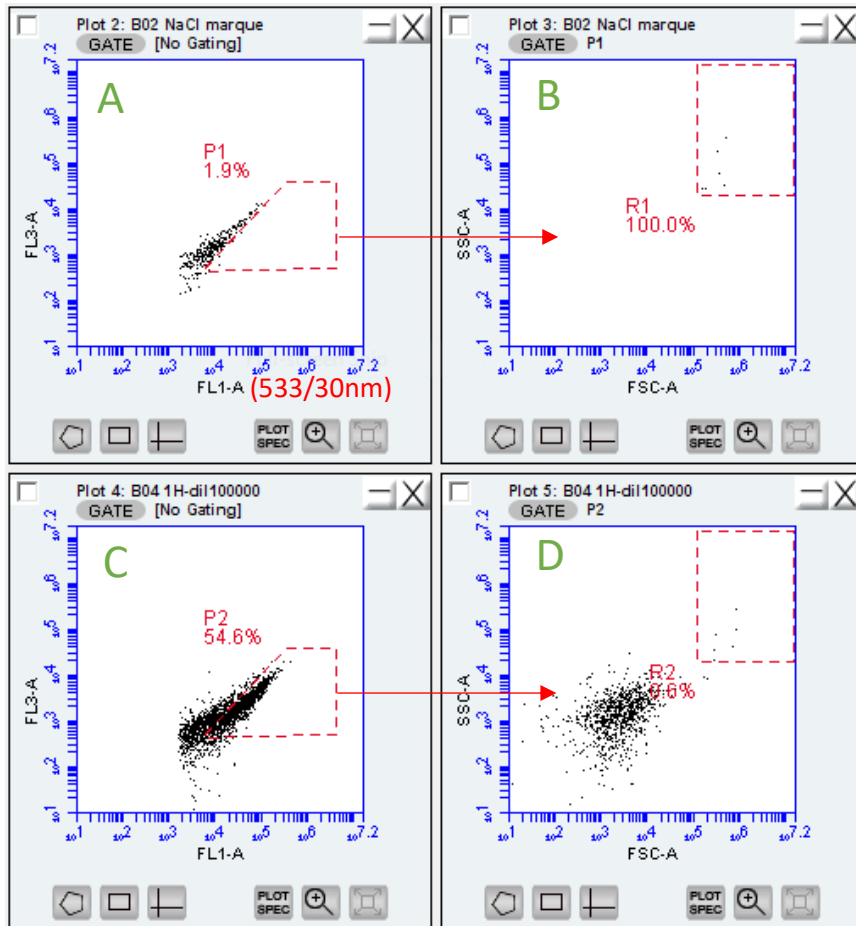


Figure 26. Méthode de gating employée lors des analyses de cytométrie pour des déterminer la concentration des selles en bactéries

(A) Intensités de fluorescence des évènements enregistrés dans une solution physiologique. (B) Granulométrie des particules détectées dans une solution physiologique. (C) Intensités de fluorescence des évènements enregistrés dans une solution fécale. (D) Granulométrie des particules détectées dans une solution fécale.

3.2.2. Concentration des bactéries dans les selles

La détection des bactéries par cytométrie a été réalisée dans l'ensemble des selles de la cohorte constituée pour ce projet. La masse mesurée pour chaque aliquot a ainsi permis d'obtenir des concentrations en bactéries pour les trois groupes de la cohorte (**Figure 27**).

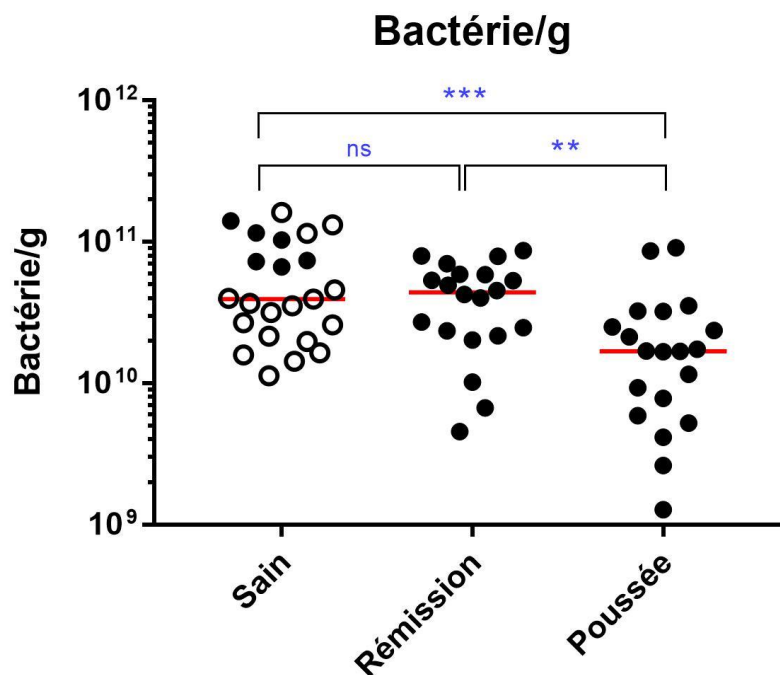


Figure 27. Concentration des selles en bactéries

Les cercles pleins correspondent à des selles ayant été conservées plus d'un an à -80°C , tandis que les cercles vides correspondent à des selles ayant été conservées moins de deux mois à -80°C . Un test de Mann-Whitney a été réalisé pour comparer les différents groupes (ns = $0,1234 > p\text{-value} > 0,0332$; * = $0,0332 > p\text{-value} > 0,0021$; ** = $0,0021 > p\text{-value} > 0,0002$; *** = $0,0002 > p\text{-value} > 0,0001$; **** = $p\text{-value} < 0,0001$). Les barres horizontales rouge représentent la médiane de chaque groupe.

Chez les individus sains, la concentration en bactéries varie de $1,13 \times 10^{10}$ à $1,62 \times 10^{11}$ bactéries par gramme de selle, avec une valeur médiane de $3,74 \times 10^{10}$ bactéries par gramme de selle a été mesurée. Une forte variabilité interindividuelle, d'un facteur 10, peut être relevée.

Chez les patients en phase de rémission, des concentrations similaires à celles des individus sains ont été relevées. En effet, les concentrations varient de $4,55 \times 10^9$ à $8,67 \times 10^{10}$ bact/g, pour une médiane de $4,39 \times 10^{10}$ bact/g. Aucune différence significative entre les individus sains et les patients en rémission n'a été trouvée.

Cependant, une concentration moins élevée en bactéries a été observée dans les selles des malades en période de crise. En effet, avec une valeur médiane de $1,68 \times 10^{10}$ bact/g, une différence significative avec le groupe des individus sains et des patients en rémission apparaît. De plus, la variabilité interindividuelle est augmentée chez les patients en crise : les concentrations oscillent en effet de $1,28 \times 10^9$ bact/g à $9,08 \times 10^{10}$ bact/g, soit près d'un facteur 100 de différence entre les deux concentrations les plus extrêmes. Nous ne disposons cependant pas des scores de Bristol des selles

issues des cohortes et ne pouvons donc pas établir de corrélation entre les concentrations en bactéries et la consistance des selles.

Les concentrations en bactéries obtenues permettront néanmoins la comparaison des quantités en phages et en vésicules des différents groupes entre eux par l'établissement de ratios.

3.2.3. Discussion

Chez les individus sains, la concentration en bactéries varie de $1,13 \times 10^{10}$ à $1,62 \times 10^{11}$ bactéries par gramme de selle et une valeur médiane de $3,94 \times 10^{10}$ bactéries par gramme de selle a été mesurée. Cette forte variabilité interindividuelle, d'un facteur 10, avait déjà observée (Vandeputte et al. 2017). Cependant, les auteurs de ces travaux décrivent une concentration médiane de bactéries un peu plus élevée chez les individus sains ($1,5 \times 10^{11}$ bactéries par gramme de selle). Cette différence peut être due à l'utilisation de selles fraîches par cette équipe, bien que ceux-ci n'aient pas remarqué de différence de concentration après une étape de congélation à -80°C ($r=0,91$ au test de Pearson). Dans mon cas, les échantillons ont pu être conservés plusieurs années. De plus, le recueil des selles d'individus sains a été réalisée par les volontaires eux-mêmes, ce qui a pu conduire à des temps plus ou moins longs entre le recueil et la congélation des selles, conduisant à la mort de certaines bactéries. Cela aurait pu conduire à la disparition de certaines bactéries, mais ce n'est pas ce qui est observé dans les selles d'individus sains, dont les durées de congélation à -80°C varient de quelques semaines à près de 10 ans.

3.3. Préparation des filtrats fécaux

Pour quantifier les phages et les VE présents dans les selles, il est nécessaire dans un premier temps de les séparer des objets de plus grande taille. C'est dans cette optique que l'on prépare en première étape des filtrats fécaux, c'est-à-dire des solutions de selles débarrassées de leurs bactéries et de leurs débris de taille micrométrique, afin de ne conserver que les objets biologiques de taille nanométrique. Pour ce faire, un protocole utilisé par notre équipe a été suivi (Cornuault et al. 2020) (cf. Matériel et Méthodes).

Parallèlement à la quantification de ces populations, les filtrats fécaux peuvent être observés en MET. Les échantillons ont d'abord été fixés à l'aide de glutaraldéhyde, à une concentration finale de 0,5% v/v. Ils ont alors été déposés, par gravité, sur des grilles carbonées, puis colorés négativement par une solution d'acétate d'uranyle. L'observation en MET des filtrats fécaux permet de se rendre compte de la diversité des nano-objets biologiques présents au sein de ceux-ci

Figure 28). Des vésicules et des fragments de flagelles ont été détectés, ainsi que des agrégats protéiques et/ou lipidiques à l'origine de l'effet de maillage noir présent sur le fond de certaines grilles. Si la présence de phages et de virus est également soupçonnée au sein des filtrats fécaux, leur morphologie, les rendant visuellement très similaires à certaines VE, ne permet pas de les identifier avec certitude dans la majorité des cas.

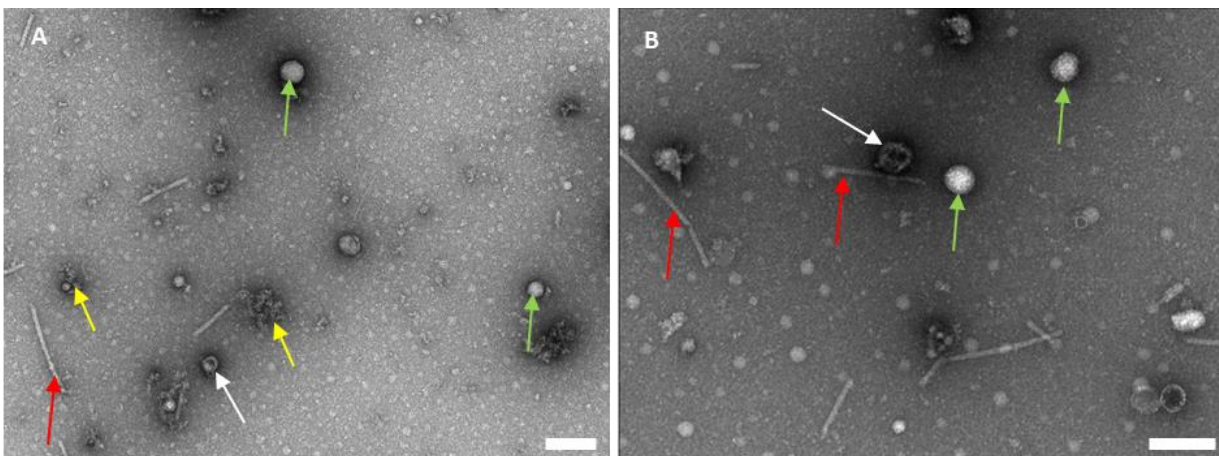


Figure 28. MET de filtrats fécaux d'individus sains

Filtrats fécaux obtenus à partir des selles de (A) l'individu S13 et de (B) l'individu S9. Les flèches vertes pointent des potentiels phages ou virus. Les flèches blanches pointent des VE. Les flèches rouges pointent des flagelles. Les flèches jaunes pointent des agrégats.

3.4. Détection et dénombrement des phages dans les filtrats fécaux

D'après la littérature citée en introduction, une augmentation du nombre de phages ou de VLPs, c'est-à-dire du nombre de particules pouvant être assimilées à des virus, pourrait être observée au niveau de la muqueuse intestinale et dans les selles au cours des MICI (Lepage et al. 2008). Comme présenté dans l'article 1, il est possible de quantifier des phages purifiés, même pour les plus petits d'entre eux comme les *Microviridae*, par épifluorescence. Afin de pouvoir énumérer les phages dans

les filtrats fécaux préparés à partir de la cohorte, il fallait cependant s'assurer que les phages soient, d'une part, bien extraits des selles et, d'autre part, détectables au sein des filtrats fécaux.

3.4.1. Détection des phages dans les filtrats fécaux

3.4.1.1. Choix du protocole

Le choix du protocole de préparation des filtrats fécaux (cf. 3.3. Préparation des filtrats fécaux) n'a pas été le premier envisagé. En effet, depuis quelques années, l'ensemble des échantillons de la cohorte étaient stockés à -80°C après une étape de resuspension dans du RNA later, un tampon contenant entre autres du citrate de sodium, du citrate d'ammonium et de l'EDTA, ce qui permet de meilleurs résultats pour la métagénomique et la métatranscriptomique du microbiote. Il a donc fallu comparer les filtrats fécaux préparés selon le protocole « RNA later » employé à l'hôpital Saint-Antoine, soit selon le protocole de l'équipe décrit précédemment. De plus, un second tampon a été testé avec le protocole communément utilisé au sein de l'équipe, à savoir le PBS, par crainte de subir une instabilité des phages et/ou des VE dans les filtrats fécaux à base de Tris, à cause d'un choc osmotique. Les filtrats fécaux ont par conséquent été préparés selon les protocoles décrits en (cf. Matériel et Méthodes)

Des aliquots issus d'une même selle humaine ont été utilisés pour cette expérience de comparaison de protocoles. Le protocole de quantification des VLPs par microscopie par épifluorescence décrit dans l'article 1 a été employé pour énumérer les VLPs des trois filtrats fécaux obtenus (Classique-Tris, Classique-PBS et RNA later). Aucune mesure n'a pu être réalisée dans les échantillons issus du protocole utilisant le RNA later, du fait d'un bruit de fond trop important. Les résultats obtenus avec le protocole classique, indiquent que la concentration en VLPs est presque identique avec les tampons Tris et PBS ($1,4 \times 10^9$ VLPs par gramme en Tris contre $7,7 \times 10^8$ VLPs par gramme en PBS). Bien qu'un facteur 2 sépare ces valeurs, la différence n'est pas significative (pas de réplicats). Parce que le Tris est classiquement utilisé dans l'équipe et parce qu'il semblerait que le PBS gêne la coloration par l'acétate d'uranyle en MET (résultats non publiés), le protocole classique-Tris, présenté précédemment, a été suivi pour le reste des travaux présentés ici.

La conservation des selles en RNA later n'est donc pas adaptée à l'observation des phages en épifluorescence. Heureusement, il restait dans la banque de selles Suivithèque d'anciens échantillons conservés « secs » (i.e échantillons conservés directement à -80°C sans étape en RNA later), qui ont pu être utilisés pour cette étude.

3.4.1.2. Contrôles de la spécificité de la détection des phages dans les filtrats fécaux

Bien que les VE ne soient a priori que peu marquables par le Sybr Gold (cet intercalant de l'ADN, hydrophile, doit dans un premier temps franchir la membrane lipidique pour trouver du matériel génétique, pas obligatoirement présent dans les VE), il a été rapporté que cela était cependant possible (Forterre et al. 2013), conduisant à une surestimation des VLPs. Afin de s'assurer que les centres de fluorescence détectés en microscopie par épifluorescence proviennent bien des virus et non des vésicules, ces deux populations ont été purifiées à partir de filtrats fécaux d'individus sains et de patients. Des gradients d'iodixanol, dont le protocole est détaillé dans l'article 1, ont été réalisés sur des filtrats fécaux issus de selles de trois individus sains et d'un patient en poussée. Ce protocole étant très consommateur en matière fécale (1 g), il n'a pas été reproduit avec les selles d'autres patients. Les différentes fractions issues des gradients de densité ont été marquées au Sybr Gold selon le protocole décrit dans l'article 1 et les VLPs ont été quantifiées par microscopie par épifluorescence (**Figure 29**). Malgré la présence de détections dans la fraction 2, dont la densité est celle des VE, les concentrations en VLPs dans la fraction 4, dont la densité correspond à celle des phages, est 10 fois supérieure. L'origine de la selle (individu sain ou patient en poussée) n'a pas d'impact sur la quantité de VLPs détectée dans la fraction 2, ce qui signifie que les VE, au cours de l'inflammation, ne sont pas susceptibles d'être plus marquées par le Sybr Gold. Bien qu'une perte du nombre de VLPs soit observée entre le filtrat fécal et la fraction 4, probablement pendant le gradient, la majorité des détections obtenues par microscopie par épifluorescence provient des virus.

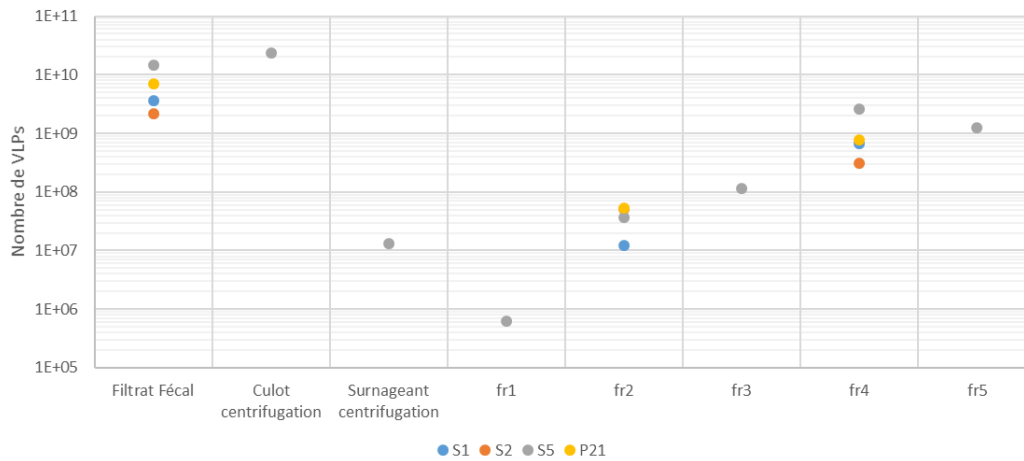


Figure 29. Nombre de VLPs détectées dans les fractions

Concentrations en VLPs détectées par épifluorescence, dans les fractions issues du gradient d'iodixanol. Les échantillons S1, S2 et S5 proviennent d'individus sains et l'échantillon P21 d'un patient en poussée. La quantité de VLPs n'a pas été déterminée dans les fractions 1, 3 et 5, ni dans le culot et le surnageant de centrifugation des échantillons S1, S2 et P21.

Le contenu des fractions 4 a été vérifié en MET pour quelques échantillons. Des caudophages de toutes les familles ont pu être observés, parmi des vésicules et fragments flagellaires caractéristiques de cette fraction. Des myophages et des podophages ont ainsi été détectés dans la fraction virale de l'individu S2 (**Figure 30A**), tandis qu'un siphophage a été vu dans la fraction virale de l'individu S9 (**Figure 30B**). Il est cependant plus difficile de différencier les phages non caudés des vésicules, malgré leur abondance non négligeable rapportée dans les viromes du microbiote intestinal humain.

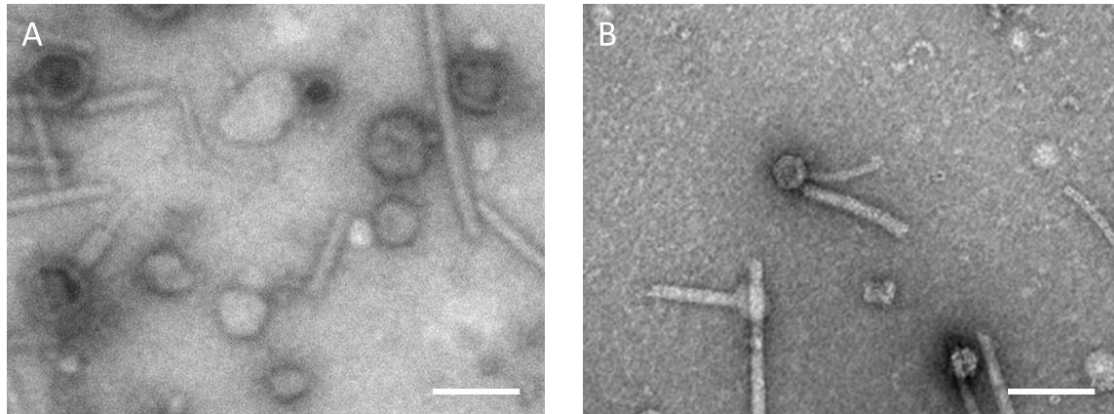


Figure 30. MET de fractions virales de filtrats fécaux d'individus sains

Fractions virales (fraction 4) de filtrats fécaux des individus sains (A) S2 et (B) S9 observés en MET, après coloration par acétate d'uranyle. Les barres d'échelle représentent 100 nm.

3.4.1.3. Contrôles de la quantification des phages dans les selles

Bien que les phages soient dénombrables par microscopie par épifluorescence (2.1.1. Détection des phages), il a fallu s'assurer qu'il en soit de même dans les filtrats fécaux humains. Ces milieux étant plus complexes, la détection de VLPs pourrait être affectée, d'une part par une éventuelle autofluorescence des filtrats fécaux, et d'autre part par l'agrégation de phages entre eux ou à des molécules présentes dans les selles, comme les mucines.

Des quantités de phages connues (issus des stocks de phages préparés pour l'article 1 et dont les concentrations ont été à nouveau quantifiées par microscopie par épifluorescence la veille de l'expérience) ont été enrichies dans des selles d'individus sains présentant des concentrations faibles en VLPs. Après préparation des filtrats fécaux, les quantités de phages présentes dans ceux-ci ont été estimées par microscopie par épifluorescence. Le myophage T4 et le siphophage SPP1 ont ainsi été ajoutés, à trois concentrations différentes en triplicats, dans les selles S1, S6 et S9. Le Φ crAss001, podophage majoritaire du microbiote intestinal a quant à lui été ajouté à trois concentrations différentes (en triplicats) dans les selles S5, S7 et S8. Enfin, le phage Φ X174, un *Microviridae*, a été ajouté dans l'ensemble de ces selles (**Figure 31A**). Pour les phages T4, SPP1 et Φ crAss001, les observations en épifluorescence montrent une forte corrélation entre le nombre de phages déposés dans les selles et le nombre de VLPs quantifiées par épifluorescence. Ainsi, l'environnement fécal ne dérange pas les quantifications de ces phages. Pour Φ X174, une adaptation protocole a été nécessaire pour obtenir une bonne quantification. Tout d'abord, une approche d'imagerie. Plutôt que de rechercher directement les maximum d'intensité dans les images, une soustraction du bruit de fond par région de l'image a été appliquée en soustrayant à l'image un double d'elle-même ayant

été traité par un flou gaussien d'un rayon de 12 pixels. L'image ainsi obtenue est donc nettoyée du gradient de variation d'intensité dû à l'optique et au relief des membranes Anodisc utilisées.

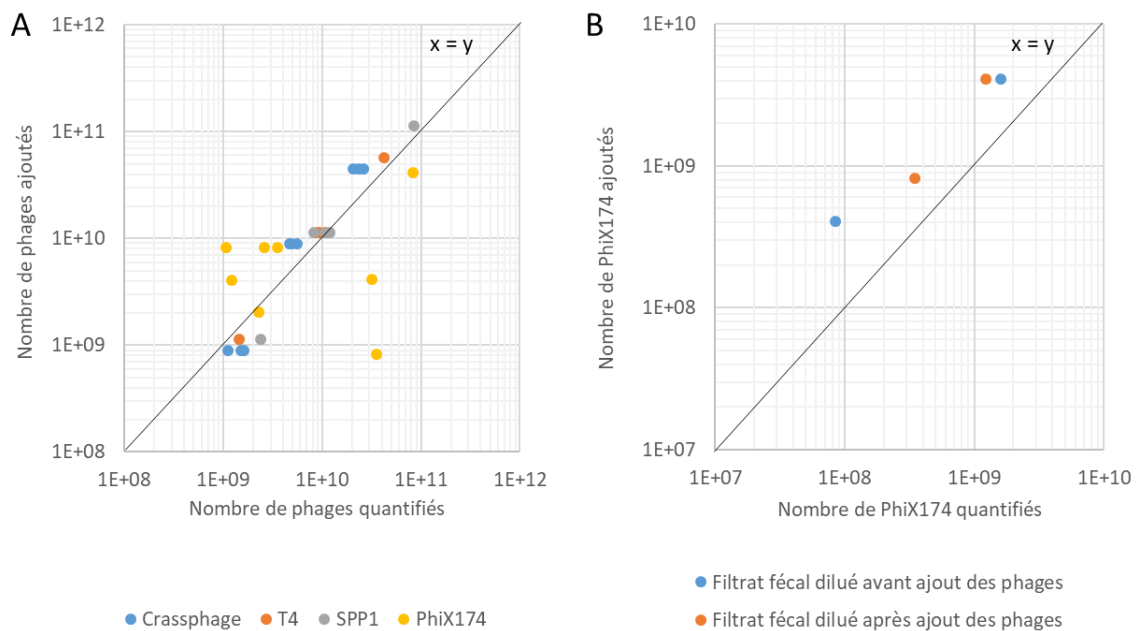


Figure 31. Inoculation de phages en quantités connues dans des filtrats fécaux

(A) Inoculations de T4, SPP1, Φ crAss001 et Φ X174 (quantités inoculées déterminées par microscopie par épifluorescence) dans des selles, dont les VLPs des filtrats fécaux ensuite produits ont été quantifiées par microscopie par épifluorescence. (B) Inoculations en Φ X174 dans un filtrat fécal dilué au 10e ou dans une selle dont le filtrat fécal a été dilué au 10e par la suite. Les VLPs présentes dans les filtrats fécaux ainsi obtenus ont été quantifiées par microscopie par épifluorescence.

3.4.1.4. Impact des dilutions

Une dilution des filtrats fécaux est nécessaire pour détecter correctement Φ X174 (**Figure 31B**). Après dilution, les quantités de Φ X174 quantifiées sont proportionnelles aux quantités de phages ajoutées (et environ trois fois inférieures). Le même résultat est obtenu que le phage soit ajouté dans un filtrat fécal préalablement dilué ou directement dans une selle dont le filtrat fécal a été dilué par la suite. Malgré cette amélioration, Φ X174 reste proche de la limite de détection pour la quantification par épifluorescence, est très difficile à énumérer et ne peut être dénombré que dans les filtrats fécaux suffisamment dilués.

Le dénombrement des phages peu fluorescents après marquage par le SYBR Gold pourrait être affecté de la même façon que Φ X174, et il est possible que les concentrations en VLPs dans les filtrats fécaux, déterminées par la suite, soient légèrement sous estimées du fait de cette limite.

Le dénombrement des VLPs dans un nombre plus important de filtrats fécaux a également permis de découvrir d'autres limites, non anticipées et ayant nécessité des ajustements de protocole. Ce fut notamment le cas concernant l'obligation de diluer certains échantillons trop concentrés. Ainsi, comme le montre la **Figure 32** *Erreur ! Source du renvoi introuvable.* qui se trouve être un cliché représentatif de l'importance de la dilution des filtrats fécaux lors de la quantification des VLPs par microscopie par épifluorescence, des tâches floues de fluorescence sont observées avant dilution. Après dilution du filtrat fécal, des spots de fluorescence bien définis sont observés. Il est possible que les virus aient tendance à former des agrégats sur les surfaces des membranes Anodisc lorsqu'ils sont trop nombreux. Une dilution supplémentaire, ici au 100^e, permet de s'en affranchir et d'observer des VLPs individuelles, faciles à compter. Par ailleurs, l'absence d'homogénéité du bruit de fond et du focus, déjà relevés précédemment nécessite de rogner certains clichés sur des zones où le focus est optimal.

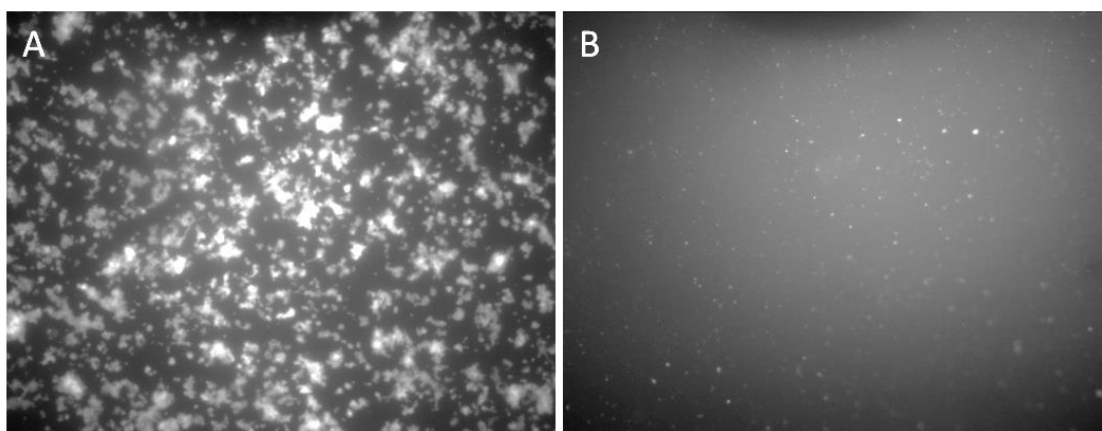


Figure 32. Filtrat fécal R2 observé par microscopie par épifluorescence

(A) Dilution au 100^e ; (B) dilution au 10000^e.

3.4.1.5. Impact de l'utilisation de DNase I

Malgré les dilutions, il s'est révélé parfois encore difficile de quantifier les VLPs de certains filtrats fécaux, du fait de la présence de tâches de fluorescence plus ou moins floues (**Figure 33A**). L'hypothèse que ces nuages, différents des points d'intensité provoqués par les phages marqués, étaient dus à de l'ADN libre présent dans les filtrats fécaux est rapidement apparue. Un traitement

par la DNase I (1 UI pour 10 µL de filtrat fécal dilué à une concentration adéquate, incubation 30 minutes à 37°C) a donc été ajouté au protocole de préparation des lames. Suite à ce traitement, presque la totalité des tâches diffuses disparaît (**Figure 33B**), confortant l'idée qu'il s'agit bien d'ADN libre marqué, pointant la nécessité d'inclure systématiquement une étape de traitement des filtrats fécaux par la DNase I au cours de leur préparation pour la quantification des VLPs.

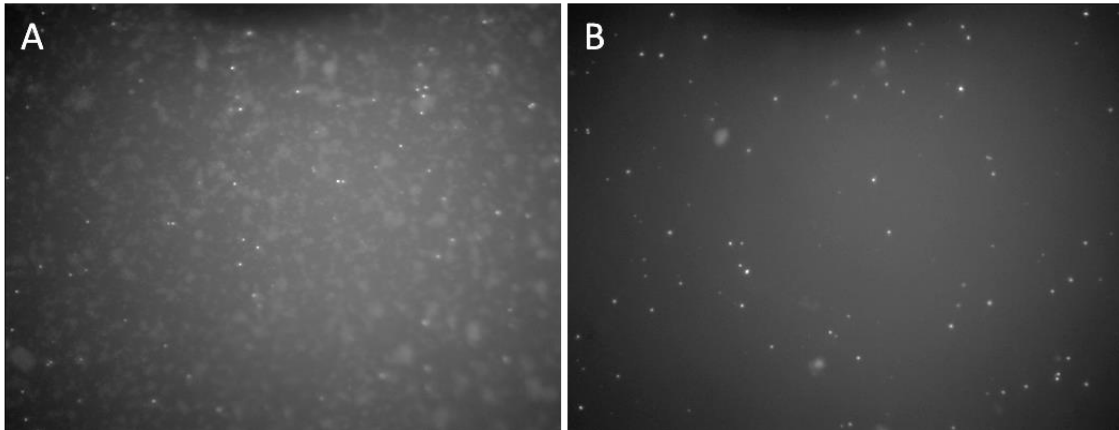


Figure 33. Filtrat fécal R9 observé par microscopie par épifluorescence

(A) Sans DNase I ; (B) Après traitement par la DNase I.

3.4.2. Dénombrement des phages dans les filtrats fécaux

Les concentrations en VLPs par gramme de selle ont ainsi été évaluées selon le protocole décrit plus loin (cf. Matériel et Méthodes) au sein des trois groupes constituant notre cohorte (**Figure 34Erreur ! Source du renvoi introuvable.**). Une valeur médiane de $8,36 \times 10^9$ VLPs par gramme de selle a été obtenue chez les individus sains, contre $1,59 \times 10^{10}$ chez les patients en période de rémission de la maladie de Crohn, soit une différence significative de près du double de VLPs par gramme. Une concentration bien plus importante du nombre de VLPs par gramme de selle est observée dans les selles des patients atteints de la maladie de Crohn en période de poussée, avec une concentration médiane à $3,38 \times 10^{11}$ VLPs par gramme de selle, soit de plus de 40 fois plus que dans les selles des individus sains.

Ces résultats se caractérisent également par une importante variabilité au sein des différents groupes, s'étalant sur 2 logs chez les individus sains, 3 logs chez les patients en poussée et 5 logs chez les patients en période de rémission.

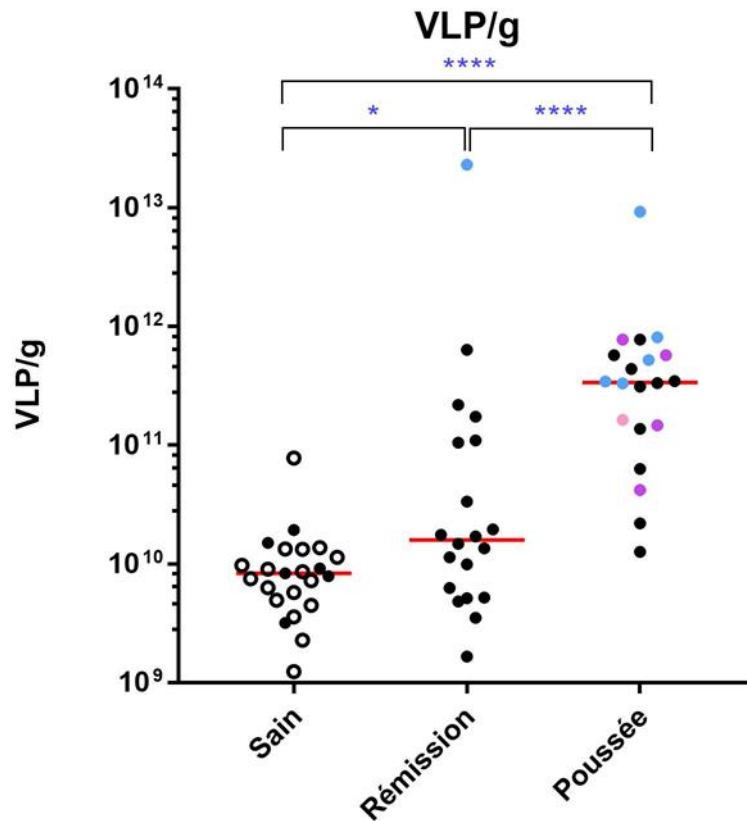


Figure 34. Dénombrement des VLPs dans les filtrats fécaux

Toutes les mesures (en VLPs/g de selle) ont été obtenues par comptage des VLPs par microscopie par épifluorescence. Les différentes couleurs représentent les dilutions des filtrats fécaux utilisées pour la détection des VLPs (noir : 1/100 ; rose : 1/500 ; violet : 1/1000 ; bleu : 1/10000). Les cercles pleins correspondent à des selles ayant été conservées plus d'un an à -80°C , tandis que les cercles vides correspondent à des selles ayant été conservées moins de deux mois à -80°C . Un test de Mann-Whitney a été réalisé pour comparer les différents groupes (ns = $0,1234 > p\text{-value} > 0,0332$; * = $0,0332 > p\text{-value} > 0,0021$; ** = $0,0021 > p\text{-value} > 0,0002$; *** = $0,0002 > p\text{-value} > 0,0001$; **** = $p\text{-value} < 0,0001$). Les barres horizontales rouges correspondent aux médianes de chaque groupe.

3.4.3. Discussion

Rares sont les travaux ayant permis le dénombrement de phages ou de VLPs au sein de selles, qu'elles proviennent d'individus sains ou de patients atteints de MICI. L'équipe de Patricia Lepage (Lepage et al. 2008) rapporte cependant une augmentation d'un facteur 20 du nombre de VLPs détectées dans les biopsies de zones inflammatoires de patients par rapport à des zones épargnées. Bien que le microbiote soit différent entre la paroi épithéliale et les selles, la tendance observée par les auteurs est confirmée par nos propres observations, puisque nous rapportons ici une augmentation de la quantité absolue de VLPs d'un facteur 40 dans le groupe de patients en période

de poussée inflammatoire, par rapport au groupe d'individus déclarés comme étant sains. Nos résultats se dirigent également vers la même direction que ceux d'une étude métagénomique réalisée sur des biopsies iléales montrant une augmentation de l'abondance phagique chez les patients (Wagner et al. 2013). Une autre étude, basée cette fois sur une approche métaprotéomique, présente quant à elle une corrélation positive entre la présence de protéines phagiques et l'inflammation/maladie (Zhang et al. 2018). Nos résultats de quantification de la concentration absolue des VLPs dans les selles d'individus sains ($8,36 \times 10^9$ VLPs/g) sont cependant légèrement plus élevés que ceux obtenus par Lesley Hoyles et co-auteurs (Hoyles et al. 2014), qui ont quantifié une moyenne de $2,96 \times 10^9$ VLPs/g de selles d'individus sains par microscopie par épifluorescence. Nos valeurs de concentrations médianes de VLPs sont, d'autre part, près de 10 fois plus élevées que celles estimées par métagénomique virale quantitative (par inoculation de phages connus) par Andrey Shkoporov et co-auteurs (Shkoporov et al. 2021), qui s'élèvent à environ 1×10^9 génomes par gramme de selle chez le Macaque. Si les quantifications par métagénomique peuvent contenir de nombreux biais, il faut également relever que ces expériences d'estimation de concentrations de virus ont été réalisées chez des macaques, qui, bien que phylogénétiquement proches de l'être humain, présentent évidemment d'importantes différences anatomiques et microbiotiques. Ces approches basées sur le séquençage permettent cependant de s'affranchir de la notion de VLPs et d'affirmer la présence de phages ou de virus eucaryotes au sein des écosystèmes, ce qui ne peut pas être permis avec le travail jusqu'à présent réalisé pour cette thèse. Nous prévoyons cependant de réaliser un séquençage des viromes fécaux dans le but de lever le voile sur l'identité des VLPs détectées.

Le protocole de séquençage des viromes fécaux nécessite cependant d'être optimisé par rapport à ce qui est couramment réalisé actuellement. Des expériences d'ajout de phages à ADN simple brin et en quantités connues réalisées au sein de l'équipe (Ilias Theodorou, résultats non publiés) ont mis en avant une amplification plus ou moins importante des génomes phagiques en fonction de leur nature nucléique et la structure de leur ADN au cours de la confection de bibliothèques en vue du séquençage des viromes. Un projet visant à la résolution de ce problème est actuellement en cours au sein de notre équipe.

Il est également nécessaire de souligner que nos résultats de quantifications absolues de VLPs puissent être sous-estimés du fait de la difficulté à détecter les *Microviridae*, famille représentant près de 31% des phages du microbiote intestinal (Shkoporov et al. 2019), par microscopie par épifluorescence. Leur génome, constitué d'environ 4k pnb d'ADN simple brin, est peu marqué par le Sybr Gold et voit son intensité de fluorescence proche de celle du bruit de fond dans les filtrats

fécaux. Ainsi, dans les meilleures conditions de manipulation, nous estimons que seulement un tiers de phages Φ X174, représentants de cette famille, peuvent être détectés par la méthode de dénombrement des VLPs utilisée au cours de ma thèse (**Figure 31** Erreur ! Source du renvoi introuvable.**B**). De plus, les virions des *Inoviridae*, dont le génome est réparti tout au long de leur manteau protéique, devraient ne pas être détectables par cette technique. Malgré ces limites concernant la fiabilité des quantifications absolues en VLPs dans les selles, la comparaison relative des différents groupes reste possible. L'observation de particules virales par MET dans les échantillons les plus concentrés en VLPs permet cependant de conforter nos résultats de quantification. Il est également à noter que nos résultats de dénombrement ne tiennent compte ni ne reflètent les populations de prophages présents au sein du microbiote fécal. Une augmentation de la proportion des caudophages chez les patients par rapport aux individus sains a par ailleurs été précédemment rapportée (Clooney et al. 2019). Ces phages étant facilement détectables par microscopie par épifluorescence, il est possible qu'une partie de l'augmentation des VLPs au sein de notre groupe de patients en poussée soit due à une abondance augmentée des représentants des caudophages. À l'inverse, aucune corrélation n'a pu être établie entre le nombre de VLPs détectées par microscopie par épifluorescence et la quantité d'ADN mesurée par le QuBit™ dans les extraits d'ADN réalisés en prévision du séquençage. Au contraire de l'hypothèse précédente, une augmentation du nombre de Microviridae pourrait passer inaperçue, dans la mesure où le Qubit™ n'est pas optimal pour quantifier les ADN à simple brin. De plus, une augmentation du nombre de virions des *Microviridae*, dont les génomes sont parmi les plus petits chez les phages, pourrait être masquée par d'autres phages intestinaux dont les génomes peuvent atteindre des masses jusqu'à 270 fois plus élevées dans le cas des Lak-phages (Devoto et al. 2019).

3.5. Dénombrement et caractérisation des VE dans les selles

3.5.1. Détermination de la nature des VE dans les selles d'individus sains grâce à la métaprotéomique

Nous avons souhaité comparer la diversité protéique portée par les VE, qu'elles soient d'origine eucaryote ou procaryote, entre contrôles et patients. L'objectif de cette entreprise était d'une part de s'assurer de la présence de VE dans les filtrats fécaux et dans les fractions de VE purifiées, et d'autre part de déterminer les populations productrices de VE chez les individus sains et

les patients. Nous nous sommes donc tournés vers des analyses métabolomiques d'échantillons issus des filtrats fécaux.

La présence de protéines d'origine autre que vésiculaire, comme les protéines de flagelles peuvent, si elles sont trop nombreuses, masquer les protéines des vésicules au cours de l'analyse. Afin d'obtenir une profondeur d'analyse optimale, les VE ont été purifiées selon le protocole présenté dans l'article 1, de façon à s'affranchir des protéines libres et des protéines appartenant à d'autres nanoparticules (flagelles, agrégats, phages...) présentes dans les filtrats fécaux. Ce travail nous a également permis de nous assurer de la qualité de nos fractions purifiées de vésicules et de la présence de VE bactériennes et eucaryotes dans les fractions 2 des gradients d'iodixanol, en plus de renseigner des indications sur la provenance des VE.

3.5.1.2. Préparation des fractions étudiées

3.5.1.2.1. Intérêt de l'étape de centrifugation

Le protocole de purification détaillé dans l'article 1 a été suivi pour purifier des VE à partir de filtrats fécaux d'individus sains. Brièvement, les filtrats fécaux sont soumis à une longue centrifugation (16h, 20 000 g) permettant l'obtention d'un culot, dont les nanoparticules sont ensuite purifiées sur gradient d'iodixanol. En plus de permettre la concentration des nanoparticules fécales, cette étape de centrifugation débarrasse, en théorie, les échantillons d'une certaine proportion de leurs protéines libres, qui ne culottent pas avec les nanoparticules du fait de leur plus faible densité. Cependant, cette durée et vitesse de centrifugation ne sont pas communes et l'une des méthodes principalement utilisées par la communauté pour la purification des VE se base sur des colonnes de chromatographie par exclusion de taille (SEC).

Afin d'être certain d'utiliser le protocole permettant la meilleure purification des VE, les ratios nanoparticules/ μ g de protéines ont été comparés sur un filtrat fécal soumis à cette étape de centrifugation ou passé sur une colonne SEC (cf. Matériel et Méthodes).

Les mêmes mesures ont été réalisées dans le culot resuspendu issu d'un filtrat fécal provenant d'une même selle, mais concentré par centrifugation (20 000 g, 16h) et dont le culot a été resuspendu dans du Tris 10 mM.

La centrifugation a permis de recueillir 67% des nanoparticules initialement présentes dans le filtrat fécal (**Figure 35A**). Cette valeur est nettement supérieure au nombre de nanoparticules récoltées dans les fractions d'intérêt issues de la colonne SEC (12%). En effet, la majorité des

nanoparticules se retrouvent dans le liquide de lavage. Cela peut être dû à une mauvaise optimisation du protocole ou à une détection limitée des petites VE par le Videodrop. Cependant, la pureté des échantillons, en termes de nanoparticules par mg de protéines, est deux fois supérieure dans les culots de centrifugation que dans les fractions d'intérêt obtenues suite à la chromatographie d'exclusion de taille (**Figure 35B**). Cet argument, ajouté au gain de temps et à la praticité non négligeable offerte par la centrifugation, a conduit à l'emploi de cette étape de centrifugation pour concentrer les nanoparticules des filtrats fécaux et les débarrasser d'une partie des protéines libres présentes dans ces derniers.

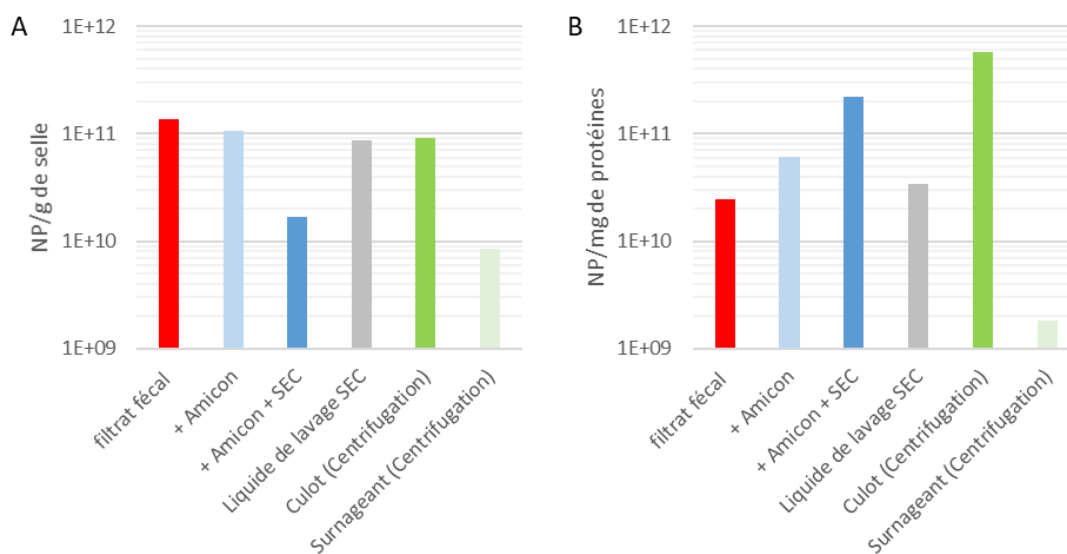


Figure 35. Comparaison de la SEC et de la centrifugation

Nombre de nanoparticules/g de selle (A) et de nanoparticules/mg de protéines (B) dans le filtrat fécal (rouge), après passage du filtrat fécal sur filtres Amicon 100K (bleu clair), puis dans les fractions d'intérêt (bleu foncé) et dans le liquide de lavage (bleu foncé) issus de la colonne SEC ; ou dans le culot (vert foncé) et le surnageant (vert clair) issus de la centrifugation (16h, 20 000 g) appliquée au filtrat fécal. Les dénombrements de nanoparticules ont été mesurés par le Videodrop et les masses de protéines par le QuBit. Ces mesures n'ont pas été réalisées en répliquats.

3.5.1.2.2. Rendement des gradients d'iodixanol pour la purification de VE

La fraction 2 issue des gradient d'iodixanol étant en théorie celle qui contient les VE les plus purifiées, elle a été sélectionnée pour la suite des analyses de protéomique. Des mesures de dénombrement ont été réalisées par le Videodrop, afin d'estimer les quantités de nanoparticules retrouvées dans chaque fraction, en particulier dans la fraction 2, afin d'estimer la proportion de VE

perdues au cours de cette étape (**Figure 36**). En moyenne, 82% des nanoparticules détectées dans les filtrats fécaux sont retrouvées dans le culot de centrifugation (16h, 20 000 g). 80% des nanoparticules détectées dans le culot sont également retrouvées dans l'une des cinq fractions obtenues à l'issue du gradient. Si la fraction 4 est la plus concentrée en nanoparticules, celle-ci voit s'y côtoyer VE, virus, flagelles et agrégats. Ce n'est pas le cas de la fraction 2, qui, d'après mes observations en MET, ne semble contenir que des VE (**Figure 37**). Bien que le nombre de nanoparticules soit relativement faible dans la fraction 2 (seulement 12% des nanoparticules détectées dans les filtrats fécaux le sont également dans la fraction 2), elle bénéficie d'une grande pureté apparente sur les grilles de MET. En effet, aucun virus ni flagelle n'a pu formellement être identifié sur les grilles observées. De rares agrégats, de nature indéterminée, ont cependant parfois pu être observés. Les observations des fractions issues du gradient d'iodixanol réalisées en MET semblent être en accord avec les concentrations de nanoparticules mesurées par le Videodrop au sein de celles-ci.

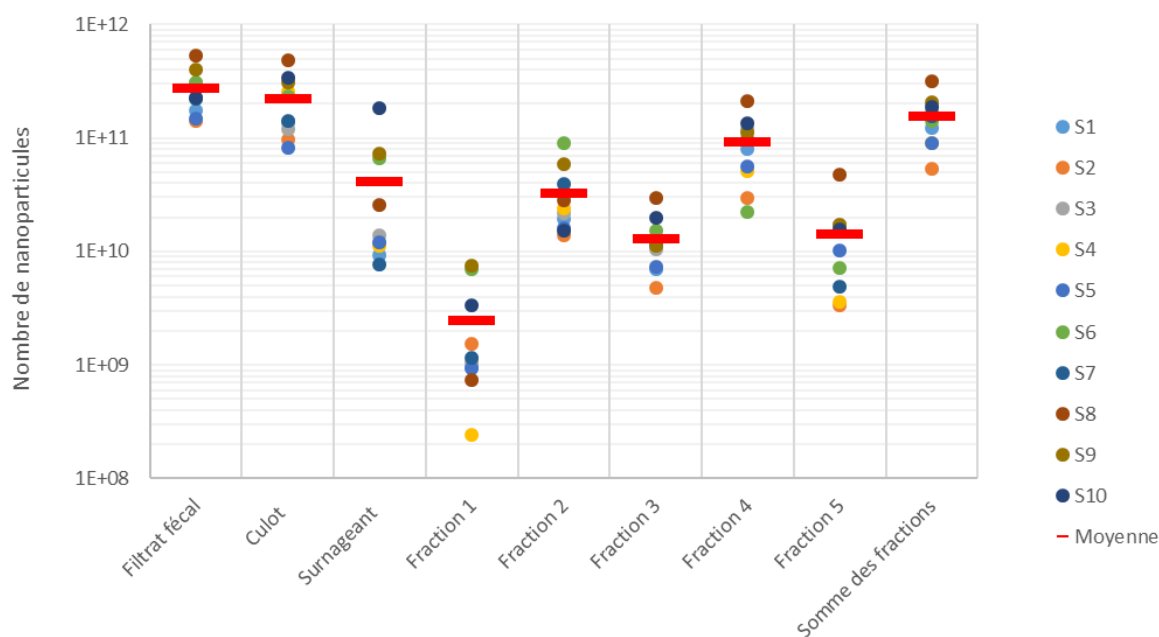


Figure 36. Quantités de nanoparticules au cours du processus de purification des VE

Nombre de nanoparticules par gramme de selle, déterminés par le Videodrop. Ces mesures ont été réalisées à partir de filtrats fécaux issus de selles d'individus sains (S1 à S10). La barre horizontale rouge correspond à la moyenne des mesures pour chaque fraction étudiée.

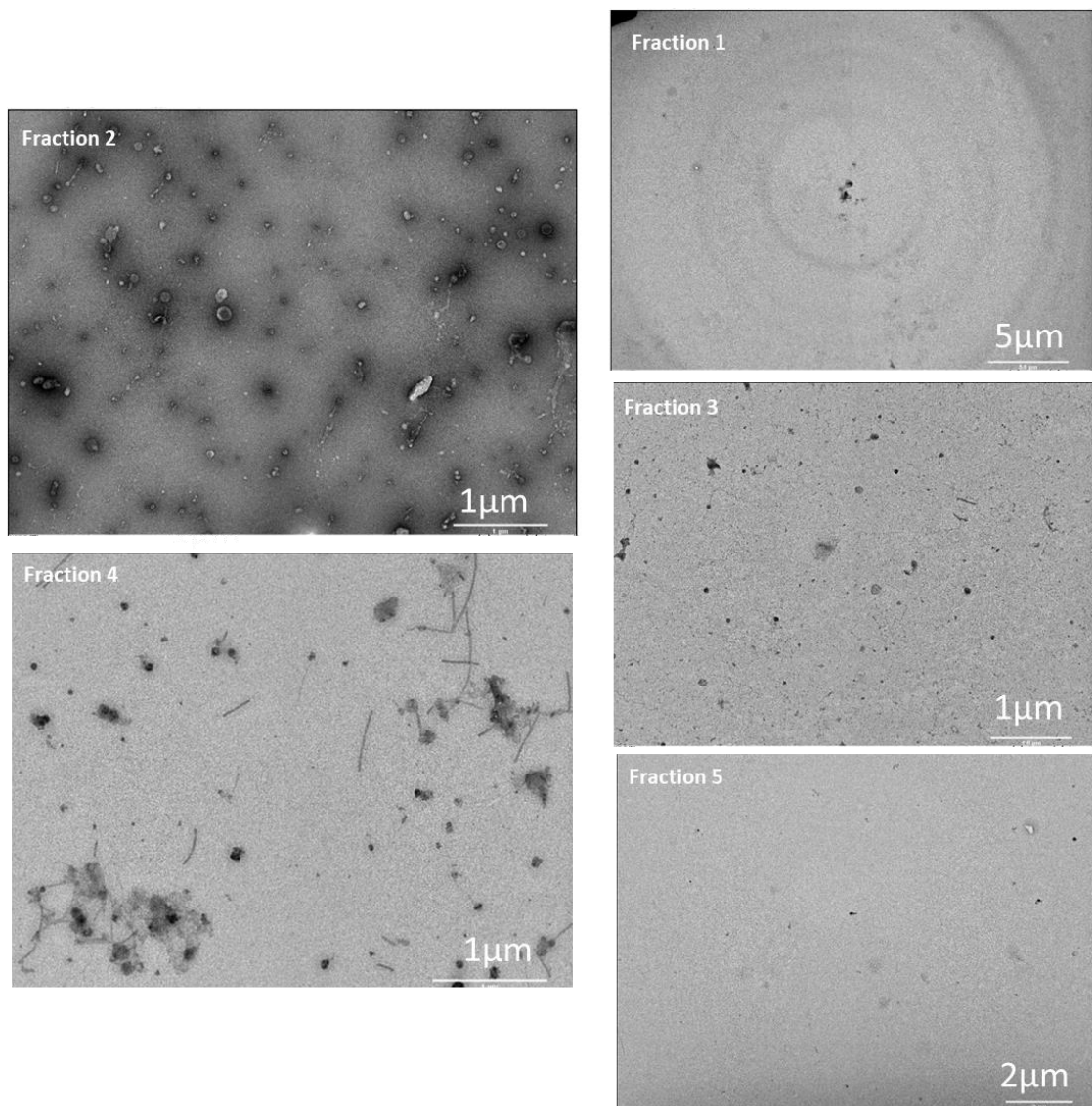


Figure 37. Observation des fractions issues du gradient d'iodixanol par MET

Une goutte de 10 μL de chaque fraction a été déposée sur une grille carbonée, puis colorée à l'acétate d'uranyle, avant d'être observée en MET. Sur la grille correspondant à la fraction 1, aucun objet n'a pu être observé, suggérant une concentration basse voire inexistante de nanoparticules au sein de cette fraction. Dans la fraction 2, de nombreuses VE sont visibles, ainsi que quelques agrégats de nature indéterminée. Dans la fraction 3, des nanoparticules de nature indéterminée ont été observées. Malgré une coloration positive, des particules ont pu être également observées dans la fraction 4, au sein de laquelle se trouvent des flagelles et de potentielles VE. Celles-ci sont difficiles à différencier des virus, en théorie également présents dans cette fraction. Enfin, aucune particule n'a été détectée au sein de la fraction 5.

3.5.1.3. Clustering protéomique des échantillons

Bien que les fractions 2 issues du gradient d'iodixanol soient les plus purifiées en VE, nous nous sommes demandé si, d'une part, cette affirmation était confirmée par analyse métabotéomique et si, d'autre part, certaines populations de vésicules n'étaient pas perdues au cours du gradient de densité. Pour ce faire, les fractions 2 et les culots post-centrifugation de trois filtrats fécaux (S7, S9, S10) ont été confrontés. Ils ont été soumis à la spectrométrie de masse selon le protocole détaillé ci-après (cf. Matériel et Méthodes) et les résultats bruts ont été traités selon le mode itératif. L'ensemble des analyses décrites métabotéomiques décrites ci-après ont été réalisées par Marianne De Paepe.

Le premier contrôle réalisé fut de s'assurer de la cohérence entre les échantillons provenant des mêmes filtrats fécaux. Pour ce faire, les échantillons ont été clusterisés sur la base de leur contenu en termes de sous-groupes, c'est-à-dire de protéines identifiées avec le même set de peptides. Les groupes, quant à eux, correspondent aux ensembles de sous-groupes possédant au moins un peptide en commun. L'arbre ainsi obtenu (**Figure 38** *Erreur ! Source du renvoi introuvable.*) permet de confirmer que les paires d'échantillons se ressemblent, comme attendu.

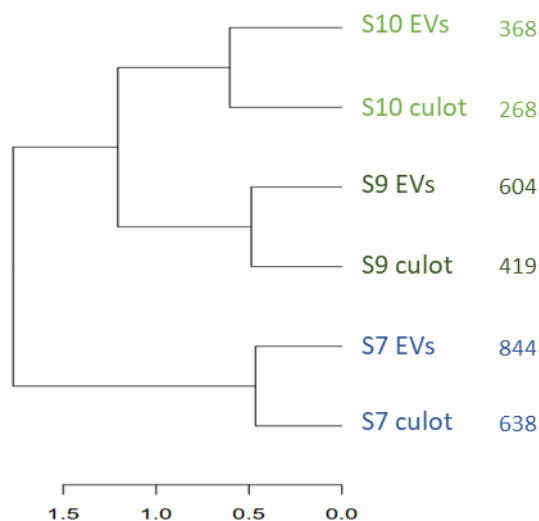


Figure 38. Clustering des échantillons selon leur contenu, en sous-groupes

3.5.1.4. Origine des protéines bactériennes

Les localisations subcellulaires des protéines bactériennes identifiées ont été déterminées par pSORTB, à partir de leurs séquences fasta. Une importante différence des répartitions des protéines en fonction de leur localisation est observée entre les culots et les fractions purifiées (**Figure**

39Erreur ! Source du renvoi introuvable.) Cela s'explique par la présence de protéines de flagelles majoritaires dans les culots et bien moins importante dans les fractions purifiées de VE, du fait de leur densité plus élevée (**Figure 19**). Ce résultat corrobore les observations réalisées par MET, qui montraient déjà une importante quantité de flagelles dans les culots de filtrats fécaux. Un enrichissement des protéines bactériennes a été observé dans les fractions purifiées, par rapport aux culots et justifie une étape de purification des VE par gradient d'iodixanol pour leur analyse protéomique. La majorité des protéines identifiées dans la fraction de VE purifiées est bien associée à une localisation potentiellement vésiculaire.

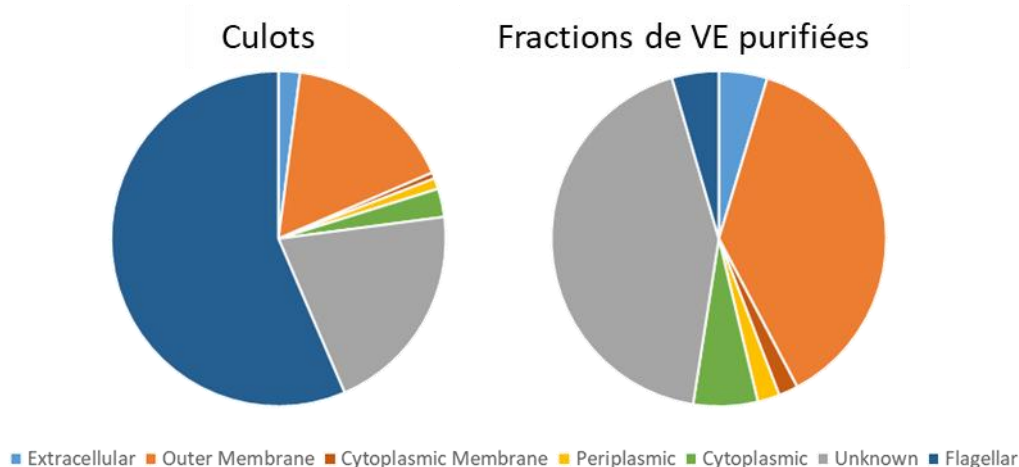


Figure 39. Proportion des spectres en fonction de l'origine de leurs protéines bactériennes associées

Une analyse metaprotR a été initiée dans le but d'identifier les genres bactériens représentés parmi les protéines identifiées dans la fraction de VE purifiées du filtrat fécal de l'individu sain S2 (**Figure 40Erreur ! Source du renvoi introuvable.**). La distribution ainsi obtenue est dominée par une proportion majoritaire de protéines issues des *Bacteroidetes*, phylum bactérien parmi les plus représentés au sein du microbiote intestinal humain (Arumugam et al. 2011, Sabino et al. 2016). La famille *Faecalibacterium*, pourtant l'une des plus abondantes en termes de répartition bactérienne dans le microbiote intestinal, est moins fortement représentée en termes de spectres associés à des protéines au sein des fractions de VE purifiées. La structure de la membrane de ces bactéries à Gram positif en est probablement responsable, dans la mesure où leur absence de la membrane externe limite leurs voies de production des VE par rapport aux bactéries à Gram négatif. L'analyse protéomique d'échantillons supplémentaires sera nécessaire à la validation de ces résultats.

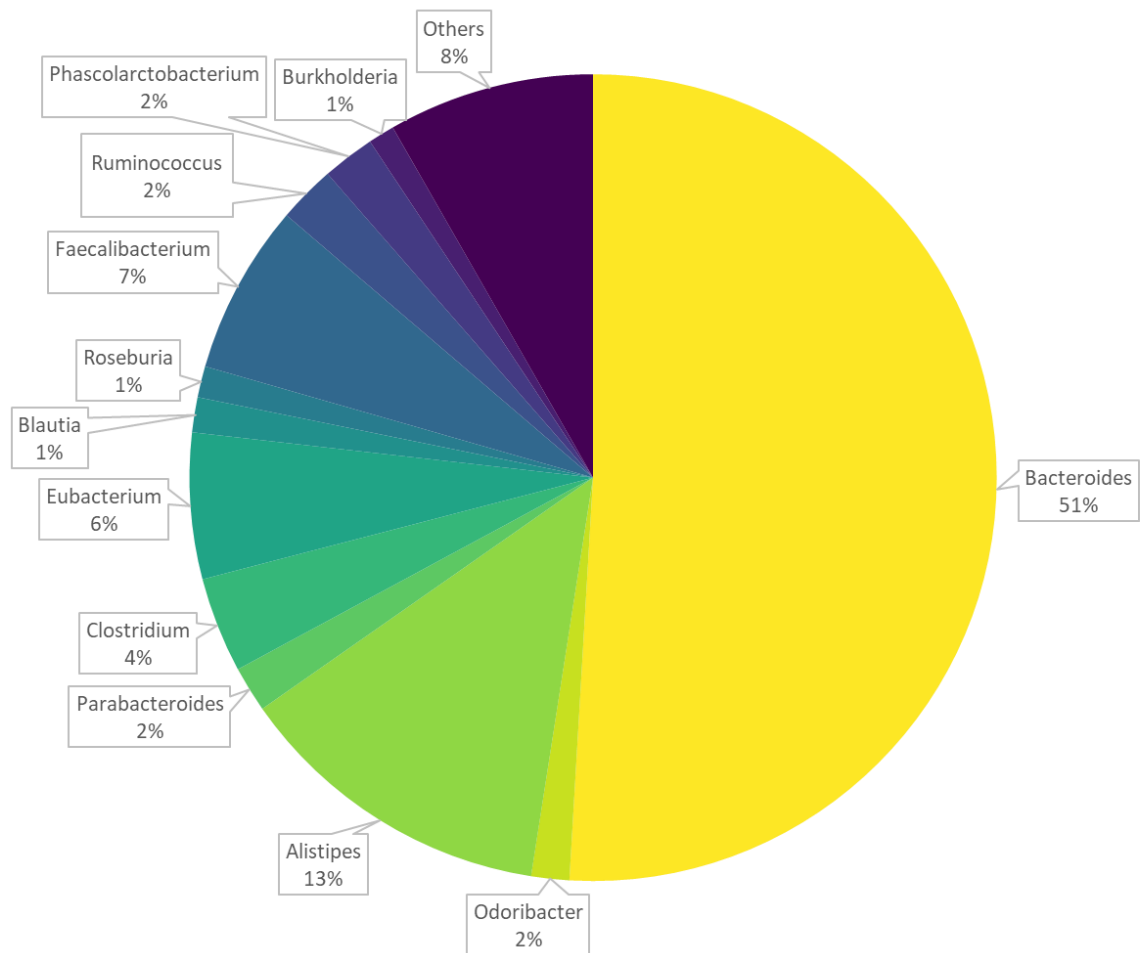


Figure 40. Répartition des protéines en genres bactériens dans la fraction purifiée de VE le filtrat fécal de l'individu sain S2

Répartitions réalisées par metaprotR.

3.5.1.5. Discussion

Les résultats présentés ci-dessus suggèrent fortement une meilleure profondeur d'analyse des protéines vésiculaires dans les fractions de VE purifiées que dans les culots issus des filtrats fécaux. Ceci est surtout dû à l'élimination de la majorité des protéines flagellaires au cours du processus de purification. De plus, la proportion de spectres associés à des protéines de la membrane externe est similaire à la répartition attendue chez des vésicules issues de bactéries à Gram négatif (Lee, Kim, and Gho 2016). C'est pour ces raisons que les études de métabotéomique à suivre devront être réalisées sur des fractions purifiées de VE, malgré le caractère chronophage de leur préparation.

Ces travaux de mise au point et ces premiers résultats pavent la voie vers de nouveaux projets de métabotéomique orientés vers l'étude des VE. Il pourrait en effet être pertinent de comparer les

compositions protéiques des fractions de VE purifiées issues de selles d'individus sains et de patients atteints de la maladie de Crohn, afin d'éventuellement y détecter des modifications de la proportion des protéines eucaryotes et procaryotes. Compte tenu de l'augmentation de la concentration en VLPs observée chez les patients en phase de poussée et de l'inflammation intestinale présente chez ces individus, il serait en effet envisageable d'observer une modification de la quantité de protéines vésiculaires eucaryotes due à l'inflammation. D'autre part, l'étude de l'origine bactérienne des protéines issues de ces VE pourrait également permettre de révéler des modifications des bactéries productrices de VE dans le cadre de la maladie, ainsi que la prévalence augmentée de vésicules bactériennes directement issues de la lyse par les phages.

Enfin, le séquençage des bactéries présentes dans les selles semble également être nécessaire pour une meilleure compréhension des acteurs de la production bactérienne de VE dans le microbiote intestinal, qu'il soit en état inflammatoire ou non.

3.5.2. Dénombrement des VE dans les selles

3.5.2.1. Utilisation du Videodrop pour le dénombrement de VE les selles humaines

En plus des comparaisons présentées dans l'article 1 concernant le dénombrement de VE purifiées à partir de cultures bactériennes et de microbiotes murins, le Videodrop a été comparé au ZetaView et au NanoSight pour le dénombrement de VE isolées de selles humaines.

Les VE des selles des individus sains S1, S5 et S8 ont été purifiées selon le protocole présenté dans l'article 1 et dénombrées par le Videodrop, le ZetaView et le NanoSight (**Figure 41**). Un maximum d'un facteur 2,5 entre les concentrations données par les différents appareils a été observée, avec celle déterminée par le Videodrop étant la plus basse dans deux échantillons sur trois. Comme discuté dans l'article 1, et les fractions observées ici étant par nature de tailles hétérogènes, il est probable que le ZetaView et le NanoSight soient affectés par l'effet d'écran conduisant un défaut de détection des plus petites VE, masquées par d'autres, plus grosses. Pour sa part, si le Videodrop est en théorie moins affecté par l'effet d'écran, il souffre d'une limite de détection plus haute concernant la taille des VE.

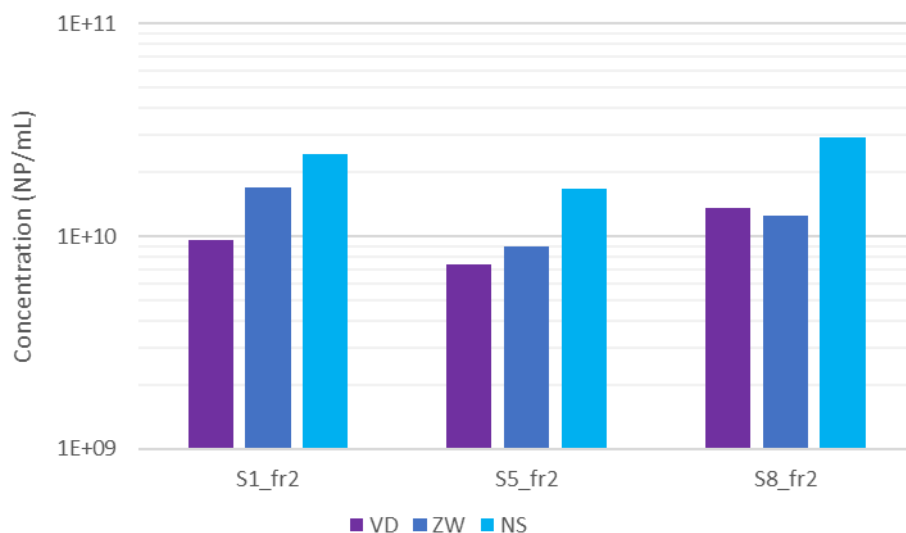


Figure 41. Concentrations en NP dans les fractions 2 (VE purifiées) de selles d'individus sains

Concentrations en NP/mL, selon le Videodrop (VD, violet), le ZetaView (ZW, bleu foncé) et le NanoSight (NS, bleu clair).

3.5.2.2. Intérêt de la protéinase K pour le dénombrement des VE dans les filtrats fécaux

Bien que l'utilisation du NP-40 permette la purification partielle des phages mélangés à des VE (cf. 2.2.2. Impact des composés utilisés sur la concentration en phages et en), cette méthode ne peut pas s'appliquer à la quantification de VE au sein de filtrats fécaux. Cela est dû à plusieurs raisons. Premièrement, comme expliqué plus haut, un peu plus de 20% des VE détectables par le Videodrop sont résistantes au NP-40. Or, dans les filtrats fécaux, nos clichés de MET suggèrent que les VE puissent être plus nombreuses que les phages (cf.3.3. Préparation des filtrats fécaux) et donc que leur nombre, même diminué par le traitement par le NP-40 soit encore suffisamment significatif pour fausser le dénombrement des phages. De plus, l'observation des filtrats fécaux en MET a révélé la présence, en plus des phages et des VE, de fragments de flagelles et d'agrégats de nature inconnue, probablement protéique et lipidique. Enfin, contrairement à ce que j'espérais au début de ma thèse, les avancées technologiques concernant le Videodrop n'ont pas été suffisantes pour permettre la détection de phages dont le diamètre hydrodynamique est inférieur à 100 nm, ne permettant donc pas le dénombrement des phages du microbiote intestinal.

Fort de ce constat, il a fallu essayer de trouver un protocole opposé à celui du traitement par le NP-40, à savoir éliminer le plus de nanoparticules possibles à part les VE dans les filtrats fécaux, puis quantifier ces dernières à l'aide du Videodrop. La protéinase K, une enzyme clivant les liaisons peptidiques aux extrémités carboxyles, a été sélectionnée à cet effet. Utilisée notamment pour

l'élimination de flagelles au sein de préparations issues de cultures bactériennes (Yu et al. 2021), nous nous sommes demandé si cette enzyme pouvait éliminer également les phages, dans l'éventualité de découvrir des filtrats fécaux au sein desquels le nombre de phages serait supérieur à celui des VE. En effet, en théorie, la protéinase K devrait être capable de détruire les capsides protéiques des phages, ainsi que les agrégats non identifiés présents dans les filtrats fécaux ; tout en épargnant les membranes lipidiques des VE.

3.5.2.2.1. Impact de la température

Traditionnellement utilisée à 37°C pour une activité optimale (Bajorath, Saenger, and Pal 1988), des expérimentations ont été réalisées afin de déterminer s'il était possible de diminuer la température d'incubation lors du traitement par la protéinase K. En effet, il était supposé que les VE issues des filtrats fécaux puissent être sensibles à cette température. Dans un premier temps, une incubation à 30°C a été testée.

Des VE purifiées de *Faecalibacterium prausnitzii* et de *Bacillus subtilis* ont été soumises à un traitement par la protéinase K (20 µg/mL, Macherey Nagel) pendant 6h, à 30°C et à 37°C. En parallèle, les mêmes échantillons ont été soumis à des incubations identiques, mais sans traitement par la protéinase K. La concentration dans les échantillons a été relevée au début et à la fin de l'expérience à l'aide du Videodrop. 91% des VE de *Faecalibacterium prausnitzii* étaient encore détectables après le traitement et l'incubation à 37°C, contre 70% pour l'échantillon soumis à l'incubation à 30°C. Concernant les VE de *Bacillus subtilis*, une tendance similaire a été observée : 75% étaient encore détectables après le traitement et l'incubation à 37°C, contre 65% pour l'échantillon soumis à l'incubation à 30°C.

Cette baisse conséquente du nombre de VE détectées par le Videodrop nous apporte une information importante : certaines VE sont plus sensibles à une incubation à 30°C qu'à 37°C. Nous avons donc décidé de conserver cette incubation à 37°C. Il faudra cependant tenir compte de la perte d'une partie des VE lors de l'incubation à cette température lors de la quantification des VE dans les filtrats fécaux.

3.5.2.2.2. Effets de la protéinase K sur les flagelles

L'efficacité de la protéinase K a d'abord été testée sur les flagelles. Nous avons tenté de reproduire des résultats précédemment décrits (Yu et al. 2021). Pour ce faire, nous avons voulu

produire des flagelles purifiés à partir d'une culture d'*E. coli* MG1655, en suivant un protocole précédemment décrit (Sherman, Soni, and Yeger 1988). Malheureusement, l'observation en MET des fractions supposées contenir les flagelles n'a révélé que la présence de petite VE. Dans un second temps, nous avons tenté de purifier des flagelles des fractions 4 des gradients d'iodixanol des filtrats fécaux S3 et S8 (S3_4 et S8_4). L'observation en MET des fractions re-purifiée a révélé la présence de flagelles dans celles-ci, mais également de nombreuses VE. N'arrivant pas à purifier des flagelles, nous nous sommes tournés vers nos échantillons les plus concentrés en ceux-ci afin d'observer les effets de la protéinase K sur ces objets.

Le filtrat fécal de souris conventionnelles (préparé selon le protocole décrit dans l'article 1) et la fraction virale de S2 (filtrat fécal de l'individu Sain 2), particulièrement riches en flagelles, ont été traités avec de la protéinase K à une concentration finale de 20 µg/mL pendant 6h à 37°C. Les échantillons avant et après traitement ont été observés par MET (**Figure 42**~~Erreur ! Source du renvoi introuvable.~~). La majorité des flagelles semblent avoir été éliminés par la protéinase K et bien que la MET ne soit pas une approche quantitative, il semblerait que les VE initialement présentes dans les échantillons le soient encore après le traitement.

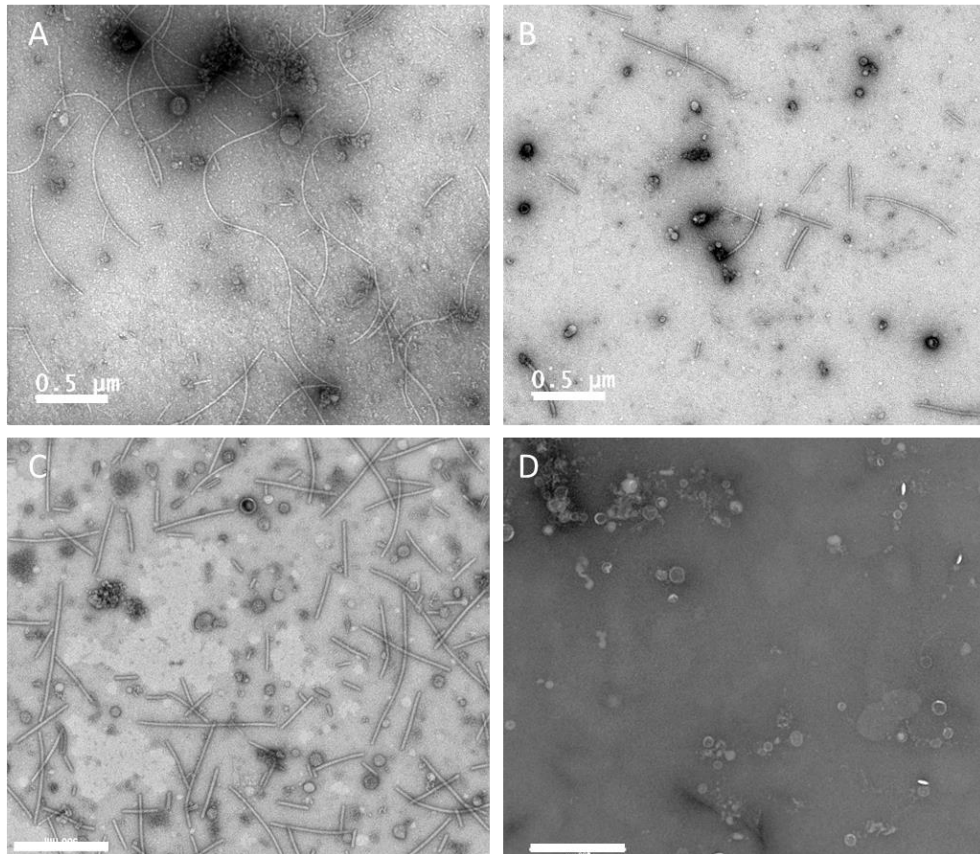


Figure 42. Effets de la protéinase K sur les flagelles

Observations réalisées en MET, après dépôt des échantillons sur grilles carbonées et coloration négative par l'acétate d'uranyle. (A) Filtrat fécal de souris conventionnelles. (B) Filtrat fécal de souris conventionnelle après traitement par la protéinase K. (C) Échantillon S2_4. (D) Échantillon S2_4 après traitement par la protéinase K. Toutes les barres d'échelle représentent 500 nm.

3.5.2.2.3. Effets de la protéinase K sur les phages

Contrairement aux flagelles, de nombreux échantillons de phages purifiés sont à disposition et plusieurs techniques permettant leur quantification. L'étude des effets de la protéinase K sur ces derniers, qu'il serait souhaitable d'éliminer afin d'énumérer uniquement les EVs dans les filtrats fécaux, s'en est ainsi trouvée plus simple. Les effets de la protéinase K sur la concentration de phages purifiés connus, à savoir T4 et P1, ont été mesurés à l'aide du Videodrop et ont mis en évidence une chute d'environ 70% de la concentration (**Figure 43**). Cette diminution s'est révélée encore plus importante dans les filtrats fécaux de patients en poussée, dans lesquels la concentration en VLPs a été estimée par microscopie par épifluorescence (cf. 3.4.2. Dénombrement des phages dans les filtrats fécaux **Erreur ! Source du renvoi introuvable.**), puisque jusqu'à 97% de VLPs n'ont plus été détectées après le traitement.

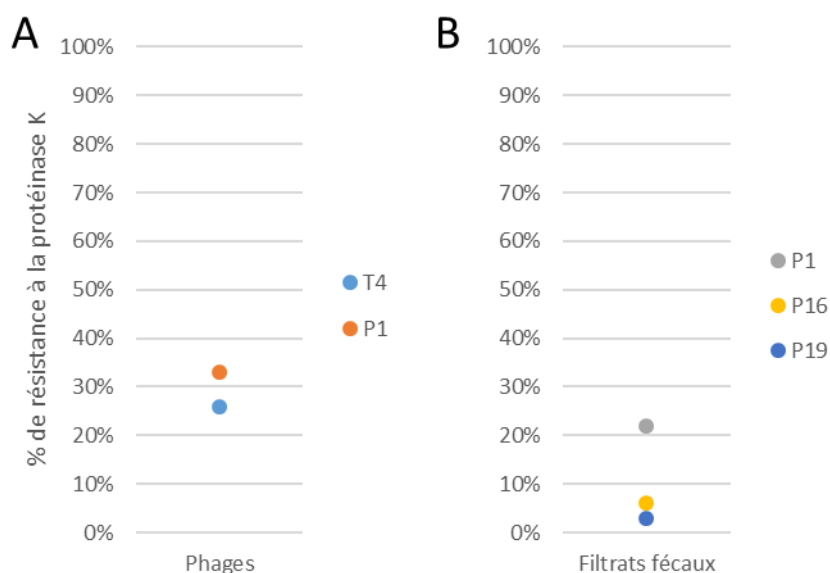


Figure 43. Proportions de phages et de VLPs résistants à la protéinase K

Proportions de (A) phages purifiés et (B) VLPs détectées dans des filtrats fécaux de trois patients en poussée résistants à un traitement à 20 µg/mL de protéinase K, après incubation de 6h à 37°C. Les concentrations en phages ont été mesurées par le Videodrop, tandis que les concentrations en VLPs dans les filtrats fécaux ont été obtenues par microscopie par épifluorescence, après marquage par le Sybr Gold.

Ces résultats sont en accord avec des études publiées sur les effets de la protéinase K sur les phages. La destruction de plus de 80% des phage (Sieber, Pluckthun, and Schmid 1998), ainsi que la diminution de l'infectiosité du phage MS2 de plus d'un facteur 10 après traitement par la protéinase K (Langlet et al. 2018) avaient en effet déjà été publiés.

3.5.2.2.4. Effets de la protéinase K sur les VE

Tout comme pour les phages, je dispose d'échantillons de VE purifiées utilisables pour évaluer l'efficacité de la protéinase K sur celles-ci. Étant principalement composées d'une couche lipidique dans laquelle s'insèrent diverses protéines, il peut raisonnablement être attendu que la protéinase K n'affecte que peu les vésicules. C'est en effet ce qu'il se produit dans les échantillons de VE purifiées issues de cultures bactériennes (et dont le protocole de purification est détaillé dans l'article 1) (**Figure 44** Erreur ! Source du renvoi introuvable.). La résistance des VE à la protéinase K est un phénomène déjà connu (Tabak et al. 2021). Cependant, d'après plusieurs travaux (Amosse et al. 2018, Shelke et al. 2014), la protéinase K ne serait pas sans effet sur les vésicules puisqu'elle raserait les protéines présentes à leur surface, sans toutefois atteindre celles plongées au cœur de leur couche

lipidique. Si certains de leurs effets s'en voient ainsi diminués par l'action de la protéinase K, leur quantification n'en est pas affectée pour autant.

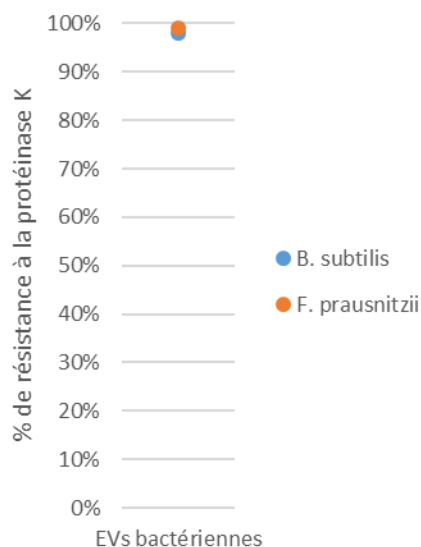


Figure 44. Proportions de VE résistantes à la protéinase K

Proportions de VE bactériennes purifiées résistantes à un traitement à 20 µg/mL de protéinase K, après incubation de 6h à 37°C. Les concentrations en VE ont été mesurées avec le Videodrop.

3.5.2.2.5. Effets de la protéinase K sur les filtrats fécaux

3.5.2.2.5.1. Impact de la durée du traitement

Afin de minimiser l'impact de l'incubation sur les VE mais également dans un souci de praticité, la durée de l'incubation a été étudiée. Plusieurs filtrats fécaux issus de selles d'individus sains ou de patients en période de rémission ont été soumis au traitement par la protéinase K (**Figure 45** **Erreur ! Source du renvoi introuvable.**). Les concentrations en nanoparticules ont été relevées à différents points de temps par le Videodrop. De la même façon que pour les VE bactériennes purifiées, les contrôles négatifs réalisés nous confirment que l'incubation est délétère pour une fraction des nanoparticules détectées par le Videodrop. De plus, qu'il s'agisse de filtrats fécaux de patients ou d'individus sains, la concentration en nanoparticules se stabilise après environ deux heures d'activité de la protéinase K. L'incubation de 2h, à 37°C, a donc été retenue pour le traitement par la protéinase K dans les filtrats fécaux. Le nombre de nanoparticules ne variant plus après 2h d'incubation, il a été considéré que passé ce délai, la protéinase K n'agissait plus sur les nanoparticules présentes dans les filtrats fécaux détectables par le Videodrop (à savoir certaines VE,

virus, flagelles et agrégats). Par conséquent, les contrôles positifs réalisés ci-dessus sur les phages et les VE purifiés n'ont pas été reproduits avec cette durée d'incubation.

Une diminution de la concentration en nanoparticules détectées par le Videodrop a été observé suite à l'incubation des filtrats fécaux d'individus sains, mais pas chez ceux des patients (**Figure 45**). Une différence de nature des nanoparticules détectées entre les groupes d'individus étudiés pourrait être à l'origine de ce phénomène.

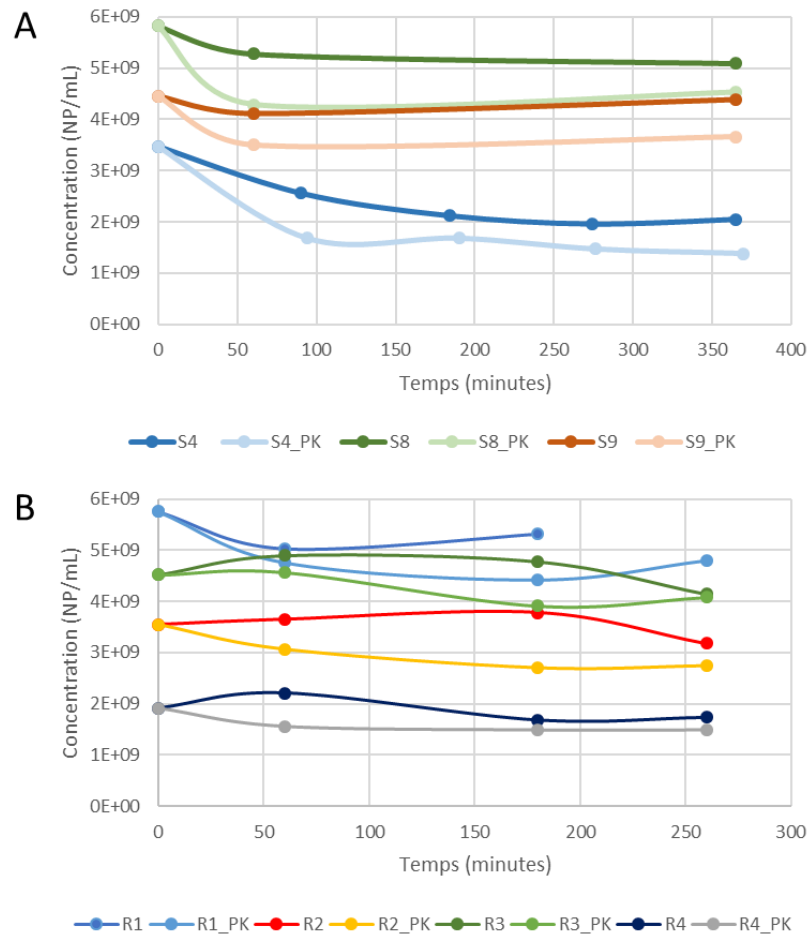


Figure 45. Effet de la durée de l'incubation lors du traitement de filtrats fécaux par la protéinase K

Concentration en nanoparticules détectées par le Videodrop dans des filtrats fécaux issus (A) d'individus sains et (B) de patients Crohn en période de rémission.

3.5.2.2.5.2. Observation de filtrats fécaux après traitement par MET

L'action de la protéinase K dans les filtrats fécaux a également été évaluée visuellement en MET (**Figure 46***Erreur ! Source du renvoi introuvable.*). En accord avec un constat précédent (3.5.2.2.2. Effets de la protéinase K sur les flagelles), la majorité des flagelles semblent détruits par la protéinase K. Les VE paraissent préservées malgré le traitement. La présence d'agrégats sur les grilles après le traitement a cependant été constatée. Il pourrait s'agir essentiellement d'agrégats lipidiques dans la mesure où la protéinase K peut agir sur des agrégats protéiques à la concentration utilisée (Swietnicki et al. 2000, Tzaban et al. 2002). Ces agrégats pourraient également provenir de gros agrégats protéiques n'ayant pas entièrement digérés par la protéinase K. N'ayant pas d'échantillon d'agrégats purifiés, cette hypothèse ne peut pas être validée.

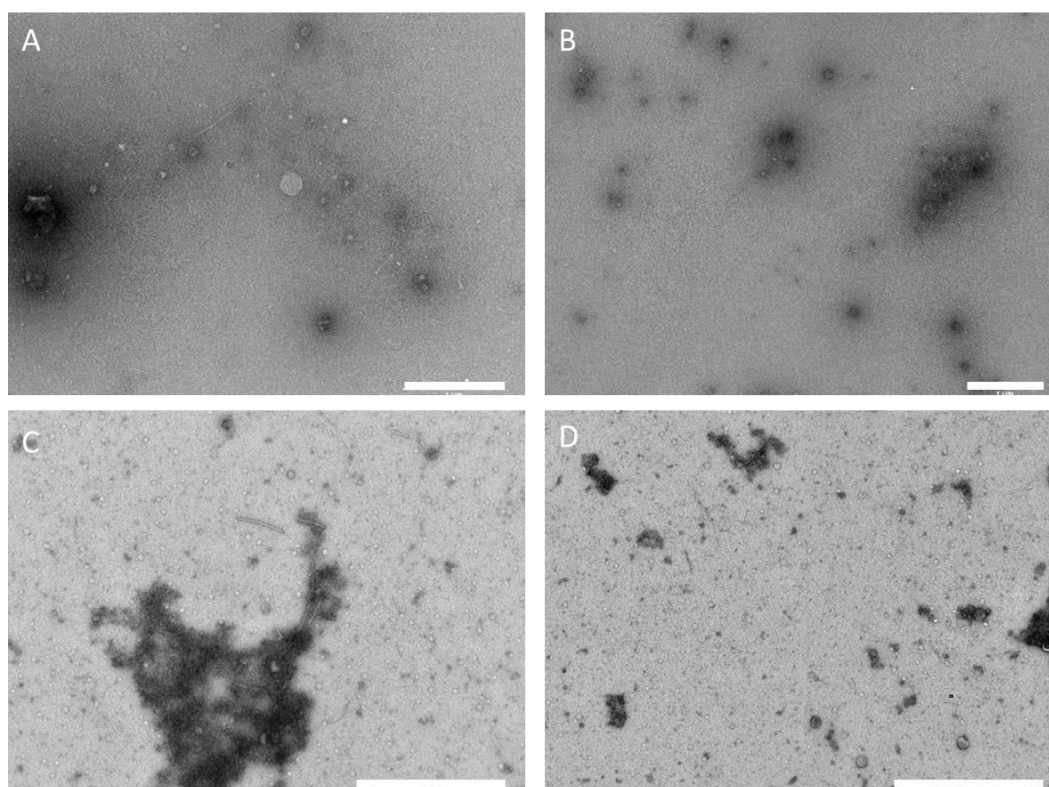


Figure 46. Effets de la protéinase K sur les filtrats fécaux

Observations réalisées en MET, après dépôt des échantillons sur grilles carbonées et coloration négative par l'acétate d'uranyle. (A) Filtrat fécal S3. (B) Filtrat fécal S3 après traitement par la protéinase K. (C) Filtrat fécal S7. (D) Filtrat fécal S7 après traitement par la protéinase K. Toutes les barres d'échelle représentent 1000 nm.

Malgré la présence certaine de quelques agrégats et flagelles résistants au traitement des filtrats fécaux par la protéinase K, l'observation d'une tendance entre la proportion de VE observées en MET et le pourcentage de nanoparticules résistantes au traitement semble conforter l'utilisation de la protéinase K pour purifier les VE dans les selles. Bien que cette purification ne soit pas totale, elle permet néanmoins de s'approcher au mieux d'une estimation du nombre de VE présentes dans les filtrats fécaux, sans avoir à recours à des processus de purification chronophages et coûteux en échantillons comme les gradients d'iodixanol.

3.5.2.3. Quantités de VE dans les selles

Au vu des résultats précédents, il a été nécessaire de réaliser trois mesures Videodrop par échantillon pour obtenir une approximation du nombre de VE présents dans les filtrats fécaux humains (**Figure 47**). La première mesure est celle de la concentration en nanoparticules dans les filtrats fécaux (il s'agit donc de réaliser l'acquisition sur le filtrat fécal brut). La deuxième mesure est réalisée après traitement par la protéinase K selon le protocole décrit ci-avant et après incubation à 37°C pendant deux heures. La troisième mesure est quant à elle réalisée après deux heures d'incubation de l'échantillon à 37°C, mais sans protéinase K. Ainsi, le ratio mesure 3 / mesure 1 nous permet d'obtenir une proportion de nanoparticules sensibles à l'incubation à 37°C. Ce ratio est un facteur correctif à appliquer à la concentration en VE (ou en nanoparticules résistantes à la protéinase K) obtenue lors de la mesure 2. Une estimation de la concentration en VE présentes du filtrat fécal peut ainsi en être déduite.

$$\text{Nb d'EVs dans le FF} = \frac{\overbrace{\text{Nb de NP après PK à } 37^{\circ}\text{C}}^{\text{Mesure 2}} \times \overbrace{\text{Nb de NP dans le FF non traité}}^{\text{Mesure 1}}}{\underbrace{\text{Nb de NP dans le FF non traité et incubé à } 37^{\circ}\text{C}}_{\text{Mesure 3}}}$$

Figure 47. Formule du calcul du nombre de VE dans les filtrats fécaux

Nb : nombre ; NP : nanoparticule ; FF : filtrat fécal

Ces mesures et calculs ont été réalisées dans tous les filtrats fécaux issus des trois groupes de la cohorte (**Figure 48** **Erreur ! Source du renvoi introuvable.**). Aucune différence significative n'a

été relevée entre les groupes dont les valeurs médianes varient de $2,52 \times 10^{11}$ VE/g de selle patients en période de rémission à $4,70 \times 10^{11}$ VE/g de selle pour les individus sains, en passant par $3,36 \times 10^{11}$ VE/g de selle patients en poussée. La quantité de VE mesurée avec le Videodrop ne semble pas être corrélée à l'inflammation intestinale, ni à une potentielle variation de la concentration en phages au cours de celle-ci.

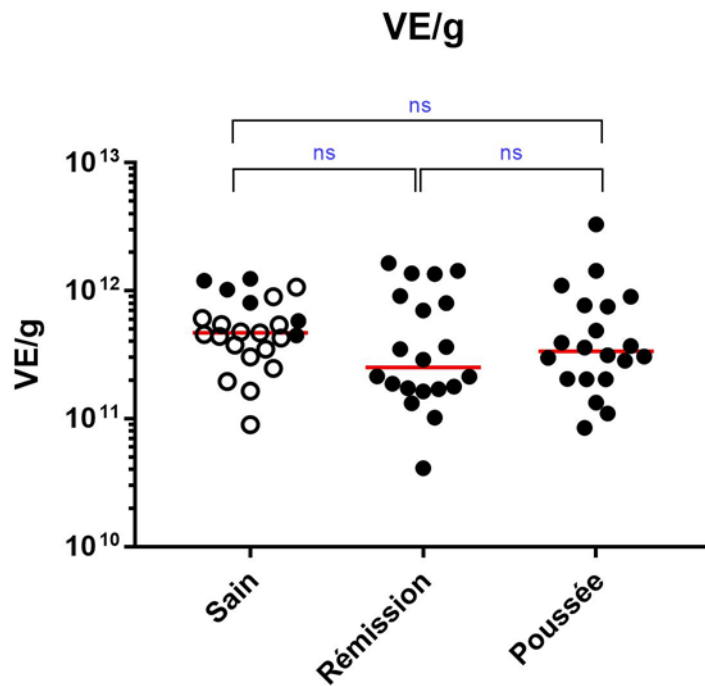


Figure 48. Dénombrement des VE dans les filtrats fécaux

Les cercles pleins correspondent à des selles ayant été conservées plus d'un an à -80°C , tandis que les cercles vides correspondent à des selles ayant été conservées moins deux mois à -80°C . Toutes les mesures ont été réalisées avec le Videodrop. Un test de Mann-Whitney a été réalisé pour comparer les différents groupes (ns = $0,1234 > p\text{-value} > 0,0332$; * = $0,0332 > p\text{-value} > 0,0021$; ** = $0,0021 > p\text{-value} > 0,0002$; *** = $0,0002 > p\text{-value} > 0,0001$; **** = $p\text{-value} < 0,0001$). Les barres horizontales rouges correspondent aux médianes de chaque groupe.

3.5.3. Discussion

Aucune différence significative de la concentration en nanoparticules résistantes à la protéinase K et détectables par le Videodrop n'a été relevée entre les différents groupes. Leurs valeurs médianes varient de $2,52 \times 10^{11}$ NP/g de selle patients en période de rémission à $4,70 \times 10^{11}$

NP/g de selle pour les individus sains et $3,36 \times 10^{11}$ NP/g de selle patients en poussée. Si la quantité de NP ne semble pas être corrélée à l'inflammation intestinale, ces valeurs de concentrations en VE dans les selles d'individus sains sont cependant très largement supérieures à celles obtenues par Nader Kameli et co-auteurs (Kameli et al. 2021). Les auteurs relèvent en effet 2×10^9 VE par gramme de selle, soit près de cent fois moins que la concentration médiane obtenue au cours de cette expérience. Il est probable que la méthode de purification des VE utilisée par les auteurs, à savoir la chromatographie par exclusion de taille, ait été une source de perte importante de vésicules comme cela a été le cas au cours de ma thèse (cf. 3.5.1.2.1. Intérêt de l'étape de centrifugation). De plus, la TRPS, utilisée par les auteurs pour procéder au dénombrement des VE, s'avère inadaptée pour l'analyse d'échantillons de tailles polydisperses (Turkki et al. 2021, van der Pol et al. 2014).

Cependant, le dénombrement des nanoparticules des filtrats fécaux par la méthode utilisée ici ne donne qu'une estimation de la concentration en VE. L'une des sources d'imprécision repose sur l'incapacité de la protéinase K à digérer l'ensemble des agrégats dans les filtrats fécaux, ce qui a pour effet de surestimer légèrement la concentration en VE dans ces derniers. L'hypothèse qu'une importante proportion de VE, si ce n'est la majorité de celles-ci, soient de trop petit diamètre pour être détectées par le Videodrop, peut être émise. En effet, le Videodrop n'est pas capable de suivre les VE les plus petites sur plusieurs images successives et donc d'en déduire leur taille réelle. Si le seuil de détection des VE par le Videodrop se situe probablement aux alentours de 70 nm, il est évident qu'un nombre important de vésicules, dont le diamètre peut descendre jusqu'à 15 nm (Article 1), ne sont pas détectées et donc prises en compte lors des calculs de concentration.

Toutefois, ces mesures nous permettent d'établir des estimations de concentrations en VE par échantillon et de remarquer que les concentrations en nanoparticules résistantes à la protéinase K et détectables par le Videodrop ne sont pas modifiées au cours de l'inflammation. Cependant, rien n'indique qu'il en soit de même pour les VE de diamètre inférieur à 70 nm, question à laquelle je n'ai pas pu répondre au cours de ma thèse au vu de ces limitations techniques.

3.6. Nombre de VLPs et de VE par bactérie dans les selles

Nous avons choisi d'utiliser les concentrations en bactéries dans celles-ci afin de normaliser les échantillons entre eux et d'affiner leurs comparaisons. Aucune différence significative du nombre de VE par bactérie n'a été observée dans les selles des individus sains et des patients en période de rémission (**Figure 49***Erreur ! Source du renvoi introuvable.*). En effet, leurs valeurs médianes sont, respectivement, de 11,8 et de 9,5 VE/bactérie. Une augmentation significative d'environ trois fois plus de VE a cependant été détectée chez les patients en poussée, avec une médiane de 32,2 VE/bactérie.

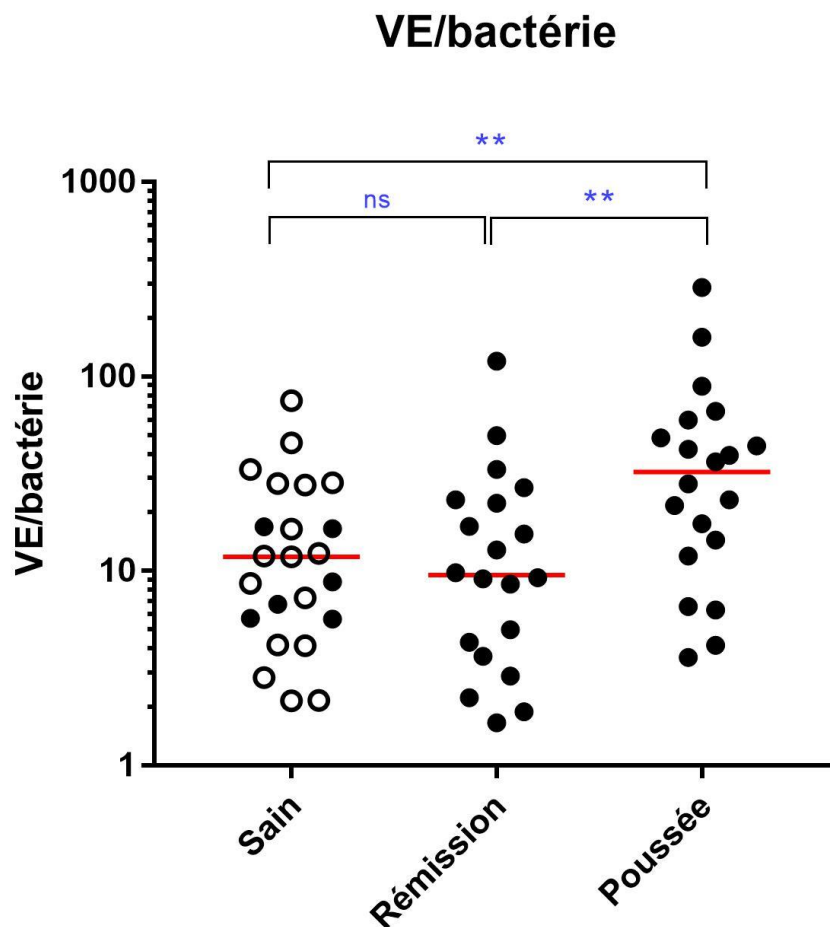


Figure 49. Nombre de VE par bactérie

Les cercles pleins correspondent à des selles ayant été conservées plus d'un an à -80°C , tandis que les cercles vides correspondent à des selles ayant été conservées moins de deux mois à -80°C . Un test de Mann-Whitney a été réalisé pour comparer les différents groupes ($ns = 0,1234 > p\text{-value} > 0,0332$; $* = 0,0332 > p\text{-value} > 0,0021$; $** = 0,0021 > p\text{-value} > 0,0002$; $*** = 0,0002 > p\text{-value} > 0,0001$; $**** = p\text{-value} < 0,0001$). Les barres horizontales rouges correspondent aux médianes de chaque groupe.

De la même façon, le nombre de VLPs par bactérie a été exploré. Une augmentation très significative de ce ratio a été observée entre le groupe d'individus sains et de patients en pleine poussée, avec, respectivement, des médianes de 0,187 et de 21,4 VLPs détectées par bactérie, soit plus de 100 fois plus. Ces valeurs indiquent également une inversion radicale de la dynamique d'écologie présente dans ces écosystèmes. Une légère augmentation (d'un facteur 2) du nombre de VLPs par bactérie est également observée chez les patients en phase de rémission par rapport aux donneurs sains, avec un ratio médian s'élevant à 0,45, ce qui reste bien en-deçà de celui observé dans les échantillons prélevés lors des poussées.

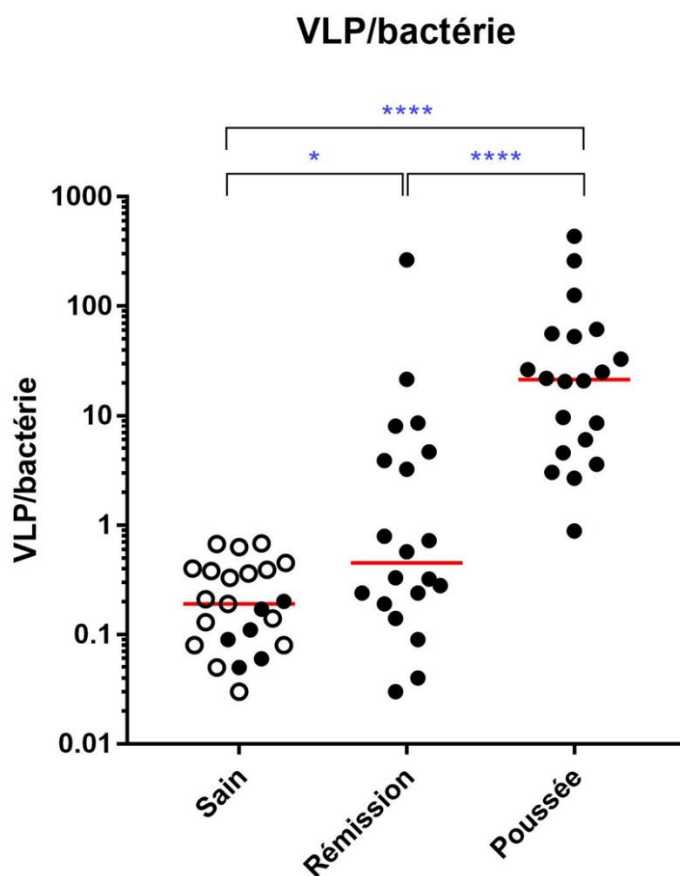


Figure 50. Nombre de VLPs par bactérie

Les cercles pleins correspondent à des selles ayant été conservées plus d'un an à -80°C, tandis que les cercles vides correspondent à des selles ayant été conservées moins de deux mois à -80°C. Un test de Mann-Whitney a été réalisé pour comparer les différents groupes (ns = 0,1234 > p-value > 0,0332 ; * = 0,0332 > p-value > 0,0021 ; ** = 0,0021 > p-value > 0,0002 ; *** = 0,0002 > p-value > 0,0001 ; **** = p-value < 0,0001. Les barres horizontales rouges correspondent aux médianes de chaque groupe.

Une importante variabilité interindividuelle est présente au sein de chaque groupe d'individus (**Figure 51** **Erreur ! Source du renvoi introuvable.**). Les échantillons les plus concentrés en VLPs de chaque groupe ont été observés en MET afin de constater ou non la présence de virus et en particulier de phages, au sein de ceux-ci. J'ai en effet pu remarquer, dans plusieurs filtrats fécaux issus des trois groupes, la présence d'objets assimilables à des virus (**Figure 52** **Erreur ! Source du renvoi introuvable.**). Bien que ceux-ci soient difficiles à différencier des VE, c'est leur nombre et leur homogénéité de tailles (entre 30 et 55 nm de diamètre dans la majorité des cas), ainsi que la qualité de la coloration, qui me poussent à penser qu'il s'agit de virus et non de vésicules. Ces virus ne semblent pas présenter de queue, il pourrait donc s'agir, par exemple, de *Microviridae* ou de virus eucaryotes. Tous semblent morphologiquement semblables, même d'un échantillon à l'autre, ce qui suggère qu'il s'agit d'une unique famille représentée par plusieurs espèces. A l'inverse, le filtrat fécal du patient en rémission R2 ($2,30 \times 10^{13}$ VLPs/g) n'a pas révélé une importante quantité de virions sur la grille observée en MET, malgré une importante concentration de VLPs détectée en microscopie par épifluorescence. La présence de nombreux agrégats non identifiés a cependant été révélée par la MET. Ces agrégats pourraient être la source de la fluorescence détectée lors de la quantification des VLPs ou bien engluer les virions, rendant leur observation par MET plus complexe.

Il est désormais évident que le microbiote des patients atteints de la maladie de Crohn se trouve fondamentalement bouleversé, en particulier au cours de la poussée. Une augmentation importante du rapport VLPs/bactérie a cependant été observée chez un individu sain (S22), ainsi que chez quelques patients en phase de rémission. Concernant ces derniers, il se pourrait que ceux-ci se trouvent tout simplement en début de poussée pas encore diagnostiquée ou à la fin d'une phase d'inflammation aigue. Malheureusement, aucune donnée établie lors de la constitution de la cohorte ne nous permet de situer les patients par rapport à la date de survenue de la prochaine poussée. Chez les individus sains, il n'est pas non plus très surprenant de trouver des individus chez lesquels ce ratio VLPs/bactérie est élevé. Il pourrait s'agir d'individus sujets au passage temporaire d'un phage virulent ou subissant une induction de prophages suite à un léger stress, non perçu comme étant symptomatique par ces personnes, qui n'ont pas déclaré être malades lors du recueil des selles. Une autre hypothèse est que certains individus possèdent un microbiote dont l'état de stabilité se définirait par un ratio phages/bactérie bien supérieur à la moyenne de la population. Seule une étude longitudinale permettrait de répondre à cette question.

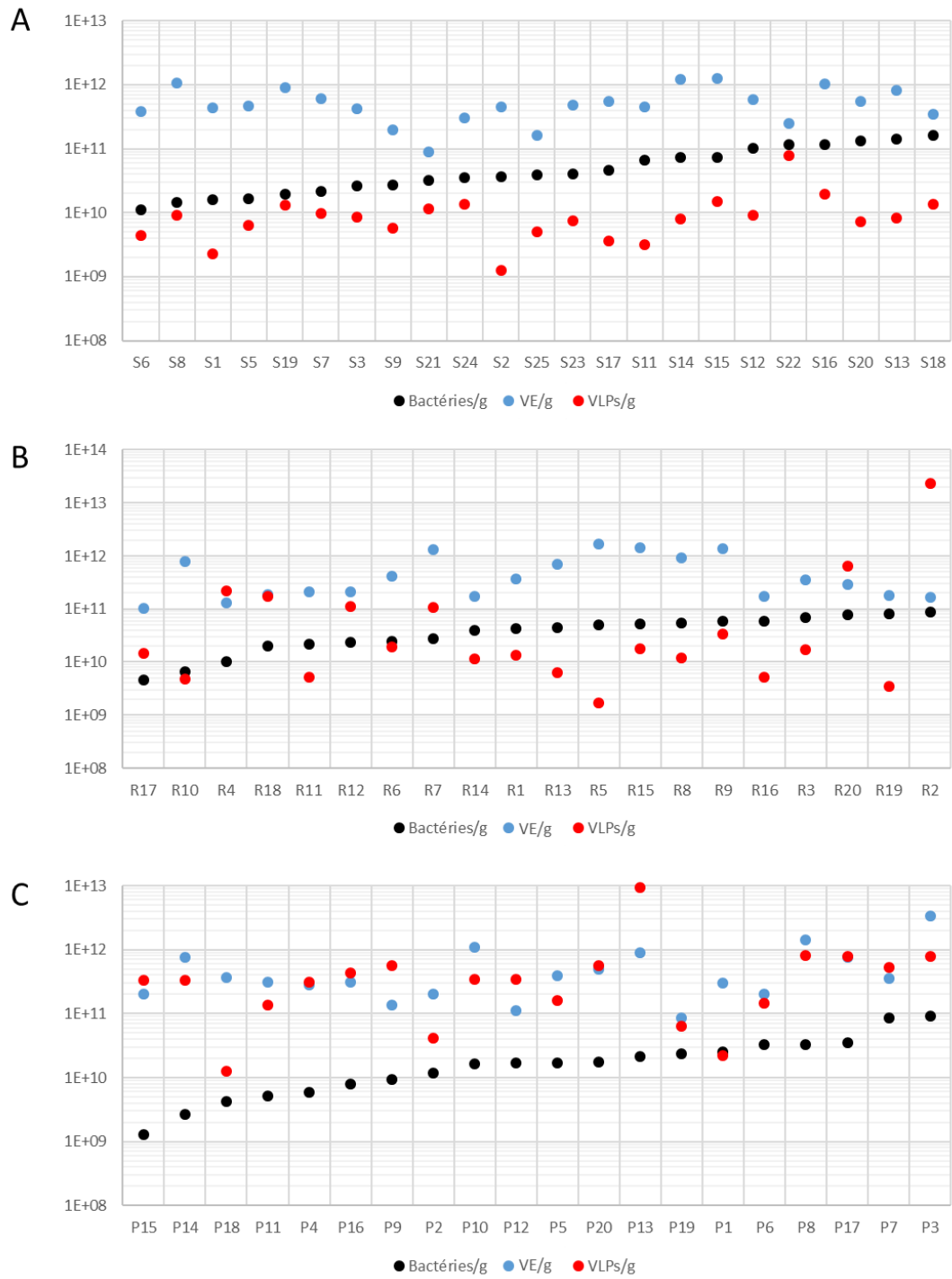


Figure 51. Concentrations individuelles de bactéries, VLPs et VE

(A) Dans les selles d'individus sains. (B) Dans les selles de patients en période de rémission. (C) Dans les selles de patients en période de poussée.

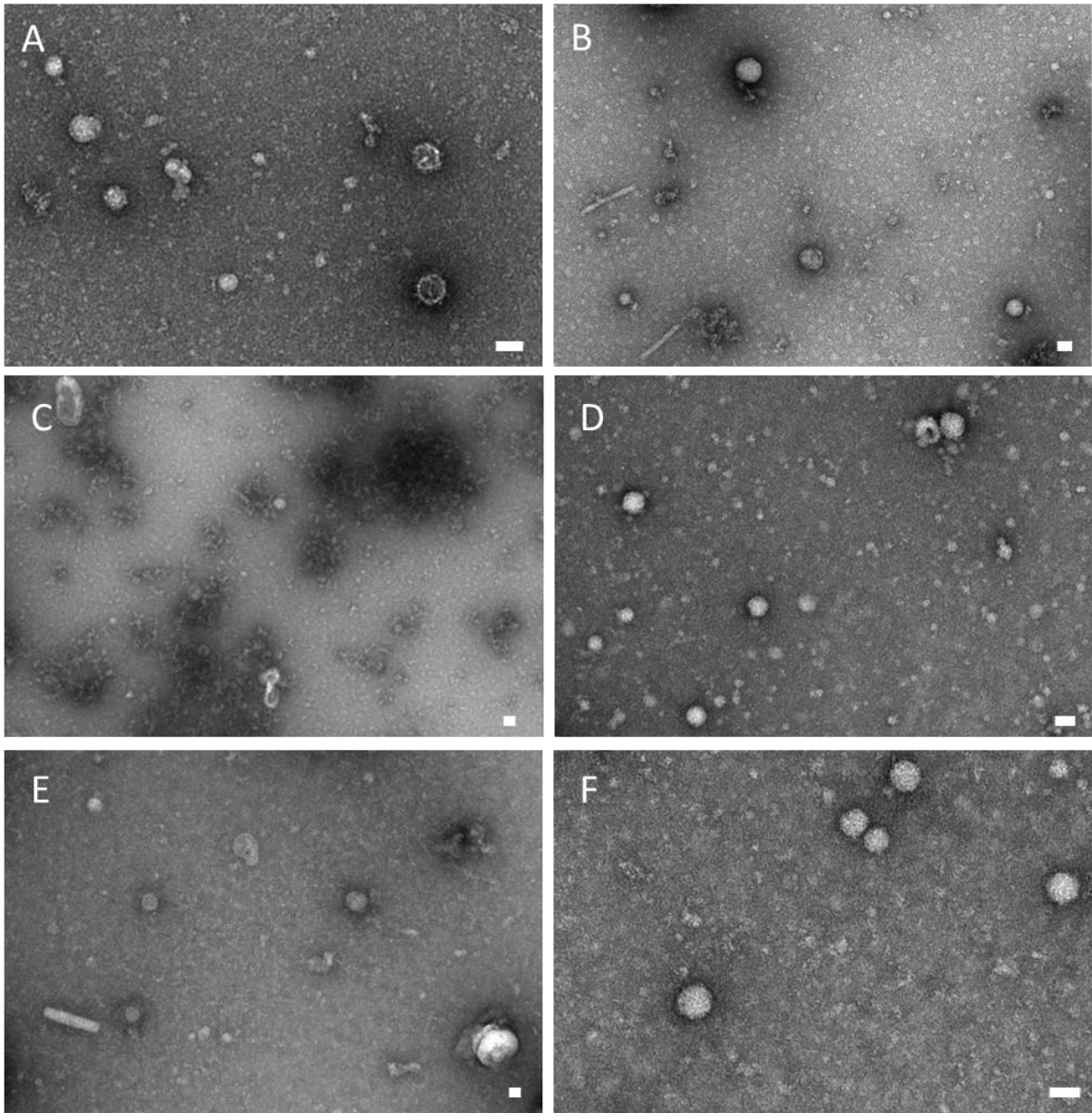


Figure 52. MET de filtrats fécaux d'individus dont la concentration en VLPs est élevée

(A) Filtrat fécal S22. (B) Filtrat fécal S13. (C) Filtrat fécal R2. (D) Filtrat fécal R18. (E) Filtrat fécal P3. (F) Filtrat fécal P13. Les barres d'échelle représentent 50 nm.

4. Recherche de phages porteurs de gènes d'antibiorésistance dans des viromes de lapins d'élevage

4.1. Introduction

En mars 2020, le contexte de pandémie de covid-19 a mis le monde en quarantaine, nous forçant brutalement à ralentir, voire stopper nos activités de recherche qui ne pouvaient plus être conduites que depuis chez soi. Ces deux mois de confinement ont été l'occasion de participer à un nouveau projet, totalement déconnecté de mon sujet de thèse, grâce auquel j'ai pu m'initier aux bases de la bioinformatique, en vue d'annoter 30 viromes, dont l'ADN avait été préparé et envoyé à séquencer avant mon arrivée sur le projet.

Le projet sur lequel j'ai été invité à participer est mené par Marie-Agnès Petit et Olivier Zemb, de l'INRAe de Toulouse, qui s'intéresse aux potentiels transferts de gènes bactériens d'antibiorésistance par les phages. Contrairement à l'idée répandue que les génomes de phages codent parfois pour des gènes d'antibiorésistance et permettent ainsi leur transfert horizontal fréquent entre bactéries, un projet similaire mené au sein de l'équipe (Billaud et al. 2021) a montré, en caractérisant 14 viromes fécaux de porcs, qu'aucun des contigs viraux identifiés n'était porteur de gènes d'antibiorésistance (ARG). Pour élargir ces observations, nous nous sommes également interrogés sur la présence de gènes d'antibiorésistance codés par les phages, en étudiant les viromes de fèces de 26 lapins provenant de 26 élevages différents, ayant été traités (75% des élevages) ou non (25%) aux antibiotiques dans les 3 mois précédant l'échantillonnage.

4.2. Méthodes

Des échantillons fécaux provenant de 26 lapins d'élevage ont été collectés et traités afin d'extraire l'ADN total des particules virales. Après la collecte, 1 à 2 g de fèces ont été conservées à -80°C jusqu'au traitement. Les échantillons fécaux ont d'abord été remis en suspension dans 45 mL de solution saline tamponnée au phosphate (PBS), puis homogénéisés pendant 1 h à 4°C. Deux cycles de centrifugation ont été effectués, un premier de 7 minutes à 5 000 g et 4°C pour éliminer les débris, puis un second de 45 minutes à 8000 g à 4°C pour éliminer les contaminations non virales restantes. Le surnageant a été filtré à 0,45 µm. Après ajout de NaCl et de Polyéthylène Glycol 8000 à des concentrations finales de 0,5M et 10 % w/v, respectivement, et incubation à 4°C toute une nuit, les échantillons ont ensuite été centrifugés deux fois à 5000 g pendant 7 minutes à 4°C. Les culots

résultants ont été remis en suspension dans 500 µL de tampon SM et dialysés contre 1 L de tampon SM pendant la nuit pour éliminer les résidus de NaCl et de PEG8000. Les échantillons ont alors été filtrés sur 0,45 µm et tous les volumes ont été ajustés à 500 µL avec du tampon SM. L'ADN libre a été éliminé par un traitement par la DNase I. Pour cela, les solutions de phages ont été complétées par du CaCl₂ (0,5 mM) et 0,5 µL de DNase I Ambion a été ajouté. Les échantillons ont ensuite été incubés à 37°C pendant 30 minutes. L'ADN viral a ensuite été extrait par phénol/chloroforme, suivi d'une précipitation à l'éthanol. L'ADN final a été remis en suspension dans un volume de 50 µL (dans du Tris pH 8, 10 mM). La concentration d'ADN a été estimée avec le kit Qubit HS (8 ng à 2 µg par échantillon).

Tous les échantillons ont été amplifiés avec le kit Illustra Genomiphi V2 après dialyse sur de petits filtres de 0,025 µm pour éliminer les inhibiteurs de la réaction de réplification. Si des quantités suffisantes d'ADN étaient générées après 75 minutes d'amplification, la réaction a été stoppée; sinon, 30 minutes d'incubation supplémentaire ont été ajoutées. Pour trois échantillons (numéros 15, 17 et 18), aucune amplification n'a pu être obtenue, même après une incubation d'une nuit. À cette étape, les échantillons ont été envoyés à séquencer chez Eurofins. Après fragmentation de l'ADN, des bibliothèques Truseq ont été préparées pour le séquençage Illumina Hiseq (2x125 bp) (25 millions de lectures par échantillon).

A partir des reads bruts reçus de la plateforme de séquençage, mes travaux ont consisté à préparer puis assembler ces reads en contigs, séparer les contigs reconnus comme viraux des autres (les viromes sont fréquemment contaminés avec de l'ADN bactérien), les annoter, rechercher les ARGs et analyser leur environnement génétique, pour confirmer leur origine phagique.

L'étape de nettoyage des reads bruts des échantillons amplifiés a été effectuée à l'aide de Trimmomatic v0.39 (Bolger, Lohse, and Usadel 2014), avec les paramètres suivants : ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:30 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN 100. La qualité des reads résultants a été confirmée en utilisant FastQC v0.11.8. De plus, ViromeQC v1.0 (Zolfo et al. 2019) a été utilisé pour estimer la prévalence des reads d'origine virale par rapport aux reads contaminants bactériens dans chaque échantillon. Les reads ont été dérépliqués avec l'outil fastx-unique (version gratuite de la suite Usearch9 (R. Edgar, <http://drive5.com/usearch>)). Les paires de reads ont ensuite été reconstruites à l'aide de la commande fast_qpair (Edwards lab) et assemblées en contigs à l'aide de SPAdes v3.14.0 (Nurk et al. 2013) en utilisant comme paramètres "--only-assemble -k 21,33,55,77,99,127". Les contigs ont ensuite été annotés individuellement en utilisant Prokka v1.14.5 (Seemann 2014) et la présence d'ARGs dans ces séquences a été recherchée en

comparant le protéome total à la base de données Resfams (Gibson, Forsberg, and Dantas 2015) en utilisant Hmmer v3.2.1, avec la commande "hmmsearch --cut_ga".

Sur l'ensemble de contigs, nous avons ensuite recherché lesquels avaient une forte probabilité d'être viraux, à l'aide de VIBRANT v1.2.1 (Kieft, Zhou, and Anantharaman 2020). En effet, les travaux précédents de l'équipe avaient montré que VIBRANT était un logiciel particulièrement adapté pour trier les contigs viraux de tous les autres (plasmidiques ou bactériens, (Billaud et al. 2021)). Cependant, VIBRANT n'accepte en entrée que les contigs ayant une taille supérieure à 2 kb (seuil également fixé pour nos analyses de contigs de lapins) et codant pour au moins quatre ORFs. Dans notre jeu de données, certains des contigs ARG+ (> 2 kb) avaient néanmoins moins de 4 ORFs. Nous avons donc procédé en deux étapes : Pour les contigs contenant au moins quatre ORFs, le résultat VIBRANT a été utilisé pour prédire les contigs d'origine phagique. Dans le cas des contigs portant trois ORF ou moins, ils ont été comparés par BLASTn (option megablast) à la banque de données non redondantes du NCBI (<https://www.ncbi.nlm.nih.gov/>) afin de proposer leur origine.

Dans une dernière étape, afin de supprimer la redondance entre tous les contigs issus des 30 assemblages, les contigs de longueur supérieure à 2kb ont également été soumis à CD-HIT v4.8.1 (Fu et al. 2012, Li and Godzik 2006) pour un clustering à 95% d'identité (divergence correspondant à l'échelle de l'espèce phagique) avec les paramètres suivants "c 0.95 -G 0 -g 1 -aS 0.9 -circle 1 -prog blastn". Le logiciel CloneManager v10.1 a également été utilisé pour identifier chaque ARG et son environnement génétique de façon individuelle.

4.3. Résultats

4.3.1. L'assemblage des viromes révèle une majorité de contigs phagiques

L'analyse ViromeQC des reads de chaque échantillon a montré un très bon enrichissement dans les échantillons amplifiés. L'échantillon A33 est celui ayant montré le plus mauvais taux d'enrichissement, bien que celui-ci reste acceptable (15,6 x sur l'ARNr SSU). Nous concluons donc que les échantillons sont suffisamment enrichis pour lancer une analyse de viromes.

L'assemblage Spades, réalisé séparément pour chaque échantillon, a permis de générer entre 143 et 2374 contigs d'une taille > 2kb par échantillon. Les plus grands contigs possédaient des tailles de 7194 à 168170 pb (**Tableau III**).

L'analyse VIBRANT de tous les contigs individuels a révélé la prévalence des contigs phagiques par rapport aux autres contigs (plasmidiques, bactériens, ou viraux mais non reconnus

comme tels par VIBRANT). Ainsi, 51,88% (9 194 sur 17 723) des contigs ont été triés comme étant des phages putatifs par VIBRANT, la somme des longueurs des contigs phagiques de l'ensemble des échantillons s'élevant à 69 986 061 pb. La présence de contigs bactériens ou autres que viraux était cependant plus élevée que celle relevée précédemment par notre équipe dans les viromes de porcs (Billaud et al. 2021), où nous avons obtenu une proportion de 77,3% de contigs phagiques, en utilisant essentiellement la même approche. Cette pureté plus importante pourrait s'expliquer par l'ajout d'une étape de purification des viromes sur gradient d'iodixanol, qui n'a pas été réalisé ici.

4.3.2. Sur les 15 585 vOTUs non redondants issus des viromes de lapins, 35 sont ARG positifs.

Afin de procéder à la recherche exhaustive d'ARGs, y compris sur les contigs non déclarés comme viraux, la collection de 17 723 contigs (> 2kb) a été organisée en 15 585 contigs non redondants appelés « vOTU » (pour viral taxonomic unit), au sens large. En effet, l'étape d'agrégation a été effectuée à 95% d'identité, seuil correspondant au niveau taxonomique de l'espèce phagique (ou bactérienne). De manière intéressante, seulement 7,3% (1144 sur 15 585) des vOTU contenaient plus d'un contig, suggérant une très grande diversité virale inter-individuelle. Ceci est une constante dans les analyses de viromes, ainsi les viromes de porcs analysés de la même façon avaient également produit seulement 4,8% de vOTU correspondant à des contigs redondants.

Parmi ces 15 585 vOTUs issus de viromes lapins, seuls 35 (0,22%) portaient des ARG homologues à ceux présents dans la base de données Resfams. Pour comparaison, une analyse réalisée de la même façon sur les viromes de porcs avait elle aussi permis d'identifier seulement 0,1% de contigs ARG positifs. Parmi ces 35 vOTUs, 6 d'entre eux, n'ayant pas suffisamment d'ORFs, n'avaient pas été classés par VIBRANT. Parmi les 29 autres, aucun n'était situé sur des contigs reconnus comme phagiques par Vibrant (**Tableau IV**).

Afin de rechercher si les six derniers contigs porteurs d'ARG pouvaient être phagiques, une analyse manuelle a été effectuée (voir Méthodes). Trois d'entre eux n'avaient aucun gène typique de phage. Les trois derniers, codant un gène conférant la résistance à la kanamycine, portaient en revanche un gène *lysIA*, codant une endolysine du phage Idefix, infectant *Enterococcus faecalis*. Un complément d'analyse a permis de reconnaître ces contigs comme des fragments de deux plasmides présents au laboratoire. Ceci met en évidence une contamination des viromes, qui aurait pu avoir lieu lors de l'étape de l'extraction d'ADN. Nous concluons donc que ces six derniers contigs, eux non plus, ne sont pas d'origine virale.

4.4. Conclusion

En conclusion, aucun des ARGs que nous avons trouvés n'est situé sur un contig phagique. Ceci permet d'estimer une fréquence d'ARG par contigs inférieure à $1/9194$ ($< 1,08 \times 10^{-4}$), ou plus précisément moins d'une ORF sur les 108 994 ORFs totales présentes sur tous les contigs viraux ($< 0,92 \times 10^{-5}$). Ce chiffre est semblable à celui obtenu pour les viromes de porc ($< 1,2 \times 10^{-5}$ ORF) (Billaud et al. 2021). Ces résultats suggèrent que les génomes de phages contribuent très peu aux transferts horizontaux d'ARGs dans les échantillons analysés. Il n'est cependant pas exclu que des phénomènes de transduction généralisée, c'est-à-dire d'emballage d'ADN bactérien dans les capsides phagiques, à la place du génome phagique, puissent avoir lieu, et contribuer à la diffusion d'ARGs, à une fréquence qui reste à déterminer.

Ces travaux seront intégrés dans une publication à venir.

Échantillon	SSU		Nombre de contigs > 2kpb	Taille max des contigs	Nombre d'ORFs totaux portés par les contigs >2kpb de l'échantillon	Nombre de contigs viraux >2kb	Nombre de contigs >2kpb porteurs d'ARGs
	%	x					
A1	0,0008	100	181	7 194	723	99	2
A2	0,001	100	352	115 732	2 056	184	0
A3	0,0024	100	424	116 324	2 472	216	1
A4	0,0005	100	393	83 586	2 301	186	2
A5	0,0019	100	961	107 820	5 948	482	3
A7	0,0016	100	162	8 496	545	70	1
A8	0,0011	100	191	58 146	846	103	1
A10	0,0014	100	391	81 033	2 190	208	1
A11	0,0015	100	1 079	98 703	6 584	549	4
A12	0,0011	100	196	12 779	689	88	0
A13	0,0013	100	803	171 952	6 264	386	4
A14	0,0028	88,4	797	54 046	4 137	458	2
A19	0,0041	59,8	2 374	167 919	17 266	1 268	3
A21	0,0039	63,9	603	99 958	2 893	316	1
A22	0,0032	76,8	1 457	71 788	7 264	861	1
A23	0,0012	100	582	133 638	3 897	301	2
A24	0,0009	100	626	63 242	3 265	286	0
A25	0,0011	100	162	100 510	850	78	2
A26	0,001	100	810	148 578	5 464	394	2
A27	0,001	100	434	46 589	2 117	194	4
A28	0,0007	100	692	67 302	3 956	395	1
A29	0,0034	73,2	1 434	56 827	6 093	633	8
A30	0,0007	100	143	16 662	697	86	0
A31	0,0019	100	1 117	127 543	8 995	620	3
A32	0,0026	96,1	416	65 565	2 449	250	1
A33	0,0159	15,6	943	168 170	8 413	483	3

Tableau III. Caractéristiques des échantillons de viromes de lapin.

SSU : % de reads s'alignant à la petite sous-unité d'ARN ribosomique, utilisée comme marqueur de contamination bactérienne.

Contigs porteurs d'ARGs	Nature	Contigs*	Taille des contigs (pb)	Types d'ARGs	Résistance		
Analysés manuellement (6)	Plasmidique (3)	A4_NODE239	6136	Aminoglycoside 3'-phosphotransferase	kanamycine		
		A10_NODE120 A5_NODE288	5191 5407	Aminoglycoside 3'-phosphotransferase	kanamycine		
		A23_NODE349	3261	Aminoglycoside 3'-phosphotransferase	kanamycine		
	Inconnue (3)	A5_NODE766	2467	Nucleotide-binding protein ExpZ	macrolides		
		A29_NODE1248	2235	tetracycline efflux MFS transporter Tet(K)	tétracycline		
		A31_NODE480	4327	Nucleotide-binding protein ExpZ	macrolides		
Analysés par VIBRANT (29)	Bactérienne ou autre (29)	A3_NODE349	2467	Aminoglycoside 6-adenylyltransferase	streptomycine		
				Streptothricin acetyltransferase A	streptothricine		
				Aminoglycoside 3'-phosphotransferase	kanamycine		
		A8_NODE111	3445	Chloramphenicol acetyltransferase	chloramphénicol		
		A11_NODE136 A13_NODE68 A1_NODE118	7419 15079 3734	Aminoglycoside 3'-phosphotransferase	kanamycine		
		A11_NODE191 A13_NODE197	6240 6044			Aminoglycoside 6-adenylyltransferase	streptomycine
		A11_NODE239	5785			rRNA adenine N-6-methyltransferase	macrolides
		A11_NODE277 A19_NODE699 A21_NODE132 A27_NODE123 A7_NODE30	5543 5534 5534 5534	Aminoglycoside 3'-phosphotransferase	kanamycine		
		A13_NODE230	5648			Chloramphenicol acetyltransferase	chloramphénicol
		A13_NODE323 A27_NODE134 A5_NODE145	4779 5363 7026			rRNA adenine N-6-methyltransferase	macrolides
		A14_NODE125	6105			Aminoglycoside 3'-phosphotransferase	kanamycine
		A14_NODE196 A1_NODE51 A33_NODE733	5534 5397 2480	Aminoglycoside 3'-phosphotransferase	kanamycine		
		A19_NODE372	8734			tetracycline resistance ribosomal protection protein Tet(W)	tétracycline
		A19_NODE709	5481	Chloramphenicol acetyltransferase	chloramphénicol		
		A22_NODE446 A29_NODE311 A32_NODE111	5046 5534 5534	Aminoglycoside 3'-phosphotransferase	kanamycine		
		A23_NODE137 A4_NODE70	6011 6136			Aminoglycoside 3'-phosphotransferase	kanamycine
		A25_NODE13	6445	rRNA adenine N-6-methyltransferase	macrolides		
		A25_NODE5	7401	Aminoglycoside 3'-phosphotransferase	kanamycine		
		A26_NODE223 A29_NODE98	5574 8924	rRNA adenine N-6-methyltransferase	macrolides		
		A26_NODE371	4532			tetracycline resistance ribosomal protection protein Tet(44)	tétracycline
		A27_NODE177	4879	Aminoglycoside 6-adenylyltransferase	macrolides		
		A27_NODE81	6136	Aminoglycoside 3'-phosphotransferase	kanamycine		
		A28_NODE82	7134	Nucleotide-binding protein ExpZ	macrolides		
		A29_NODE106	8504	Aminoglycoside 3'-phosphotransferase	kanamycine		
		A29_NODE428	5013	Chloramphenicol acetyltransferase	chloramphénicol		
		A29_NODE724	3650	Nucleotide-binding protein ExpZ	macrolides		
		A29_NODE851	3161	rRNA adenine N-6-methyltransferase	macrolides		
		A29_NODE99	8893	rRNA adenine N-6-methyltransferase	macrolides		
		A31_NODE327	5691	Chloramphenicol acetyltransferase	chloramphénicol		
		A31_NODE331 A33_NODE921	5647 2044	tetracycline resistance ribosomal protection protein Tet(44)	tétracycline		
		A33_NODE296	5461			Nucleotide-binding protein ExpZ	macrolides

Tableau IV. Caractéristiques des ARGs dans les échantillons analysés

*Les contigs sont regroupés par vOTUs, le cas échéant

Matériel et Méthodes

Destruction des VE

Le chloroforme, couramment utilisé dans les lysats pour prévenir des contaminations bactériennes, a été utilisé à une concentration finale de 50% v/v. Les échantillons ont ensuite été soumis à un vortexage de 30 secondes afin d'obtenir une émulsion. Le Vertrel (Sigma-Aldrich), également utilisé pour la purification de phages (Mendez et al. 2000), a quant à lui été utilisé à une concentration finale de 30% v/v, avant agitation forte avec un vortex pendant 2 secondes, de façon à obtenir une émulsion. Après une incubation de 30 minutes à température ambiante, la phase supérieure est prélevée. Le Sarkosyl (Sigma-Aldrich) a également été testé. Utilisé dans les protocoles de purification de protéines issues de vésicules bactériennes (Valguarnera et al. 2018), il est peu délétère pour les phages T4, T7 et λ (Jurczak-Kurek et al. 2016), à une concentration finale de 0,1% m/v. Après une brève agitation forte réalisée à l'aide d'un vortex, la préparation est incubée 10 minutes à température ambiante. Deux détergents ayant montré des effets de destruction des VE prometteurs (Osteikoetxea et al. 2015, Valguarnera et al. 2018), le Tween 20 (Sigma-Aldrich) et le Triton X-100 (Merck), ont été utilisés à des concentrations finales respectives de 1% v/v et de 0,075% v/v. Ces mélanges ont été suivis d'un vortexage de 30 secondes, afin de bien homogénéiser les préparations. Enfin, le NP-40 (Sigma-Aldrich), connu pour désagréger les membranes eucaryotes et procaryotes (Davies et al. 1984, Salton and Netschey 1965), a été utilisé à une concentration finale de 1% v/v. Les préparations ont ensuite été soumises à une brève étape de vortexage de 2 secondes. Les mesures ont été réalisées après une incubation de 30 minutes dans de la glace.

Dénombrement des bactéries dans les selles

Après décongélation, des aliquots de selles préalablement pesées et d'environ 100 mg ont été homogénéisées puis diluées 100 000 fois dans une solution physiologique de NaCl (8,5 g/L) préalablement filtrée à 0,22 μ m. Les solutions fécales ont été filtrées à l'aide de filtres stériles de pores de 5 μ m afin d'éliminer les plus gros débris des homogénats obtenus. 1 μ L d'une solution de SYBR Green I (de concentration 1 000x) a été ajouté à 1 mL de solution fécale ainsi filtrée. Pour chaque série de selles ainsi préparée, un contrôlé négatif de solution physiologique a été réalisé.

Après 15 minutes d'incubation à l'abri de la lumière, la quantification des bactéries par cytométrie en flux (FCM) a été réalisée à l'aide d'un cytomètre en flux C6 Accuri (BD Biosciences). Les événements de fluorescence ont été enregistrés en paramétrant les détecteurs optiques de l'appareil de la façon suivante : FL1=533/30 nm, and FL3>670 nm. Un seuil minimal d'intensité de 2 000 a été

utilisé sur le canal FL1. La lumière diffusée vers l'avant (Forward Scattered, FSC, qui donne une information sur la taille des particules) et la lumière diffusée sur le côté (Side Scattered, SSC, qui donne une information sur la granulométrie des particules) ont également été collectées et analysées. La vitesse du flux a été réglée sur 14 $\mu\text{L}/\text{min}$ et chaque acquisition a été réglée pour s'arrêter après l'enregistrement de 10 000 événements ou, à défaut, après une période de 30 secondes.

Préparation des filtrats fécaux

Le protocole dit « classique »

Après décongélation d'aliquots de selles humaines de masse connue, ceux-ci sont manuellement homogénéisés dans un volume équivalent (4,5 mL pour 100 mg de selles) de Tris 10 mM (pH 7,5) préalablement filtré à 0,22 μm . Après une seconde homogénéisation de 15 minutes sur roue à 4°C, les solutions sont soumises à une centrifugation de 30 minutes, 5 200 g, à 4°C, afin de culoter les bactéries et les plus gros objets présents dans les homogénats. Après filtration du surnageant sur 0,22 μm , les filtrats fécaux ainsi obtenus sont conservés à 4°C, en attendant la quantification rapide des phages et des VE qu'ils contiennent.

Ce protocole a également été testé en remplaçant le Tris 10 mM par du PBS 1X.

Le protocole dit « RNA later »

Une masse connue de selles humaines est broyée à l'aide d'œuses dans un volume équivalent (4,5 mL pour 100 mg de selles) de RNA later. Après une courte étape de centrifugation (10 minutes, 3 000 g, 4°C), les 25% du mélange les plus proches de la surface du surnageant sont retirés à l'aide d'une pipette, permettant l'élimination des plus gros débris flottants. Le surnageant restant et le culot sont alors homogénéisés à l'aide de billes de verre et d'un court vortexage. Après retrait des billes de verre de la solution, celle-ci est centrifugée à 20 000 g pendant 15 minutes, à 4°C. Afin de mimer une étape de conservation à -80°C, comme appliquée à l'ensemble de la cohorte, le culot issu de cette étape de centrifugation est alors congelé une nuit à cette température, puis décongelé le lendemain. Après resuspension dans du Tris 10 mM, la solution est homogénéisée sur roue pendant 15 minutes, à 4°C. Comme pour le protocole classique, une centrifugation de 30 minutes, 5 200 g, à 4°C est réalisée. Le surnageant est enfin filtré à 0,22 μm .

Dénombrement des VLP dans les filtrats fécaux

1 UI de DNase I est ajouté à un mélange de 10 µL de filtrat fécal et de 970 µL de Tris 10 mM préalablement filtré. Le tout est ainsi vortexé quelques secondes avant une incubation de 30 minutes à 37°C. Après ajout de 20 µL de glutaraldéhyde à 25% (pour atteindre une concentration finale de 0,5%) et incubation à 4°C pendant 15 minutes, les échantillons sont soumis à une étape de congélation rapide dans l'azote liquide. Suite à la décongélation, 4 mL de tampon SM sont ajoutés à chaque échantillon, avant d'être filtrés sur des filtres Anodisc de 0,02 µm. Chaque filtre est ensuite incubé sur une goutte de 50 µL de SYBR Gold 200X dans l'obscurité pendant 15 min, avec le côté virus vers le haut. Après retrait de la goutte, les filtres sont laissés dans l'obscurité pour séchage pendant 15-20 minutes avant d'être montés avec une goutte de Fluoromount-G entre une lame et une lamelle de verre. Les lames sont conservées à -20°C jusqu'à observation. Les observations microscopiques sont réalisées à l'aide d'un Nikon Ti-E équipé d'un objectif à huile 100× Apo TIRF (NA, 1.49 ; Nikon) avec un système de couplage laser iLas2 de Gataca Systems (150 mW, 488 nm). Dix images sont prises par lame, à des positions aléatoires sur celle-ci, avec un canal de champ clair et de fluorescence GFP (avec une longueur d'onde de filtre d'excitation de 472/30 nm et une longueur d'onde de filtre d'émission de 520/35 nm). Les émissions sont collectées à l'aide de filtres interférentiels et imagées sur une paire de caméras sCMOS (Orca Flash 4.0 v2 sCMOS ; Hamamatsu), avec un gain défini à 300 fixé à un objectif de grossissement $\times 2,5$; et avec un temps d'exposition réglé de 100 ms à 200 ms selon l'intensité de fluorescence des VLPs de chaque échantillon. La dimension finale des pixels est de 64 nm de côté. Le logiciel Metamorph v.7 a été utilisé pour contrôler et traiter l'acquisition des images, qui ont ensuite été analysées avec ImageJ (v1.52a). En cas de nécessité, les images ainsi acquises ont pu être rognées et traitées en soustrayant à celles-ci un double d'elles-mêmes ayant été traitée par un flou gaussien d'un rayon de 12 pixels. Les images sont ainsi nettoyées de leur bruit de fond et l'utilisation de la fonction « find maxima » permet alors le comptage rapide des VLPs. Le nombre de VLPs sur l'ensemble des filtres est alors calculé en multipliant les comptages moyens par le quotient de la surface du filtre par la surface des images.

Passage des filtrats fécaux sur colonnes SEC

Le passage sur les colonnes SEC ne pouvant se réaliser que millilitre par millilitre, le filtrat fécal utilisé a préalablement été concentré à l'aide de tubes filtrants Amicon (100K). Le volume total du filtrat fécal étant de 36 mL, trois passages successifs (15 mL, 15 mL, puis le reste) ont été nécessaires. Lors de chaque passage, le tube a été soumis à une centrifugation de 10 minutes, à

4 000 g, à une température de 4°C. Un concentré de 3 mL a pu être récupéré, auquel se sont ajoutés 2 mL de PBS (1X), utilisés pour laver l'unité de filtration et en détacher d'éventuelles nanoparticules restées piégées. Ainsi, 5 mL de filtrat fécal concentré ont pu être passés sur les colonnes SEC (qEVs) de la façon suivante.

Après retrait de l'azide de sodium permettant de protéger la colonne, celle-ci a été lavée par 5 mL de PBS filtré, à quatre reprises. L'échantillon de filtrat fécal concentré a ensuite été injecté, millilitre par millilitre. Pour chaque millilitre ayant traversé la colonne, 1 mL de PBS a été ajouté dans celle-ci afin d'éviter son assèchement. Ainsi, après 3 mL sortis de la colonne (correspondant au volume mort), 1,3 mL est collecté. Il s'agit de la fraction d'intérêt. Une fois cette fraction recueillie, trois fois 5 mL de PBS ont été injectés dans la colonne afin de la nettoyer. L'ensemble de ces opérations a dû être répété pour chaque millilitre passé dans la colonne, soit un total de cinq fois. Le nombre de nanoparticules détectables par le Videodrop, ainsi que la masse de protéines, ont été mesurées dans l'ensemble des fractions recueillies, respectivement par le Videodrop et le QuBit.

Extraction des peptides des VE

450 µL des fractions 2 issues de filtrats fécaux ont été lyophilisés en speedvac puis resuspendus dans 30 µL d'eau stérile. Dans chaque échantillon, 10 µL de bleu Laemmli 4X et 1,57 µL de bêta-mercaptol ont été ajoutés. Après incubation pendant 15 minutes à 91°C, l'ensemble des solutions a été chargé sur une électrophorèse sur gel 1D à courte migration (NuPAGE® 4 12 % Bis-Tris Gel, novex). Les protéines ont été visualisées à l'aide de Coomassie G-250 (SimplyBlue™ SafeStain, Invitrogen) et les blocs de protéines obtenus ont été découpés en petits morceaux. Les morceaux de gel ont été décolorés à l'aide du solvant A (acide acétique 10 % v/v, éthanol 40 % v/v) et du solvant B (bicarbonate d'ammonium 50 mM 50 % v/v, acétonitrile 50 % v/v). Les liaisons disulfurées ont été réduites avec du dithiothreitol (DTT) à une concentration finale de 10 mM pendant 30 minutes à 56°C. Les cystéines ont été alkylées avec de l'iodoacétamide à une concentration finale de 50 mM en incubant 45 minutes dans l'obscurité à température ambiante. Les bandes ont d'abord été lavées avec 30 µL de 50% ACN/50mM de bicarbonate d'ammonium puis déshydratées avec 30 µL d'acétonitrile. La digestion a été réalisée pendant la nuit à 37°C avec 10 ng de trypsine (Promega). Les peptides ont été extraits d'abord avec 30 µL d'acétonitrile 40 % / acide trifluoroacétique 0,1 % puis avec 30 µL d'acétonitrile.

Les échantillons ont été séchés sous vide (SpeedVac) puis mis en suspension dans 30 µL d'acétonitrile à 2% et d'acide trifluoroacétique à 0,1 %. 1 µL par échantillon a été injecté dans le système LC-MS/MS.

Spectrométrie de masse en tandem par chromatographie liquide

La spectrométrie de masse et l'identification des protéines ont été réalisées sur la plateforme PAPPSO (MICALIS, INRA, Jouy en Josas, France ; <http://pappso.inrae.fr/>) par Céline Henry et Aaron Millan Oropeza. Les peptides ont été analysés par le spectromètre de masse timsTOF PRO (Bruker) couplé à la HPLC nanoElute. Les peptides ont été séparés sur une colonne Odyssey de IonOpticks (1,6 µm C18, 120 Å, 75 µm ID, 25 cm) en utilisant un gradient de 30 min de 2 à 30% d'ACN avec 0,1% d'acide formique. L'acquisition MS a été réalisée en mode DDA avec PASEF, de 100 à 1700 m/z avec une exclusion active de 0,4 min. La durée totale du cycle était de 1,2 s avec 6 cycles PASEF. La sélection des ions parents a été réalisée à l'aide d'un filtre bidimensionnel m/z et 1/k0 de la zone de sélection permettant l'exclusion des ions à charge unique. La tension capillaire était réglée à 1,5 kV.

Les fichiers mgf ont été utilisés pour l'identification.

Le catalogue « Integrated nonredundant Gene » (IGC 9,9 millions de gènes provenant de 1267 échantillons fécaux individuels) ainsi que la base de données *Homo sapiens* (Uniprot, version 2020, 42367 entités) ont été utilisés. X! Tandem ALANINE (version 2017.2.1.4) a été utilisé avec la méthode itérative déjà décrite (Bassignani et al. 2021).

L'identification des protéines a été réalisée avec une tolérance de masse du précurseur de 20 ppm et une tolérance de masse du fragment de 0,02 Da. Les règles de clivage enzymatique ont été fixées à la digestion par la trypsine (après Arginine et Lysine, sauf si une Proline suit directement) et aucune règle de clivage semi-enzymatique n'a été autorisée. La modification fixe a été fixée à la carbamidométhylation de la cystéine et l'oxydation de la méthionine a été considérée comme une modification potentielle. Dans un second passage, l'acétylation N-terminale a été ajoutée comme autre modification potentielle, tandis que tous les autres paramètres précédents ont été conservés. Les protéines identifiées ont été sélectionnées à la dernière étape sur base d'une e-value inférieure à 0,05.

L'ensemble des résultats a été analysé avec le package metaprotR (<https://CRAN.R-project.org/package=metaprotR>).

Discussion

L'un des résultats phare obtenu au cours de cette thèse, si ce n'est le plus important, est bien évidemment l'augmentation du rapport VLPs/bactéries d'un facteur 100 chez les patients en poussée par rapport aux individus sains. Si l'idée que des virus eucaryotes soient à l'origine des MICI fut longtemps supposée, elle a progressivement été abandonnée au vu de l'absence de corrélation évidente entre l'un d'entre eux et ce groupe de pathologies (Liang et al. 2021). Deux hypothèses principales pourraient alors être envisagées pour expliquer l'augmentation du nombre de VLPs dans les selles de patients atteints de la maladie de Crohn.

Premièrement, le passage d'un phage virulent dans les intestins pourrait expliquer cette augmentation. Plusieurs facteurs pourraient cependant entraîner une diminution de l'efficacité de la prédation des bactéries par les phages dans l'intestin par rapport aux expérimentations réalisées *in vitro*, comme la possibilité pour les bactéries de se protéger au sein de niches spatiales inaccessibles aux phages (De Sordi, Lourenco, and Debarbieux 2019, Lourenco et al. 2020, Maura et al. 2012). L'hétérogénéité de l'environnement intestinal expliquerait également une différence de susceptibilité bactérienne à la prédation phagique. En effet, les différences de pH, de concentration en dioxygène, de concentrations en mucines, acides biliaires et acides gras à chaînes courtes impacteraient la capacité des phages à infecter leurs hôtes. L'arrivée soudaine ou l'importante multiplication d'un phage virulent dans le microbiote intestinal pourrait être à l'origine de la forte augmentation de VLPs chez les patients et déclencher des phases de poussées. L'alternance entre les périodes de crise et de rémission pourrait alors être la conséquence d'une succession de multiplications phagiques séparées par des phases de moindre multiplication.

Les phages participeraient à l'inflammation de plusieurs manières. En effet, en plus de leurs interactions avec les bactéries du microbiote intestinal, il est soupçonné que les phages puissent être en lien avec le système immunitaire de l'hôte humain et avoir un impact sur la réponse immunitaire. L'affinité des phages pour les glycoprotéines présentes dans le mucus (Barr et al. 2013) leur permettrait un contact favorisé avec les cellules épithéliales et ainsi favoriser les phénomènes de transcytose observés *in vitro* (Nguyen et al. 2017). Au-delà d'une couche de mucus moins épaisse chez les patients souffrants de MICI, l'atteinte de l'intégrité de la barrière intestinale pourrait également favoriser la rencontre des phages avec le système immunitaire de l'hôte, via la phagocytose par les cellules dendritiques (Aronow R Fau - Danon et al. , Barfoot et al. 1989, Sweere et al. 2019, Wenger, Turner, and Petriccioni 1978) ou par les splénocytes (Hodyra-Stefaniak et al. 2015). De plus, les phages provoqueraient l'augmentation de la perméabilité de la barrière épithéliale

(Tetz et al. 2017). Chez la souris, l'administration orale de phages d'*E. coli* précédant une colite induite par le DSS provoque par ailleurs une exacerbation significative de la colite, ainsi qu'une augmentation importante de l'atteinte à l'intégrité de la barrière épithéliale, observée en coupe histologiques dans le colon (Gogokhia et al. 2019). Ces phénomènes contribueraient à l'entretien de la réponse inflammatoire et à une atteinte chronique de l'intégrité de la barrière intestinale chez les patients, au regard de la forte concentration en VLPs dans les filtrats fécaux observée au cours de cette thèse.

L'augmentation du nombre de VLPs dans l'intestin, dont il est fortement supposé qu'il s'agisse de phages, pourrait être à l'une des origines de l'inflammation intestinale au cours de la maladie ou favoriser son entretien. Plusieurs travaux semblent en effet pointer vers un rôle pro-inflammatoire de certains phages. C'est notamment le cas de phages de *Staphylococcus aureus* et de *Pseudomonas aeruginosa*, qui induisent la transcription d'IL-6, d'IL-10 et de TNF, tous trois pro-inflammatoires (Van Belleghem et al. 2017). Des effets pro-inflammatoires des phages ont été observés dans une autre étude conduite sur des cellules dendritiques, qui, en présence de phages, voient leur production d'IL-6, IL-10 et IL-12 augmentée. De plus, les auteurs rapportent une stimulation de la production d'une autre cytokine pro-inflammatoire, l'IFN- γ , par les cellules T CD4⁺ (Gogokhia et al. 2019).

À l'inverse, certains phages exerceraient une activité anti-inflammatoire. C'est le cas de phages filamenteux de *Pseudomonas aeruginosa*, Pf1 et Pf4, dont l'endocytose par les cellules dendritiques enclenche, *in vivo* chez la souris, une cascade de réactions conduisant à l'inhibition de TNF, dont l'activité est pro-inflammatoire leur permet, à eux et leur hôte bactérien, d'échapper à la vigilance du système immunitaire (Sweere et al. 2019). Une diminution de la réponse inflammatoire chez la souris a également été observée suite à l'injection intrapéritonéale de LPS couplé à la gp12, une protéine de queue du phage T4, par rapport à des injections de LPS seul (Miernikiewicz et al. 2016). La gp12 ainsi que la protéine de tête Hoc du phage T4 ont par ailleurs été corrélés à une augmentation de la réponse humorale via une augmentation de la sécrétion d'IgA dans les fèces de souris, suite à leur ajout dans l'eau de boisson de ces animaux (Majewska et al. 2015). Aucun lien de causalité n'a cependant été établi entre l'administration de ces protéines et leurs effets au cours de l'inflammation, qui pourraient être médiés par une modification de la composition du microbiote intestinal due à la présence de ces protéines. En effet, la présence d'IgA favoriserait la colonisation mucoale par certaines espèces bactériennes comme *Bacteroides fragilis* (Donaldson et al. 2018). L'ensemble de ces résultats tend à mettre en évidence l'interconnexion des phages et du système immunitaire et le rôle principal que certains d'entre eux auraient au cours de l'inflammation. Aucun résultat obtenu au cours de cette thèse ne permet cependant de déterminer si la présence des VLPs

observées pourrait avoir des conséquences pro- ou anti-inflammatoires dans le contexte de la maladie de Crohn. Cependant, l'abondance accrue de ces dernières au cours des périodes de poussée pourrait en partie expliquer la survenue de crises inflammatoires et/ou l'entretien de ces dernières chez les patients.

Une seconde hypothèse pouvant expliquer l'augmentation du nombre de VLPs ou favoriser le maintien d'une importante concentration en VLPs au cours des périodes de poussée inflammatoire est justement liée à la survenue de l'inflammation intestinale et de son impact sur les prophages du microbiote intestinal. En effet, il peut tout à fait être supposé que le stress inflammatoire présent dans la maladie de Crohn et plus particulièrement au cours des poussées, puisse jouer un rôle dans l'induction de prophages, comme cela a déjà été décrit chez la souris pendant une infection avec *Salmonella thyphimurium* (Diard et al. 2017). Cette hypothèse de l'induction des prophages s'ajoute à la survenue de diarrhées pour expliquer la diminution de la concentration en bactéries dans les selles de patients en poussée. Elle est également en accord avec l'augmentation de la quantité absolue de VLPs comptées dans leurs filtrats fécaux, d'autant plus qu'une plus forte proportion de phages tempérés est observée chez les patients atteints de la maladie de Crohn par rapport aux individus sains (Clooney et al. 2019, Fernandes et al. 2019, Norman et al. 2015). Au vu de la diminution de certaines espèces de bactéries antiinflammatoires au cours de la maladie (Machiels et al. 2014, Sokol et al. 2009, Willing et al. 2010) et de la présence de certains prophages dans les génomes de celles-ci, comme *F. prausnitzii* (Cornuault et al. 2018) et *R. intestinalis* (Cornuault et al. 2020), mes résultats vont dans le sens de l'hypothèse de l'induction liée au stress de prophages au sein de celles-ci (Cornuault 2018). Ces résultats semblent également être en faveur de l'occurrence d'un modèle écologique de type « community shuffling », au moins dans la lumière des portions distales du tube digestif. Par ailleurs, l'augmentation du ratio VE/bactéries pourrait justement relever d'une lyse partielle de la population bactérienne, la concentration en bactéries étant diminuée suite à l'induction de prophages, provoquant la formation de VE lors de la mort bactérienne (Mandal et al. 2021).

D'autre part, l'induction de prophages faisant suite à l'inflammation intestinale pourrait être à l'origine de phénomènes de transduction, dont la transduction latérale conduisant à l'encapsidation d'ADN bactérien adjacent au prophage (Misson, article en cours de révision, (Chen et al. 2018, Fillol-Salom et al. 2021)). Dans un contexte d'induction massive de prophages, un tel mécanisme pourrait participer à des transferts de gènes bactériens et conduire à des phénomènes

d'évolution bactérienne plus fréquents. Bien qu'aucun transfert de gène d'antibiorésistance associé à la transduction latérale n'ait pu être détecté au sein des viromes de lapins étudiés au cours de ma thèse ni lors de précédents travaux réalisés au sein de l'équipe (Billaud et al. 2021), il ne peut être exclu que d'autres gènes bactériens soient sujets à la transduction. Celle-ci pourrait aussi bien être latérale que généralisée (mécanisme conduisant à l'encapsidation accidentelle de fragments d'ADN bactérien, peu importe leur position au sein du génome (Kleiner et al. 2020)).

L'identification des virus les plus abondants présents dans les filtrats fécaux des patients semble être la suite logique des travaux à réaliser. Le séquençage des viromes permettra alors de déterminer, d'une part, si l'augmentation du nombre de VLPs pendant l'inflammation résulte du bloom d'espèces phagiques particulières ou s'il s'agit d'une augmentation générale du nombre de virions de la fraction phagique du microbiote dans son intégralité. Cette identification permettra également de déterminer les morphologies des virions détectés et de les confronter à celles des virus observés en MET dans les filtrats fécaux, qui pourraient être des représentants des *Microviridae*. Par ailleurs, des études métabolomiques devront être réalisées pour comparer l'origine des VE du microbiote intestinal. L'augmentation du ratio VE/bactérie observé lors de l'inflammation est probablement dû à la lyse bactérienne, au stress des cellules épithéliales et à l'apoptose. Sous les effets du stress inflammatoire, certaines populations bactériennes pourraient également produire plus de vésicules à des fins de communication ou de défense. Enfin, le séquençage du microbiote devra également être réalisé afin de constater l'émergence ou le déclin de populations bactériennes au cours de la maladie de Crohn.

Les résultats obtenus au cours de cette thèse suggèrent que la concentration des VE de plus de 100 nm de diamètre ne diffère pas dans les selles des individus sains et des patients inclus au sein des cohortes étudiées. Il est cependant impossible d'affirmer la même chose concernant les plus petites VE, dont l'abondance pourrait être modifiée au cours de la maladie. Si la quantité absolue de VE dans les selles reste encore une donnée difficile à déterminer au vu des limitations technologiques existantes, des taux plus élevés de cytokines pro-inflammatoires ont été détectés dans des VE isolées du contenu intestinal de patients atteints d'IBD que chez des contrôles sains (Mitsubishi et al. 2016). Les concentrations en VE ont par ailleurs été estimées par le NanoSight au cours de ce même projet, et aucune différence significative de leur concentration n'a été observée. Le cercle vicieux de l'inflammation pourrait ainsi être entretenu par la production de VE plus concentrées en cytokines

pro-inflammatoires et par la diminution hypothétique de la production de VE aux propriétés anti-inflammatoires. Ce pourrait notamment être le cas des VE de *F. prausnitzii*, dont les propriétés anti-inflammatoires ont été rapportées *in vivo* (Jafari et al. 2019). Ainsi, en plus de son impact via la production et la synthèse de composés anti-inflammatoires comme le butyrate et le peptide MAM (Machiels et al. 2014, Sokol et al. 2009), la diminution de *F. prausnitzii* au cours de la maladie pourrait également entretenir le cercle vicieux de l'inflammation via une production diminuée de VE anti-inflammatoires.

En dépit de sa prévalence élevée, la maladie de Crohn, et plus largement les MICI, restent à ce jour encore difficiles à diagnostiquer. En effet, c'est la combinaison de données cliniques, biologiques, endoscopiques, histologiques et d'imagerie, qui permettent, après de longs mois, voire des années, à l'équipe médicale d'établir le diagnostic. Cette errance diagnostique retarde ainsi fortement l'accompagnement et la prise en charge thérapeutique des patients. La découverte d'un biomarqueur spécifique serait un véritable graal permettant la réduction de cette période d'errance. L'éventualité de découvrir une augmentation importante de la quantité de phages dans le microbiote intestinal était, au départ de cette thèse, motivée non seulement par un désir de compréhension des écosystèmes digestifs au cours de la maladie de Crohn, mais aussi par l'exploration d'une nouvelle piste menant potentiellement à la mise au point d'un outil de diagnostic pour une pathologie en manquant cruellement.

La découverte de l'augmentation d'un facteur 40 du nombre de VLPs dans les filtrats fécaux de patients en poussée pourrait mener à l'initiation de projets de recherche sur le diagnostic de la maladie de Crohn. Toutefois, le chemin est encore long avant d'en arriver là. En effet, le protocole de dénombrement des VLPs est à ce jour chronophage (une journée de manipulations pour quelques échantillons) et nécessite l'acquisition et la maîtrise d'équipements coûteux. De plus, il reste encore à déterminer si cette augmentation de la concentration en VLPs dans les selles est spécifique de la maladie de Crohn ou si elle est retrouvée dans d'autres pathologies impliquant une inflammation intestinale, à l'instar de la rectocolite hémorragique, qui voit aussi la composition de sa fraction virale profondément modifiée (Gogokhia et al. 2019, Zuo et al. 2019). Au vu de l'observation en MET et de la quantification de nombreux VLPs dans les selles d'environ 10% des individus sains inclus dans ce travail de thèse, ainsi que de l'augmentation somme toute assez modérée de la concentration en VLPs chez les patients en période de rémission, il est probable que l'augmentation du nombre de

VLPs ne soit pas un marqueur spécifique de la maladie de Crohn mais le reflet d'une dysbiose du microbiote intestinal associée à d'autres conditions, pas nécessairement pathologiques.

Conclusion générale

Ce travail proposant de nouvelles méthodologies pour étudier les nanoparticules fécales met en évidence l'importance des phages au sein du microbiote intestinal au cours de la maladie de Crohn. Que leur concentration fortement augmentée lors des crises inflammatoires soit à l'origine ou une conséquence de l'inflammation intestinale, il est désormais évident que les phages sont impliqués dans cette pathologie. Ce travail ouvre la voie à de nouvelles recherches, en plaçant les phages comme potentiels biomarqueurs de la maladie et permettra, une fois leur nature élucidée, de mieux comprendre les interactions complexes intervenant dans le microbiote intestinal au cours des MICI.

Autres projets

Article de Revue

New insights into intestinal phages



REVIEW ARTICLE OPEN

New insights into intestinal phages

R. Sausset^{1,2}, M. A. Petit¹, V. Gaboriau-Routhiau^{1,3,4} and M. De Paep¹

The intestinal microbiota plays important roles in human health. This last decade, the viral fraction of the intestinal microbiota, composed essentially of phages that infect bacteria, received increasing attention. Numerous novel phage families have been discovered in parallel with the development of viral metagenomics. However, since the discovery of intestinal phages by d'Hérelle in 1917, our understanding of the impact of phages on gut microbiota structure remains scarce. Changes in viral community composition have been observed in several diseases. However, whether these changes reflect a direct involvement of phages in diseases etiology or simply result from modifications in bacterial composition is currently unknown. Here we present an overview of the current knowledge in intestinal phages, their identity, lifestyles, and their possible effects on the gut microbiota. We also gather the main data on phage interactions with the immune system, with a particular emphasis on recent findings.

Mucosal Immunology (2020) 13:205–215; <https://doi.org/10.1038/s41385-019-0250-5>

INTRODUCTION

The human gut contains a large number of viruses, mostly bacteriophages, or phages, which infect bacteria. As other viruses, phages are classified according to their type of nucleic acid, capsid morphology - notably the presence or absence of a tail - and the presence or not of an envelope. The genetic material of phages consists of double-stranded (ds) or single-stranded (ss) DNA or RNA, and their genome sizes range from ~3.5 kb (e.g., ssRNA genome of *Escherichia coli* phage MS2) to ~540 kb (dsDNA genome of *Prevotella* LAK phages). There is considerable diversity among phages, but 95% of them are non-enveloped tailed dsDNA phages, or *Caudovirales*. Within this group, the traditional differentiation into *Siphoviridae*, *Myoviridae*, and *Podoviridae* families, based on tail types, is not fully coherent with phylogeny, and therefore progressively abandoned. In addition, new phage types are constantly discovered, and classification is currently ongoing reorganization.

Phages are present in all microbial environments and the importance of phage predation on bacteria is evidenced by the large repertoire of bacterial anti-phage defence mechanisms. Anti-phage systems include cell-surface modifications that prevent phage recognition (phage multiplication is highly dependent on the proper selection of their target bacteria, which is achieved by the recognition of a specific structure on the bacterial surface¹), but also abortive infection mechanisms that trigger cell death upon phage infection and restriction-modification or CRISPR-Cas systems that cleave invading phage genomes (reviewed in refs. 2 and 3).

The presence of phages in the intestine has been described only 2 years after their discovery by Twort⁴ when d'Hérelle⁵ independently discovered phages, and their therapeutic potential, in the stools of patients with dysentery. Before the dawn of antibiotics, but also later on in the Soviet Union, phages have been utilized to treat a variety of intestinal infections, mainly cholera⁶ and dysentery.⁷ However, the success of these treatments

has been variable and antibiotics proved to be both more efficient and cost-effective, leading to the almost abandonment of phage therapy in most countries (reviewed in ref. 8). With the rise of bacterial resistance to antibiotics, phage therapy has recently regained interest, fueling researches on applied but also basic phage biology. The relatively recent discovery of the influence of phages in aquatic bacterial ecosystems further explains the present bloom of phage studies.⁹ Finally, due to increased awareness of the importance of the gut microbiota in human health, a growing number of studies are addressing the roles of phages in the gut microbiota. Emerging views suggest that intestinal phages play important roles in health and disease by shaping the co-occurring bacteriome, but also by interacting directly with the human immune system.^{10–12} Several recent reviews have exhaustively reported different aspects of intestinal phage biology, such as its genetic diversity,^{13,14} bacterial resistance mechanisms, including CRISPR-Cas systems and other molecular mechanisms of phage–bacteria interactions,^{2,3} phage–bacterium antagonistic interactions in the gastrointestinal tract (GIT),^{15,16} lysogeny,¹¹ and phage interactions with the host immune system.^{10,11,17} Here we aim at giving a global view of current knowledge of phages in the GIT, emphasizing on new results, open questions, and technical difficulties of this rapidly growing field of research.

COMPOSITION OF THE INTESTINAL PHAGEOME

Description of intestinal phages, either from a taxonomic or lifestyle point of view, is still in its infancy compared with that of intestinal bacteria, and encounters technical difficulties. First, viral genomes lack universal marker genes such as the 16S rRNA gene used for bacterial taxonomic assignment. Second, the genetic diversity of phages remains largely unknown, preventing sequence-based identification of most intestinal phages. Typically, 75% to 99% of sequences from intestinal phages do not produce

¹Micalis Institute, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France; ²Myriade, 68 boulevard de Port Royal, 75005 Paris, France; ³Laboratory of Intestinal Immunity, INSERM UMR 1163, Institut Imagine, Paris, France and ⁴Université Paris Descartes-Sorbonne Paris Cité, 75006 Paris, France
Correspondence: M. De Paep (marianne.depaep@inra.fr)

Received: 2 September 2019 Revised: 13 November 2019 Accepted: 5 December 2019
Published online: 6 January 2020

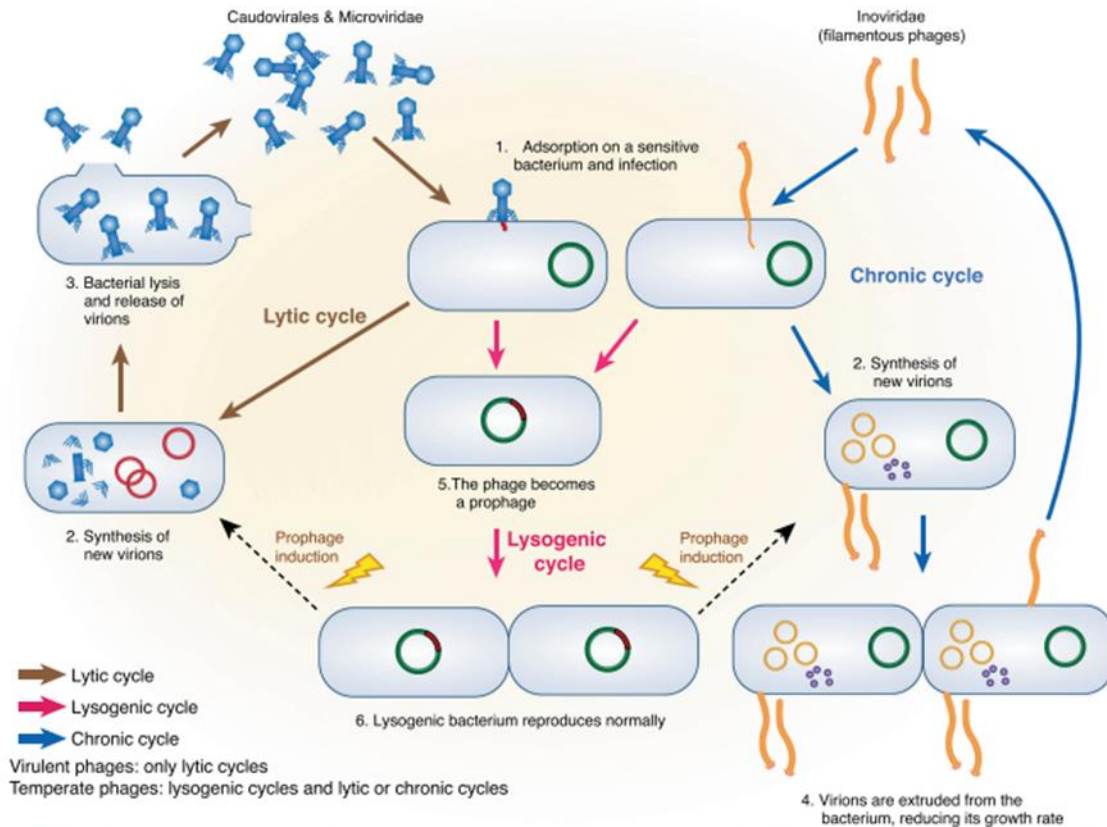


Fig. 1 Phage life cycles. The production of new virions is realized either through lytic cycles for *Caudovirales* and *Microviridae* phages (left side of the figure, brown arrows) or through chronic infection in the case of filamentous phages, or *Inoviridae* (blue arrows). Both start with the recognition and infection of the targeted bacteria (1), followed by phage DNA replication and synthesis of new virions (2). In lytic cycles, new virions are released through bacterial lysis (3), while new virions of filamentous phages exit bacteria through a dedicated secretion apparatus, without bacterial lysis (4). Phages that reproduce only through lytic cycles are called virulent. By opposition, some phages, called temperate phages, in addition to performing either lytic or chronic cycles, are able to perform lysogenic cycles (pink arrows), whereby they enter a dormant state in the infected bacteria, the prophage state (5). The prophage, either integrated within the bacterial genome or in an episomal state, is replicated with the bacterial chromosome as long as bacteria divide (6). In some bacteria, generally when submitted to a stress, the prophage is induced and the phage resumes a lytic or a chronic cycle.

significant alignments to any known viral genome.¹³ Finally, intestinal phages are very challenging to cultivate, notably because their bacterial hosts are mainly strict anaerobes that are difficult to grow. However, starting from 0.2 or 0.45 μm filtered fecal samples enriched in virions, shotgun deep sequencing has permitted access to the human free-phage content (which will be designated below virome or phageome, since it comprises mainly phages). Most phages appeared to be non-enveloped DNA viruses, either dsDNA *Caudovirales* or ssDNA *Microviridae*. In addition, a recent study indicates that ssDNA filamentous phages, or *Inoviridae*, that reproduce through chronic infection without killing their host (Fig. 1) might also constitute a significant fraction of the human gut virome.¹⁸ In contrast, RNA phages were found to be rare, if not completely absent, in the intestine.^{19,20}

In the intestine as in other environments, phages can be distinguished on the basis of their lifestyle, independently of taxonomy (Fig. 1). Virulent phages essentially complete lytic cycles, whereby each infection is followed by virion production and host cell lysis. Yet, in some conditions such as nutritional stress, virulent phage multiplication can be halted for a long time.

This phenomenon, called pseudolysogeny, is poorly described, but is suspected to exist in the intestinal environment (reviewed in ref. ²¹). Temperate phages, for their part, can use two very different lifestyles, the so-called "lysis-lysogeny choice": infection is either followed by a lytic cycle, as with virulent phages, or by lysogeny, whereby the phage enters a dormant state and is called a prophage. In this state, the expression of most phage genes is repressed, preventing phage multiplication, but the phage genome is replicated passively along with the bacterial genome. The prophage can be either integrated in the bacterial chromosome or extrachromosomal, like a plasmid. Following specific cues described below, prophages can be activated, leading to phage lytic cycle and death of the previously lysogenic bacteria. In consequence, when considering intestinal phages, one should take into account both free-phages and prophages.

Using recent assembly procedures, tens to thousands of DNA phage contigs can be assembled per virome sample, depending on sequencing depth. Yet, the functional roles of phages in the gut ecosystem remain difficult to apprehend, notably because they belong to entirely new and still uncharacterized genera and

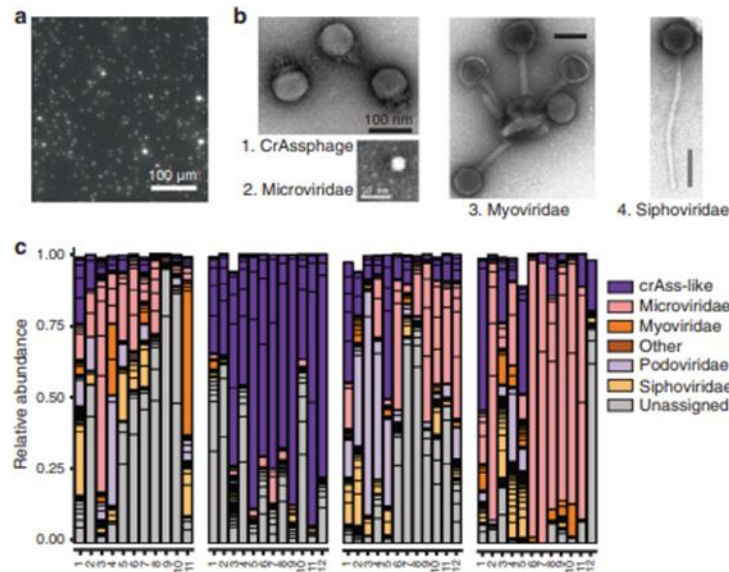


Fig. 2 Main intestinal phage types. **a** Epifluorescence microscopy image of a human fecal filtrate. Virus-like particles (VLPs) appear as bright dots following Sybr-gold staining of DNA (M. De Paep, unpublished results). **b** Transmission electronic microscopy images of major intestinal phage types. Scale bars are 100 nm except for the *Microviridae* virion, for which it is 50 nm. 1: CrAssphage PhiCrAss001 (*Podoviridae*), reproduced from ref. ⁴⁵ 2: *Microviridae* virion isolated from surface water near a coastal aquaculture site, reproduced from ref. ¹²⁶ 3 and 4: *Myoviridae* and *Siphoviridae* virions respectively, extracted from human intestinal contents. Several *Myoviridae* virions are bound to a membrane vesicle. Reproduced from ref. ⁵⁰ Copyright © 2014 Elsevier Masson SAS. **c** Inter-individual virome specificity and conservation in four human subjects over 12 months. Each phage cluster, which roughly corresponds to a phage genus, is outlined in black. Reproduced from ref. ²⁹

even families, and their bacterial hosts are unknown. Several recent reviews have summarized the main characteristics of the new findings in this emerging field.^{13,14,22} Their conclusions about the phage community in the human gut are summarized below.

- (1) Most phages appear unique to each individual.^{23–25} The existence of a “core phageome”, i.e., a small number of phages shared between individuals, is under debate. A first study indicated that across 62 healthy individuals, 23 phage contigs (0.5% of all contigs) were shared by at least 50% of individuals and 132 contigs (3%) by at least 20% of individuals.²⁴ More recently, another study found that no viral population (equivalent to viral species) was present in more than half of 132 samples from healthy individuals, and that only 1% was shared by over 20% of individuals.^{24,25} These differences depend on the criteria used to identify a given phage in a given virome sample and comparisons suffer notably from the lack of consensus on phage taxonomy. Recent remarkable progresses in genome-based phage taxonomy^{26,27} may enable to better define the most common phage “types” in the human microbiota in the near future.
- (2) Healthy individuals tend to conserve the same phages over time (tested over 1 year), especially the most abundant ones^{23,28,29} (Fig. 2). Persistent phages, i.e., phages that reside at least one year in a given individual, seem also to be more commonly shared than others. Thus, among persistent phages, 22 out of the 3639 viral clusters (0.6%; clusters might reflect genus level groupings) were shared by more than half of the 10 individuals tested.²⁹
- (3) In the human gut, temperate phages constitute at least 20% to 50 % of free phages, depending on samples but also on the methodology, as temperate phages are difficult to

distinguish from virulent ones. For example, a study focusing on the virome content of twins estimated the proportion of temperate phages on the basis of the proportion of phage contigs encoding an integrase gene, this gene being responsible for the integration of temperate phage genome into the bacterial chromosome.²³ However, temperate phages do not necessarily encode an integrase gene, as many of them coexist with their host as autonomously replicating episomes, and such method gives only a lower bound of the proportion of temperate phages. The high abundance of temperate phages is compatible with the observation that among isolated gut bacteria, most strains are lysogens and often polylysogens, notably within the dominant *Bacteroidetes*, *Firmicutes*, *Actinobacteria*, and *Proteobacteria* phyla.^{30–34} Several metagenomic studies further suggest that the majority of intestinal bacteria carry prophages.^{35,36} Yet, estimations of prophage prevalence based on genomic analyses cannot distinguish “active” prophages from “defective” ones, i.e., those that have lost the capacity to resume a lytic cycle due to mutations. However, the presence of active prophages, which produce virions, has been demonstrated experimentally in almost all gut bacterial strain tested, suggesting that a significant proportion of prophages detected by genomic analyses are active.^{33,34,37,38} High prevalence of temperate phages in human microbiota could participate in the stability of the phageome over time, as populations of lysogenic bacteria constantly produce virions.

- (4) The identity of the hosts targeted by the phages is a crucial but largely unanswered question. In 2016, Edwards et al.³⁹ compared different methods for host prediction and those that gave the best results were simple nucleotide similarity searches between phage and bacterial genomes, either with

BLASTn or by searching the longest exact nucleotide match between a phage and a bacterial genome. With these methods, the correct hosts were predicted for 37% and 40%, respectively, of the 820 complete phages included in the test. Several studies have used CRISPR spacers to predict hosts,^{29,39–41} which led to very confident predictions but was restricted to hosts encoding CRISPR–Cas systems and in which phage infection was relatively recent. Between 4% and 13% of phages could be assigned to a host this way. Finally, the program WISH bases its predictions on the similarity of the phage genome to that of its hosts. It uses a probabilistic approach that compares the composition in subsequences of nine nucleotides, or 9-mers, in phage and bacterial genomes, and reaches good prediction even for short 3 kb-long phage contigs, which is not the case for other methods. Using the same large data set as Paéz-Espino et al.,⁴¹ WISH predicted a host at family level for 59% of the contigs.⁴² Using a mix of these approaches, recent studies indicate that the spectrum of the bacterial hosts of the dominant phages reflects the microbiome composition.^{29,43} Indeed, among 180 persistent phage clusters identified, about one third could be linked to a bacterial genus, all of them belonging to abundant taxa, such as *Faecalibacterium* and *Bacteroides*.²⁹

- (5) In individuals consuming a western diet, two particularly prevalent and abundant phage taxa have been recently described, both infecting *Bacteroides* species (Fig. 2):

Gokushovirinae are temperate phages belonging to the *Microviridae* family, with a small circular ssDNA genome, and a small virion of 30 nm of diameter.³¹ Given the nature of their genome, these virions can escape scrutiny or, on the contrary, be overestimated depending on the kit and treatments used prior to sequencing (discussed in ref. 14). Therefore, their exact contribution to the phageome remains to be more accurately estimated.

CrAss-like phages constitute a completely new clade of related phages present in at least 50% of individuals from western cohorts. They can represent up to 90% of the phageome of a single individual. CrAss-like phages have a dsDNA genome of ~100 kb, a 90 nm diameter large head and a short tail (Fig. 2b). Their abundance has helped assembling many genomes of the clade, which presently span over four proposed subfamilies.⁴⁴ One of them, CrAss001, infects *Bacteroides intestinalis*.⁴⁵ Due to difficulties in cultivating the CrAssphages, their lifestyle is not clearly established yet, but the presence of an integrase in some of them suggests they may be temperate phages.²⁹

In addition, a distinct prevalent phage taxon, called LAK phages, has recently been described in individuals from Tanzania and the region of Laksam, in Bangladesh. Their genomes were assembled directly from complete microbiota samples rather than using viral fractionation.⁴⁶ Due to their remarkably large genome size (540 kb), they were coined “megaphages”, but their virions are not imaged yet. LAK phages have been shown to preferentially infect *Prevotella*, but their lifestyle is still unknown.

QUANTIFICATION OF PHAGES IN THE GIT

Defining the number of phages present in the GIT should help to predict their impact on bacteria. As already stressed, both free phages and prophages should be taken into account. Prophages are present in the majority of bacterial genomes and are thus approximately as numerous as bacteria in the GIT. Free phages, produced during lytic cycles, are generally enumerated as virus-like particles (VLPs). VLPs are nanoparticles that can be observed by epifluorescence microscopy following staining of nucleic

acids (Fig. 2a). However, if such nanoparticles are mainly virions, some are membrane vesicles containing nucleic acid that are difficult to distinguish from virions by microscopy.⁴⁷ In addition, immunologists often define VLPs as particles obtained by spontaneous assembly of viral structural proteins, antigenically indistinguishable from infectious viruses, but that do not contain the viral genome.⁴⁸ Despite these restrictions, we will use indiscriminately virions and the microscopy-based definition of VLPs below. In most ecosystems, free phages outnumber bacteria by about tenfold.⁴⁹ In contrast, the small number of available studies enumerating virions in human stools indicates that in healthy subjects, bacteria outnumber virions. Two studies reported VLP abundances ranging from 1×10^8 to 2×10^9 VLPs/g in stools,^{50,51} whereas bacterial concentration is thought to be 1×10^{11} bacteria/g.⁵² Recently, the concentration of virions was also estimated by spiking viral metagenomic samples with a known concentration of a specific virus, indicating that total viral loads may be between 2×10^8 and 8×10^{10} viral genome copies per gram of stool.²⁹ The ratio of free phages to bacteria that can be extrapolated is thus comprised between 1 and 0.001, overall suggesting lower phage-induced mortality in the human gut microbiota than in other microbial ecosystems. An alternative interpretation is that a large fraction of virions are “lost” in the GIT. Several factors may favor virion disappearance. First, virions bind irreversibly to their bacterial receptors with a very fast rate when receptor concentration is high.⁵³ Such receptors can be present on the surface of susceptible bacteria, but also on the surface of resistant bacteria or even on other structures such as membrane vesicles or bacterial debris. In these last cases, binding of virions on their receptor results in their inactivation. As the bacterial concentration in the GIT is much higher than in most other microbial ecosystems, a higher part of virions could be lost by such a mechanism. For example, binding of virions on membrane vesicles is regularly reported in intestinal samples (Fig. 2b). Second, phages may be captured inside the mucus layer. Indeed, the Ig-like protein domain on the capsid of phage T4 was shown to increase by fourfold virion binding to mucins *in vitro*.⁵⁴ As numerous intestinal phages possess similar motifs on their capsid proteins,⁵⁵ it may account for the large number of virions detected in the intestinal mucus.⁵⁶ Whether these mechanisms operate in the human microbiota remains to be investigated. Experiments in gnotoxenic mice (i.e. colonized with a limited number of bacterial strains), in which the majority of the susceptible bacteria were killed by phage, suggest that virion loss is highly dependent on individual phages considered. In some reports, free phage to bacteria ratios over 100 were observed in the feces of mice,^{37,57} whereas in others, ratios close to 1 were observed,^{58,59} suggesting either virion loss or very low viral amplification per infected bacteria, i.e., about one new virion produced per lytic cycle.

PHAGE–BACTERIUM INTERACTIONS IN THE GIT

Phage–bacterium interactions in the gut microbiota are highly complex. First, phages often interact specifically with one single bacterial strain. Given bacterial diversity in the human gut, hundreds to thousands of phage–bacteria pairs may potentially interact at any time. Second, as discussed below, the intestinal environment seems to protect a fraction of the genetically susceptible bacterial populations from phage infection. Third, both antagonistic and mutualist interactions are possible. For example, bacteria can be killed by phages, either by infection followed by lytic cycle (predation) or upon prophage induction, but on the contrary, as we will see below, prophages and filamentous phages can provide benefits to their host. Finally, due to the high evolutionary capacities of phages and bacteria, phage–bacteria interactions may change very rapidly over time. Below we discuss the phage–bacterium interactions that have

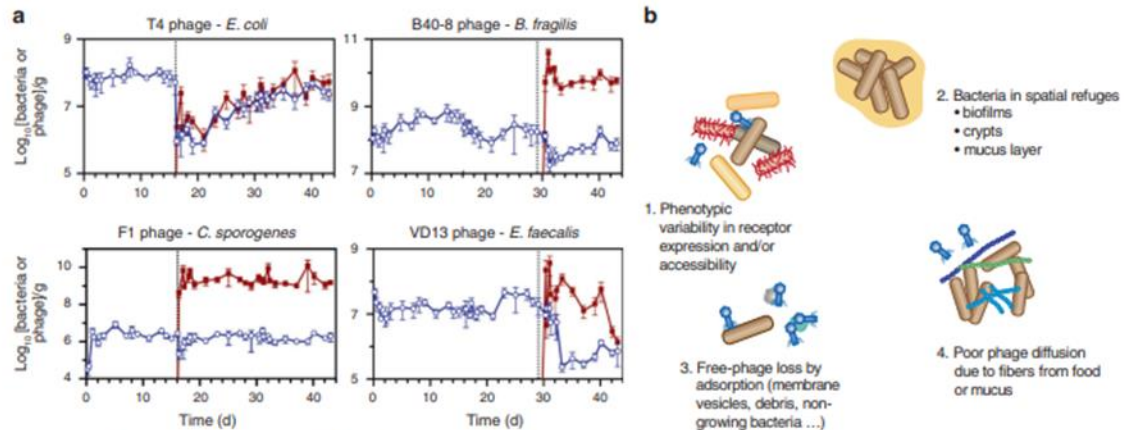


Fig. 3 Phage-bacterium interactions in the GIT. **a** Illustration of the variability in the populations of phages and targeted bacteria in feces of gnotobiotic mice, reproduced from ref. ⁵⁷ Germ-free mice were colonized with ten bacterial strains representing the major phyla of the human gut microbiota (*Clostridium sporogenes*, *E. faecalis*, *Bacteroides fragilis*, *B. ovatus*, *B. vulgatus*, *Parabacteroides distasonis*, *Klebsiella oxytoca*, *Proteus mirabilis*, *E. coli* Nissle 1917, and *Akkermansia muciniphila*). Dotted lines indicate the time of addition of 2×10^6 PFU of a virulent phage targeting one bacterial strain exclusively. Bacteria and phages were quantified by quantitative PCR (blue and red lines, respectively). In the case of *E. coli*, after a rapid drop in bacterial population after T4 phage administration, phage and bacteria coexist without selection of phage-resistant mutants. In the cases of *C. sporogenes* and *B. fragilis*, bacterial populations are transiently affected by phage administration, and phages reach very high concentrations. Finally, in the case of *E. faecalis*, the bacterial population drops and does not recover its initial level over time, due to a defect of colonization by phage-resistant mutants. **b** Hypotheses for the co-existence of phage and susceptible bacteria in the GIT. (1) Important bacterial phenotypic variability, e.g., in phage receptor expression or growth rate, would render some bacteria phage-susceptible, whereas other genetically identical bacteria would be resistant. (2) Spatial heterogeneity of the environment (such as intestinal crypts, biofilm-like structures on food debris, or the inner part of the mucus layer) would create a refuge for bacteria and prevent access to phages. (3) Important loss of virions due to specific and nonspecific adsorption would sufficiently lower free-phage concentration to protect bacteria. (4) The presence of numerous long chain carbohydrates fibers from food and mucus would hinder phage diffusion, preventing them to adsorb on bacteria. None of these mechanisms has been formally demonstrated in the GIT.

been demonstrated in the GIT and their impact on the gut microbiota composition.

Prophage induction

Most prophages are highly stable, but environmental stressors or stochastic fluctuations in phage repressor concentrations can trigger their induction, i.e., resumption of the lytic cycle and subsequent lysis of the host bacteria. In general, cellular signals triggering prophage induction are DNA damages, via the destabilization of the repressor or master regulator of lysogeny. Quinolone antibiotics, which cause DNA double-strand breaks, are the most described prophage inducers in the GIT.⁶⁰ In consequence, as *E. coli* stx prophages encode the shiga toxin, treatment of human shigatoxigenic *E. coli* infection with quinolones have significant adverse clinical consequences.⁶⁰ Spontaneous induction rates, estimated in vitro to be between 10^{-7} and 10^{-4} , were generally considered to be too low to negatively impact the lysogen's fitness (reviewed in ref. ⁶¹). Yet, several pieces of evidence suggest that induction rates are globally higher in the murine GIT than in classical in vitro growth cultures, due to more frequent activation of the DNA damage response (SOS response).^{58,62,63} In the case of *Lactobacillus reuteri*, SOS activation was proposed to result from the activation of specific bacterial metabolic pathways in the GIT.⁶² Conflicting results were reported concerning the possible increase or, on the contrary, decrease of *E. coli* stx prophage induction rates by gut metabolites such as nitric oxide or bile salts (reviewed in ref. ⁶⁴). Bile salts were also shown to induce some *Salmonella* prophages.⁶⁵ Besides, intestinal inflammation increases the induction of a *Salmonella* prophage in mice.⁶⁶ Finally, one has to mention the recent discovery of quorum-sensing pathways in phages of *Enterococcus faecalis* and *Vibrio cholerae*, suggesting that some prophages could regulate their induction rate in response to

bacterial concentration.^{67,68} Altogether, these results suggest that in the GIT, prophage induction might constitute a significant burden for their bacterial host, and alter microbiota composition. Indeed, it has been shown in mouse models that a high prophage induction rate modifies the equilibrium between bacterial strains by disfavoring lysogens.^{38,58}

Lysogenic conversion

Some prophages compensate their detrimental effects on their host by providing them beneficial traits that augment their fitness and can confer completely novel phenotypes, a phenomenon known as "lysogenic conversion." Examples of lysogenic conversion are numerous and recently reviewed by Taylor et al.⁶⁹ and Wahl et al.⁷⁰ in *Salmonella*. These new phenotypes comprise immunity to phage super-infection, resistance to other phages, tolerance to various stresses, pathogenicity, and, very rarely, antibiotic resistance. With respect to antibiotic resistance genes (ARGs), their presence within phage genomes is debated. Most reports are flawed by either excessive bacterial DNA contamination of the virome samples⁷¹⁻⁷⁴ or inappropriate thresholds used for the similarity search against ARG databases.⁷⁵ When both pitfalls were eliminated, ARG were rare in phage genomes (around or <1 gene among 10^5 genes analyzed, on phage genomes and phage metagenomic contigs respectively),⁷⁶ confirming an ancient observation.⁷⁷ Further illustrating the potential of prophages to confer fitness advantages to their bacterial hosts, a very recent study analyzing sponge-associated virome suggested that some prophages encoding ankyrin-repeat proteins may provide anti-inflammatory properties to their bacterial host and foster host-microbe symbiosis.⁷⁸

Independently of lysogenic conversion, a few prophages can function as genetic switches, a phenomenon known as "active lysogeny" (reviewed in ref. ⁷⁹). In active lysogeny, prophage

excision does not result in lytic replication, but restores the integrity of the bacterial gene or operon in which the prophage was integrated. For the intracellular pathogen *Listeria monocytogenes*, active lysogeny was shown to facilitate bacterial escape from the phagosomes.⁸⁰ In a third situation, the production of virions itself benefits the lysogen population (reviewed in ref. 61). For example, production of filamentous phage particles by *Neisseria meningitidis* promotes bacterial aggregation in vitro via the formation of bundles of phage filaments, supposedly increasing the bacterial colonization of the nasopharynx.⁸¹

Due to the beneficial traits conferred to bacteria by prophages, temperate phages are often considered as mutualists rather than parasites of their bacterial host, and the prevalence of lysogeny is often interpreted as an evidence that prophages increase the selective fitness of their hosts (reviewed in ref. 82). However, if benefits conferred by prophages to pathogenic bacteria are well demonstrated, to the best of our knowledge, there is no direct demonstration that lysogenic conversion increases the colonization ability of a gut commensal in vivo, except very transiently by phage killing of a closely related strain.^{58,83} Alternatively, the prevalence of lysogeny may reflect regular lysogenization of bacteria by temperate phages, as evidenced in the GIT of animals,^{58,66,84,85} and even in the human gut microbiota.⁸⁶

Predation

A large body of information on phage predation in the GIT comes from animals with a modified microbiota, either gnotobiotic or treated by antibiotics, and colonized with specific bacterial strains and their associated phages. In all cases, phages multiplied successfully in the GIT. Yet, the decrease in the targeted bacteria is very variable (Fig. 3a).^{57,59,87,88} In some experiments, phage-resistant bacterial mutants were observed. However, and in contrast with results obtained in vitro, these mutants never entirely replaced the susceptible bacteria.^{37,57,89} In other cases, phage-mediated bacterial mortality was relatively limited and neither depletion of the bacterial population, nor selection of phage-resistant mutants were observed.^{57,59,87,88} Overall, these data suggest that the intestinal environment provides bacteria with spatial refuges or alternative resistance mechanisms, such as phenotypic resistance (Fig. 3b). For example, bile salts were shown to inhibit the infection of *E. coli* by several phages in vitro, possibly via the repression of a phase-variable cell-surface protein.⁹⁰ Variability in phage resistance can also result from spatial heterogeneity in the bacterial population driven by gradients of abiotic factors, such as pH and oxygen, but also of molecules such as mucins, bile acids, and short-chain fatty acids. These gradients can modify the physiology of bacteria and, consequently, their susceptibility to phages, as shown ex vivo in Maura et al.⁹¹ Fluxes may also prevent phages to access some compartments, such as intestinal crypts.^{15,91} Finally, the crowded environment of the GIT may hamper phage diffusion (reviewed in refs. 92 and 93). Overall, most current reports point to a transient effect of phages on the population sizes of targeted bacteria in the GIT microbiota of animal models. Nevertheless, successful phage therapy trials realized before the 70s against intestinal bacteria indicate that phages may efficiently reduce bacterial loads in the human GIT.^{5–7}

Another related question is the importance of phage-mediated mortality in a "natural" microbiota, i.e., without the artificial ingestion of high phage doses. At the timescale of evolution, phage killing left traces in the genomes of intestinal bacteria in the form of a large repertoire of anti-phage systems and highly diverse CRISPR spacer sequences.⁴⁰ However, such traces of phage-mediated selection could result from a low level of phage-induced mortality, as even a small selective advantage can lead to allele fixation if sufficient time is given. At a shorter ecological timescale, there are also several evidences of significant natural phage-mediated bacterial mortality in humans and animals. First, a study by Seed et al.⁹⁴ strongly suggests that mutations

conferring phage resistance to *V. cholerae* occurred and were selected in two patients during cholera infection. In addition, metagenomic studies conducted on conventional animal's feces showed a correlation between increases in specific phages and reductions in particular bacterial taxa, but without definitive proofs of direct interactions.^{35,95}

Lastly, an interesting but poorly investigated phage–bacteria antagonistic interaction in the GIT is the potential ability of some phages to infect simultaneously several bacterial strains belonging, or not, to the same species. Although a large amount of work has indicated that most phages infect only a few strains of the same bacterial species, several clues suggest that intestinal phages might be more promiscuous than usually assumed. First, a phage infecting *Faecalibacterium prausnitzii* was shown to also infect *Blautia hansenii*, a species only distantly related to *F. prausnitzii*.³³ Second, metagenomic analyses of CRISPR spacers have suggested that some phages may have a broad spectrum of hosts in the human gut.²⁹ In addition, phage evolution resulting in the ability to infect a new host (host jump) was reported in a mouse model of coevolution in the GIT.⁸⁹

In conclusion, phage–bacteria interactions in the GIT appear to be highly diverse, involving partial bacterial resistance, fast evolution of phages and bacteria, lysogenization and probably indirect cascading effects on several bacterial species and their phages.

IMPACT OF PHAGES ON HOST THROUGH GUT MICROBIOTA MODULATION

Phage therapy of intestinal disease infections

The complex relationship between phages and bacteria probably explains the relative unpredictability of phage therapy trials against intestinal pathogens. Phage therapy was nearly abandoned in most western countries 60 years ago given its unpredictable outcomes and the efficacy of newly discovered antibiotics. The compassionate use of phage therapy has been extensively practiced, though, in several countries, notably in Poland. The Eliava Institute, in Georgia, has been routinely treating various infections, and notably gut infections, with phage cocktails for more than a century, but no evaluation is available. An evaluation of Polish treatments reported symptomatic amelioration and clinical recovery in 40 and 18% of patients, respectively, which is quite encouraging for patients who previously failed to respond to antibiotic treatments.⁹⁶ The mounting incidence of severe infections by multi-drug resistant bacteria, along with recent progresses in phage biology has prompted re-evaluation of phage therapy. A randomized controlled clinical trial was recently conducted against *E. coli* diarrhea in Bangladesh. It was unfortunately unsuccessful. Yet, it was subsequently shown that *E. coli* was probably not the cause of the diarrhea.^{97,98} This large scale study conducted up to stage III nevertheless permitted to establish the innocuous character of phage cocktails for humans.⁹⁸ Besides classical phage therapy with defined phage mixtures, transfer of fecal filtrates containing phages but no bacteria, was proposed as an alternative to fecal microbiota transplantation (FMT) for *Clostridium difficile* infection.⁹⁹ The transfer of fecal filtrates to five patients with chronic *C. difficile* infection was successful in all patients, suggesting that phages present in the fecal filtrates mediate many of the effects of FMT, but the mechanisms remain unknown.⁹⁹

Impact of phages on individuals with a healthy microbiota

Beyond phage therapy, given the numerous interactions between intestinal bacteria and the host, it is important to evaluate the extent of phage-induced modifications of microbiota functionality. First, as discussed previously, prophages modify numerous bacterial phenotypes, and notably virulence, thereby impacting host–bacteria interactions^{69,70,78}. In particular, the ankyphages

recently discovered in sponges might contribute to promote bacterial protection from the eukaryotic immune system.¹⁰⁰ Indeed, transformed *E. coli* bacteria expressing a phage ankyrin protein better resisted murine bone-marrow-derived macrophage phagocytosis and induced a diminished inflammatory cytokine response in these macrophages. Interestingly, phage contigs encoding similar ankyrin proteins were found in various host-associated environments, including human oral and gut viromes, suggesting some wide-range phenomenon.¹⁰⁰

Second, beyond lysogeny, phages can impact the mammalian host physiology through microbiota modulation, by depleting bacterial species important for homeostasis. However, as discussed previously, in animal models, phages never fully eradicate bacterial species in the gut, due to partial protection of bacteria and fast evolution of resistant bacterial mutants. Effects of phages on microbiota composition at the species level are therefore expected to be transient, unless they act in conjunction with other factors, such as other members of the microbiota. Studies aiming at detecting cascading effects on bacteria not targeted by the phage introduced have brought contrasting results. Studies in gnotobiotic rats and mice showed that beyond phage-mediated targeted knockdown of susceptible species in the gut, the concentration of non-targeted bacteria could be impacted, notably through modification of bacterial interactions.^{57,101} However, other studies in conventional animals showed no modification of the non-targeted microbiota.^{102,103} Similarly, a modification of the human microbiota composition has been observed upon oral administration of phages in one study,¹⁰⁴ but not in another involving children.¹⁰⁵ Such differences could be related to the difficulty of attributing shifts in microbiota composition to phages or to its natural temporal variability.

Phages could also impact indirectly the host if phage resistance affects the ability of bacteria to colonize the gut and to interact with their host. Indeed, many bacterial structures used as phage receptors, such as lipopolysaccharides, porins, flagella, fimbriae and pili, play a key role in bacterial colonization, growth, persistence, virulence and recognition by the host's immune system in the GIT (reviewed in refs. 2,16). For example, mutations in porins such as TonB or LamB, conferring phage resistance in *E. coli*, can diminish the nutritional competence of the bacteria, possibly affecting its gut colonizing ability. In particular, phage resistance was shown to negatively impact the ability of bacteria to develop antibiotic resistance, as bacterial mutants resistant to phage were more sensitive to antibiotics. For example, mutations in *Pseudomonas* conferring in vitro phage resistance affected a porin involved in multi-drug efflux,¹⁰⁶ whereas in the case of *E. faecalis*, *epa* mutations changing its exopolysaccharide composition rendered the strain more sensitive to cell-wall-targeting antibiotics in mice.¹⁰⁷ Moreover, several *E. faecalis epa* mutations could simultaneously confer phage resistance^{107,108} and defective intestinal colonization in mice.¹⁰⁹ Indeed, a promising type of phage therapy consists in selecting phages on their faculty to favor the emergence of phage-resistant bacterial mutants that become antibiotic sensitive.¹¹⁰ Phages may also indirectly benefit their mammalian host by increasing the diversity of microbial communities,^{16,111,112} a known factor of stability of the gut microbiota.¹¹³ Finally, some phages can perform generalized transduction, permitting the transfer of bacterial DNA from an infected cell to another. Some data indicate that a substantial fraction of both temperate and virulent phages can perform such transduction (reviewed in refs. 114), so that this phenomenon may take place regularly in the GIT microbiota, participating in bacterial evolution.

A role of phages on dysbioses?

Several studies have investigated correlations between virome composition and the physio-pathological state of individuals.

One study notably reported a global shift in virome composition in patients with inflammatory bowel disease (IBD), either Crohn's disease or ulcerative colitis, compared with household controls.¹¹⁵ Whether such shift implies that phages are actively involved in disease, or simply reflects the shift in microbiota composition remains to be shown. Nonetheless, re-analysis of the data focusing on the temperate phages of *F. prausnitzii*, a species generally depleted in IBD patients, showed that two *F. prausnitzii* phages were significantly more prevalent in IBD samples and two others were significantly more abundant,³³ suggesting enhanced temperate phage-mediated mortality of *F. prausnitzii* in IBD. One may speculate that the inflammatory environment of the gut might increase prophage induction, as shown in *Salmonella*,⁶⁶ thereby promoting bacterial lysis and aggravating dysbiosis, and reinforcing the inflammatory loop.

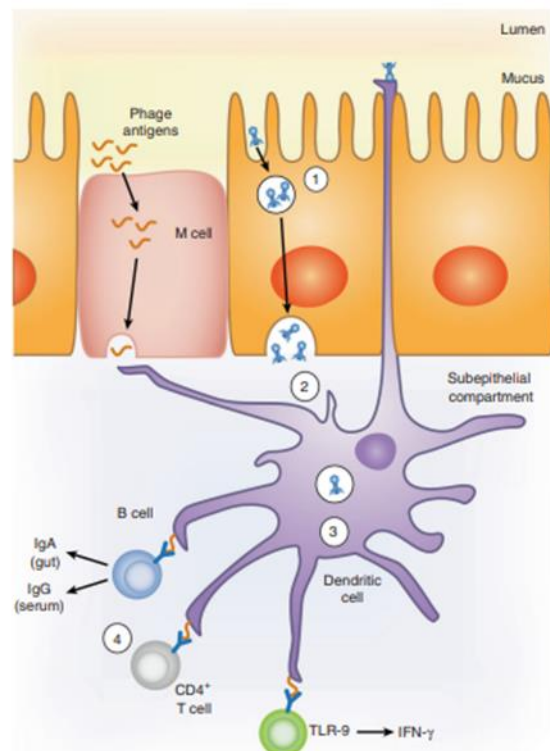


Fig. 4 Potential interactions between phages, epithelial cells, and host immune cells in the intestine. Phage tropism for the mucus may promote penetration of phages within the body through endocytosis and transcytosis in intestinal epithelial cells (1), or through sampling by dendritic cells (2). Dendritic cells endocytose phages, may be captured in the intestinal lumen via extended dendrites, or exocytosed in the subepithelial compartment. Once endocytosed, phage nucleic acids can trigger TLR pathways, notably TLR9-dependent pathways (3), and stimulate adaptive immune responses (4). Although mechanisms of B and T-cell activation by phages are not fully elucidated, recent studies showed that activation of B cells leads to the secretion of phage-specific antibodies, both in the intestine and in the systemic compartment. In addition, activation of T cells in the Peyer's patches and mesenteric lymph nodes results in production of cytokines, such as IFN-γ.

INTERACTIONS BETWEEN INTESTINAL PHAGES AND THE MAMMALIAN IMMUNE SYSTEM

In addition to their indirect impact on hosts immune responses through changes in microbiota, phages may also interact directly with the host immune system and trigger immune responses (Fig. 4). Whether intestinal phages disseminate outside the gut lumen and interact with immune cells in the intestinal mucosa or at distant sites is a first matter of debate. Some studies in rodents and humans stated that no or few phages were detected in the bloodstream after oral administration. In contrast, others demonstrated effective dissemination, largely correlating with the dose of ingested phages¹¹⁶ (and reviewed in refs. ^{11,12}). In addition, phage-neutralizing antibodies against naturally occurring intestinal phages have been detected in the sera of different mammalian species, indicating phage contact with host immune cells (reviewed in ref. ¹⁷). A recent study further showed that repeated oral exposure to high doses of the *E. coli* phage T4 induced simultaneously a specific protective secretory IgA response in the gut and a serum IgG response.¹¹⁶ Overall, these data suggest that phages can translocate through the mouse gut epithelium and activate the host immune system both locally and systemically. Translocation may occur across epithelial cells. Accordingly, recent *in vitro* data showed oriented phage transcytosis across several epithelial cell layers.¹¹⁷ Binding of phages to mucin glycoproteins⁵⁴ may further promote contact with epithelial cells and transcytosis. Alternatively, phage uptake may involve dendritic cells. Indeed, dendritic cells exposed *in vitro* to phage particles were shown to efficiently phagocytose these particles^{118–120} and dendritic cells are thought to extend dendrites through the intestinal epithelium and thereby sample luminal bacteria and particles.¹²¹ However, the incidence of this phenomenon seems rare in the steady state, suggesting that luminal phage uptake by dendritic cells may be a relatively rare event.

Although more work is required to establish how phages can get access to host immune cells *in vivo*, recent studies have analyzed their interactions with host intracellular immune pathways and their capacity to trigger immune responses *in vitro* and *in vivo*. Following the unexpected observation that the treatment of mice with a cocktail of antiviral drugs aggravated dextran sulfate sodium (DSS)-induced colitis, Yang et al.¹²² suggested that gut resident viruses, a large fraction of which being phages, protect against intestinal inflammation via a mechanism involving TLR3 and TLR7-dependent production of interferon (IFN)- β . In this study, it is however difficult to assign a role to phages in TLR triggering as the antiviral drugs used target only eukaryotic viruses. Moreover, antiviral treatment was associated with changes in the microbiota, which may have also contributed to stimulate protection. Three other studies suggest that phages also directly interact with innate immune cells. Van Belleghem et al.¹²³ observed that *in vitro* incubation of peripheral blood monocytes with purified *Staphylococcus aureus* or *Pseudomonas aeruginosa* phages induced a transcriptional response in monocytes, and notably enhanced the transcription of interleukin (IL)-1, IL-6, and tumor necrosis factor (TNF). Strengthening this *in vitro* study, Gogokhia et al.¹² observed the expansion of IFN- γ -producing CD4⁺ T cells and of CD8⁺ T cells in the Peyer's patches of germ-free mice orally treated with a lipopolysaccharide-free fraction of purified *E. coli* phages. They further showed that dendritic cells incubated with phages or phage-derived DNA could produce several cytokines including IL-12, IL-6, and IL-10, and stimulate CD4⁺ T-cell production of IFN- γ through a TLR9-dependent, but TLR3-independent, signal. In contrast with the effect of eukaryotic viruses reported by Yang et al.¹²² oral administration of the phage cocktail aggravated DSS colitis in a TLR9- and IFN- γ -dependent manner.¹² Noticeably, the authors made a parallel with their observations in patients with ulcerative colitis treated by fecal transplantation, showing that the relative intestinal abundance of *Caudovirales* bacteriophages was higher in patients who failed to

respond than in those who responded to this treatment.¹² A second recent study provided evidence of TLR3-triggering by a filamentous phage present in the *P. aeruginosa* strains infecting chronic human wounds.¹²⁴ Activation of TLR3 required endocytosis of phages by immune cells and, much surprisingly, neosynthesis of phage RNA into the immune cell, which resulted in the production of type I IFN. This cytokine, in turn, inhibited the production of TNF by macrophages, thereby impairing phagocytosis and bacterial clearance, and delaying wound healing.¹²⁴ Interestingly, this study suggests that phage uptake by mammalian immune cells, rather than simple cell-surface interactions, may be necessary to trigger host immune responses.

Overall, a limited set of data suggests that phages may exert a direct effect on immune cells. More work is however needed to define whether the effects observed *in vitro* or *in vivo* upon oral gavage with large amounts of purified virions can be recapitulated in more physiological conditions in the gut and how these effects may interfere or synergize with the changes induced by phages in the composition of the microbiota to impact mammalian health and disease, both in and out of the intestine.

CONCLUSIONS

During the last decade, viral metagenomics has shed light on the taxonomic composition and dynamics of the viral component of the gut microbiota. Deep sequencing and novel assembly methods have allowed the description of completely new phages. These approaches have notably revealed that the virome composition is highly variable, with only a small fraction of phages shared among individuals. Not surprisingly, the most abundant viruses were found to infect *Bacteroides* and *Clostridiales* species that are dominant in the microbiota. In parallel, experiments in animals with a simplified microbiota allowed to explore phage–bacteria antagonistic interactions in the gut and have uncovered variable outcomes. For some phage–bacteria pairs, the intestinal environment somehow protects bacteria and only a small fraction of genetically susceptible bacteria are killed by their specific phages. In other cases, most of the phage-susceptible bacterial population is replaced by resistant mutants within a few days, indicating very efficient phage infection. Therefore, which outcome will prevail in complex natural microbiota remains difficult to predict. Traces of phage predation can be found in gut bacterial genomes, but these traces could result from a low level of phage-induced mortality. A low level of phage predation could nevertheless be crucial to shape microbiota composition and functionality, by affecting bacterial evolution through horizontal gene transfer, but also by promoting bacterial diversity. At shorter ecological timescales, the impact of phages is less well established. In particular, the role of phages in the dysbioses that accompany various pathological conditions remains poorly defined. Thanks to recent progresses in the determination of phage–bacteria pairs, longitudinal studies can now be undertaken to identify possible relationships between temporal shifts in bacteria and their associated phages and to delineate whether phages may contribute to dysbiosis and disease or, on the contrary, help to maintain microbiota stability by preserving bacterial diversity.

Further studies are also needed to substantiate possible direct interactions of phages with immune cells and to define whether and how such direct effects may modify the composition of the bacterial microbiota and influence host health or disease. Defining the mechanisms that determine the outcome of phage–bacteria interactions in the gut is particularly instrumental in the perspective of phage therapy, which, to cite Brüssow,¹²⁵ “is without doubt an interesting approach to the antibiotic resistance problem and merits intensified research to get out of the fruitless confrontation between enthusiasm from the East and lingering Western skepticism”.

AUTHOR CONTRIBUTIONS

All authors participated to the writing of the review.

ADDITIONAL INFORMATION

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.

REFERENCES

1. Nobrega, F. L. et al. Targeting mechanisms of tailed bacteriophages. *Nat. Rev. Microbiol.* **16**, 760–773 (2018).
2. Dy, R. L., Richter, C., Salmond, G. P. & Fineran, P. C. Remarkable mechanisms in microbes to resist phage infections. *Annu. Rev. Virol.* **1**, 307–331 (2014).
3. Ofir GS, R. Contemporary phage biology: from classic models to new insights. *Cell* **172**, 1260–1270 (2018).
4. Twort, F. W. An investigation on the nature of ultra-microscopic viruses. *Lancet* **2**, 1241–1243 (1915).
5. d'Hérelle, F. On an invisible microbe antagonistic toward dysenteric bacilli. *Compte Rendu de l'Académie des Sci.* **165**, 373–375 (1917).
6. d'Hérelle, F. Studies upon Asiatic cholera. *Yale J. Biol. Med.* **1**, 195–219 (1929).
7. Babalova, E. G. et al. [Preventive value of dried dysentery bacteriophage]. *Zh. Mikrobiol. Epidemiol. Immunobiol.* **45**, 143–145 (1968).
8. Kutter, E. et al. Phage therapy in clinical practice: treatment of human infections. *Curr. Pharm. Biotechnol.* **11**, 69–86 (2010).
9. Weinbauer, M. G. Ecology of prokaryotic viruses. *FEMS Microbiol. Rev.* **28**, 127–181 (2004).
10. Chatterjee, A. & Duerkop, B. A. Beyond bacteria: bacteriophage-eukaryotic host interactions reveal emerging paradigms of health and disease. *Front Microbiol.* **9**, 1394 (2018).
11. Keen, E. C. & Dantas, G. Close encounters of three kinds: bacteriophages, commensal bacteria, and host immunity. *Trends Microbiol.* **26**, 943–954 (2018).
12. Gogokhia, L. et al. Expansion of bacteriophages is linked to aggravated intestinal inflammation and colitis. *Cell Host Microbe* **25**, 285–299 e288 (2019).
13. Aggarwala, V., Liang, G. & Bushman, F. D. Viral communities of the human gut: metagenomic analysis of composition and dynamics. *Mob. DNA* **8**, 12 (2017).
14. Shkoporov, A. N. & Hill, C. Bacteriophages of the human gut: the "known unknown" of the microbiome. *Cell Host Microbe* **25**, 195–209 (2019).
15. De Sordi, L., Lourenco, M. & Debarbieux, L. The battle within: interactions of bacteriophages and bacteria in the gastrointestinal tract. *Cell Host Microbe* **25**, 210–218 (2019).
16. Scanlan, P. D. Bacteria-bacteriophage coevolution in the human gut: implications for microbial diversity and functionality. *Trends Microbiol.* **25**, 614–623 (2017).
17. Van Belleghem J. D., Dabrowska K., Vaneechoutte M., Barr J. J., Bollyky P. L. Interactions between bacteriophage, bacteria, and the mammalian immune system. *Viruses* **11**, pii: E10 (2018).
18. Roux, S. et al. Cryptic inoviruses revealed as pervasive in bacteria and archaea across Earth's biomes. *Nat. Microbiol.* **4**(11), 1895–1906 (2019).
19. Zhang, T. et al. RNA viral community in human feces: prevalence of plant pathogenic viruses. *PLoS Biol.* **4**, e3 (2006).
20. Lim, E. S. et al. Early life dynamics of the human gut virome and bacterial microbiome in infants. *Nat. Med.* **21**, 1228–1234 (2015).
21. Los, M. W. & Wegrzyn, G. Pseudolysogeny. *Adv. Virus Res.* **82**, 339–348 (2012).
22. Manrique, P., Dills, M. & Young, M. J. The human gut phage community and its implications for health and disease. *Viruses* **9**, E141 (2017).
23. Reyes, A. et al. Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* **466**, 334–338 (2010).
24. Manrique, P. et al. Healthy human gut phageome. *Proc. Natl Acad. Sci. USA* **113**, 10400–10405 (2016).
25. Gregory A. C., Zablocki O., Howell A., Bolduc B., Sullivan M. B. The human gut virome database. *Cold Spring Harb. Lab.* <https://doi.org/10.1101/655910> (2019).
26. Low, S. J., Dzunkova, M., Chaumeil, P. A., Parks, D. H. & Hugenholtz, P. Evaluation of a concatenated protein phylogeny for classification of tailed double-stranded DNA viruses belonging to the order Caudovirales. *Nat. Microbiol.* **4**, 1306–1315 (2019).
27. Barylski, J. et al. Analysis of spounaviruses as a case study for the overdue reclassification of tailed phages. *Syst. Biol.* **69**, 110–123 (2019).
28. Minot, S. et al. The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res.* **21**, 1616–1625 (2011).
29. Shkoporov, A. N. et al. The human gut virome is highly diverse, stable, and individual specific. *Cell Host Microbe* **26**, 527–541 e525 (2019).

30. 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).
31. Krupovic, M. & Forterre, P. Microviridae goes temperate: microvirus-related proviruses reside in the genomes of Bacteroidetes. *PLoS ONE* **6**, e19893 (2011).
32. Bobay, L. M., Rocha, E. P. & Touchon, M. The adaptation of temperate bacteriophages to their host genomes. *Mol. Biol. Evol.* **30**, 737–751 (2013).
33. Cornuault, J. K. et al. Phages infecting *Faecalibacterium prausnitzii* belong to novel viral genera that help to decipher intestinal viromes. *Microbiome* **6**, 65 (2018).
34. Lugli, G. A. et al. Prophages of the genus *Bifidobacterium* as modulating agents of the infant gut microbiota. *Environ. Microbiol.* **18**, 2196–2213 (2016).
35. Waller, A. S. et al. Classification and quantification of bacteriophage taxa in human gut metagenomes. *ISME J.* **8**, 1391–1402 (2014).
36. Kim MSB, J. W. Lysogeny is prevalent and widely distributed in the murine gut microbiota. *ISME J.* **12**, 1127–1141 (2018).
37. Cornuault J. K. et al. The enemy from within: a prophage of *Roseburia intestinalis* systematically turns lytic in the mouse gut, driving bacterial adaptation by CRISPR spacer acquisition. *ISME J.* <https://doi.org/10.1038/s41396-019-0566-x> (2019).
38. Oh, J. H. et al. Prophages in *Lactobacillus reuteri* are associated with fitness trade-offs but can increase competitiveness in the gut ecosystem. *Appl. Environ. Microbiol.* **86**, e01922–19 (2019). pii.
39. Edwards, R. A., McNair, K., Faust, K., Raes, J. & Dutilh, B. E. Computational approaches to predict bacteriophage-host relationships. *FEMS Microbiol. Rev.* **40**, 258–272 (2016).
40. 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 Res.* **22**, 1985–1994 (2012).
41. Paez-Espino, D. et al. Uncovering Earth's virome. *Nature* **536**, 425–430 (2016).
42. Galliez, C., Siebert, M., Enault, F., Vincent, J. & Soding, J. Wish: who is the host? Predicting prokaryotic hosts from metagenomic phage contigs. *Bioinformatics* **33**, 3113–3114 (2017).
43. Moreno-Gallego, J. L. et al. Virome diversity correlates with intestinal microbiome diversity in adult monozygotic twins. *Cell Host Microbe* **25**, 261–272 e265 (2019).
44. 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 e656 (2018).
45. Shkoporov, A. N. et al. PhiCrAss001 represents the most abundant bacteriophage family in the human gut and infects *Bacteroides intestinalis*. *Nat. Commun.* **9**, 4781 (2018).
46. Devoto, A. E. et al. Megaphages infect *Prevotella* and variants are widespread in gut microbiomes. *Nat. Microbiol.* **4**, 693–700 (2019).
47. Forterre, P., Soler, N., Krupovic, M., Marguet, E. & Ackermann, H. W. Fake virus particles generated by fluorescence microscopy. *Trends Microbiol.* **21**, 1–5 (2013).
48. Chackerian, B. Virus-like particles: flexible platforms for vaccine development. *Expert Rev. Vaccines* **6**, 381–390 (2007).
49. Cobian Guemes, A. G. et al. Viruses as winners in the game of life. *Annu. Rev. Virol.* **3**, 197–214 (2016).
50. Hoyle, L. et al. Characterization of virus-like particles associated with the human faecal and caecal microbiota. *Res. Microbiol.* **165**, 803–812 (2014).
51. 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).
52. Sender, R., Fuchs, S. & Milo, R. Revised estimates for the number of human and bacteria cells in the body. *PLoS Biol.* **14**, e1002533 (2016).
53. Berg, H. C. & Purcell, E. M. Physics of chemoreception. *Biophys. J.* **20**, 193–219 (1977).
54. Barr, J. J. et al. Bacteriophage adhering to mucus provide a non-host-derived immunity. *Proc. Natl Acad. Sci. USA* **110**, 10771–10776 (2013).
55. Fraser, J. S., Yu, Z., Maxwell, K. L. & Davidson, A. R. Ig-like domains on bacteriophages: a tale of promiscuity and deceit. *J. Mol. Biol.* **359**, 496–507 (2006).
56. Lepage, P. et al. Dysbiosis in inflammatory bowel disease: a role for bacteriophages? *Gut* **57**, 424–425 (2008).
57. Hsu, B. B. et al. Dynamic modulation of the gut microbiota and metabolome by bacteriophages in a mouse model. *Cell Host Microbe* **25**, 803–814 e805 (2019).
58. De Paepe, M. et al. Carriage of lambda latent virus is costly for its bacterial host due to frequent reactivation in monoxenic mouse intestine. *PLoS Genet.* **12**, e1005861 (2016).
59. Weiss, M. et al. In vivo replication of T4 and T7 bacteriophages in germ-free mice colonized with *Escherichia coli*. *Virology* **393**, 16–23 (2009).
60. Zhang, X. et al. Quinolone antibiotics induce Shiga toxin-encoding bacteriophages, toxin production, and death in mice. *J. Infect. Dis.* **181**, 664–670 (2000).

61. Nanda, A. M., Thormann, K. & Frunzke, J. Impact of spontaneous prophage induction on the fitness of bacterial populations and host-microbe interactions. *J. Bacteriol.* **197**, 410–419 (2015).
62. Oh, J. H. et al. Dietary fructose and microbiota-derived short-chain fatty acids promote bacteriophage production in the gut symbiont *Lactobacillus reuteri*. *Cell Host Microbe* **25**, 273–284 e276 (2019).
63. Samuels, A. N., Roggiani, M., Zhu, J., Goulian, M. & Kohl, R. M. The SOS response mediates sustained colonization of the mammalian gut. *Infect. Immun.* **87**, e00711–e00718 (2019).
64. Jubelin G., et al. Modulation of enterohaemorrhagic *Escherichia coli* survival and virulence in the human gastrointestinal tract. *Microorganisms* **6**, pii: E115 (2018).
65. Hernandez, S. B., Cota, I., Ducret, A., Aussel, L. & Casadesu, J. Adaptation and preadaptation of *Salmonella enterica* to bile. *PLoS Genet.* **8**, e1002459 (2012).
66. Diard, M. et al. Inflammation boosts bacteriophage transfer between *Salmonella* spp. *Science* **355**, 1211–1215 (2017).
67. Rossmann, F. S. et al. Phage-mediated dispersal of biofilm and distribution of bacterial virulence genes is induced by quorum sensing. *PLoS Pathog.* **11**, e1004653 (2015).
68. Silpe, J. E. & Bassler, B. L. Phage-encoded LuxR-type receptors responsive to host-produced bacterial quorum-sensing autoinducers. *MBio* **10**, e00638–19 (2019).
69. Taylor, V. L., Fitzpatrick, A. D., Islam, Z. & Maxwell, K. L. The diverse impacts of phage morons on bacterial fitness and virulence. *Adv. Virus Res.* **103**, 1–31 (2018).
70. Wahl, A., Battesti, A. & Ansaldo, M. Prophages in *Salmonella enterica*: a driving force in reshaping the genome and physiology of their bacterial host? *Mol. Microbiol.* **111**, 303–316 (2019).
71. Rolain, J. M., Fancello, L., Desnues, C. & Raoult, D. Bacteriophages as vehicles of the resistome in cystic fibrosis. *J. Antimicrob. Chemother.* **66**, 2444–2447 (2011).
72. 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).
73. Debroas D. & Siguret C. Viruses as key reservoirs of antibiotic resistance genes in the environment. *ISME J.* **13**, 2856–2867 (2019).
74. Calero-Caceres, W., Ye, M. & Bakczar, J. L. Bacteriophages as environmental reservoirs of antibiotic resistance. *Trends Microbiol.* **27**, 570–577 (2019).
75. 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).
76. Enault, F. et al. Phages rarely encode antibiotic resistance genes: a cautionary tale for virome analyses. *ISME J.* **11**, 237–247 (2017).
77. Davies, J. & Davies, D. Origins and evolution of antibiotic resistance. *Microbiol. Mol. Biol. Rev.* **74**, 417–433 (2010).
78. Leigh, B. A. Cooperation among conflict: prophages protect bacteria from phagocytosis. *Cell Host Microbe* **26**, 450–452 (2019).
79. Feiner, R. et al. A new perspective on lysogeny: prophages as active regulatory switches of bacteria. *Nat. Rev. Microbiol.* **13**, 641–650 (2015).
80. Rabinovich, L., Sigal, N., Borovok, I., Nir-Paz, R. & Herskovits, A. A. Prophage excision activates *Listeria* competence genes that promote phagosomal escape and virulence. *Cell* **150**, 792–802 (2012).
81. Bille, E. et al. A virulence-associated filamentous bacteriophage of *Neisseria meningitidis* increases host-cell colonisation. *PLoS Pathog.* **13**, e1006495 (2017).
82. Obeng, N., Pratama, A. A. & Elsas, J. D. V. The significance of mutualistic phages for bacterial ecology and evolution. *Trends Microbiol.* **24**, 440–449 (2016).
83. Duerkop, B. A., Clements, C. V., Rollins, D., Rodrigues, J. L. & Hooper, L. V. A composite bacteriophage alters colonization by an intestinal commensal bacterium. *Proc. Natl Acad. Sci. USA* **109**, 17621–17626 (2012).
84. Frazao, N., Sousa, A., Lassig, M. & Gordo, I. Horizontal gene transfer overrides mutation in *Escherichia coli* colonizing the mammalian gut. *Proc. Natl Acad. Sci. USA* **116**, 17906–17915 (2019).
85. Sekse, C., Solheim, H., Urdahl, A. M. & Wasteson, Y. Is lack of susceptible recipients in the intestinal environment the limiting factor for transduction of Shiga toxin-encoding phages? *J. Appl. Microbiol.* **105**, 1114–1120 (2008).
86. Zhao, S. et al. Adaptive evolution within gut microbiomes of healthy people. *Cell Host Microbe* **25**, 656–667 e658 (2019).
87. Chibani-Chennoufi, S. et al. In vitro and in vivo bacteriolytic activities of *Escherichia coli* phages: implications for phage therapy. *Antimicrob. Agents Chemother.* **48**, 2558–2569 (2004).
88. Maura, D. & Debarbieux, L. On the interactions between virulent bacteriophages and bacteria in the gut. *Bacteriophage* **2**, 229–233 (2012).
89. De Sordi, L., Khanna, V. & Debarbieux, L. The gut microbiota facilitates drifts in the genetic diversity and infectivity of bacterial viruses. *Cell Host Microbe* **22**, 801–808 e803 (2017).
90. Gabig, M. et al. The cell surface protein Ag43 facilitates phage infection of *Escherichia coli* in the presence of bile salts and carbohydrates. *Microbiology* **148** (Pt 5), 1533–1542 (2002).
91. Maura, D., Galtier, M., Le Bouguenec, C. & Debarbieux, L. Virulent bacteriophages can target O104:H4 enteroaggregative *Escherichia coli* in the mouse intestine. *Antimicrob. Agents Chemother.* **56**, 6235–6242 (2012).
92. Lourenco, M., De Sordi, L. & Debarbieux, L. The diversity of bacterial lifestyles hampers bacteriophage tenacity. *Viruses* **10**, E327 (2018).
93. Brussow, H. Bacteriophage-host interaction: from splendid isolation into a messy reality. *Curr. Opin. Microbiol.* **16**, 500–506 (2013).
94. Seed, K. D. et al. Evolutionary consequences of intra-patient phage predation on microbial populations. *Elife* **3**, e03497 (2014).
95. 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).
96. Miedzybrodzki, R. et al. Clinical aspects of phage therapy. *Adv. Virus Res.* **83**, 73–121 (2012).
97. Sarker, S. A. & Brussow, H. From bench to bed and back again: phage therapy of childhood *Escherichia coli* diarrhea. *Ann. N. Y. Acad. Sci.* **1372**, 42–52 (2016).
98. Sarker, S. A. et al. Oral phage therapy of acute bacterial diarrhea with two coliphage preparations: a randomized trial in children from Bangladesh. *Ebio-Medicine* **4**, 124–137 (2016).
99. Ott, S. J. et al. Efficacy of sterile fecal filtrate transfer for treating patients with *Clostridium difficile* infection. *Gastroenterology* **152**, 799–811 e797 (2017).
100. Jahn, M. T. et al. A phage protein aids bacterial symbionts in eukaryote immune evasion. *Cell Host Microbe* **26**, 542–550 e545 (2019).
101. Tetz, G. V. et al. Bacteriophages as potential new mammalian pathogens. *Sci. Rep.* **7**, 7043 (2017).
102. Galtier, M. et al. Bacteriophages to reduce gut carriage of antibiotic resistant uropathogens with low impact on microbiota composition. *Environ. Microbiol.* **18**, 2237–2245 (2016).
103. Dissanayake, U., Ukhanova, M., Moya, Z. D., Sulakvelidze, A. & Mai, V. Bacteriophages reduce pathogenic *Escherichia coli* counts in mice without distorting gut microbiota. *Front. Microbiol.* **10**, 1984 (2019).
104. Febvre, H. P. et al. PHAGE study: effects of supplemental bacteriophage intake on inflammation and gut microbiota in healthy adults. *Nutrients* **11**, E666 (2019).
105. Sarker, S. A. et al. Oral application of *Escherichia coli* bacteriophage: safety tests in healthy and diarrheal children from Bangladesh. *Environ. Microbiol.* **19**, 237–250 (2017).
106. Chan, B. K. et al. Phage selection restores antibiotic sensitivity in MDR *Pseudomonas aeruginosa*. *Sci. Rep.* **6**, 26717 (2016).
107. Chatterjee, A. et al. Bacteriophage resistance alters antibiotic-mediated intestinal expansion of *Enterococci*. *Infect. Immun.* **87**, e00085–19 (2019).
108. Lossouarn, J. et al. *Enterococcus faecalis* countermeasures defeat a virulent picovirinae bacteriophage. *Viruses* **11**, E48 (2019).
109. Rigottier-Gois, L. et al. The surface rhamnopolysaccharide epa of *Enterococcus faecalis* is a key determinant of intestinal colonization. *J. Infect. Dis.* **211**, 62–71 (2015).
110. Kortright, K. E., Chan, B. K., Koff, J. L. & Turner, P. E. Phage therapy: a renewed approach to combat antibiotic-resistant bacteria. *Cell Host Microbe* **25**, 219–232 (2019).
111. Golomidova, A., Kulikov, E., Isaeva, A., Manykin, A. & Letarov, A. The diversity of coliphages and coliforms in horse feces reveals a complex pattern of ecological interactions. *Appl Environ. Microbiol.* **73**, 5975–5981 (2007).
112. Rodriguez-Valera, F. et al. Explaining microbial population genomics through phage predation. *Nat. Rev. Microbiol.* **7**, 828–836 (2009).
113. Le Chatelier, E. et al. Richness of human gut microbiome correlates with metabolic markers. *Nature* **500**, 541–546 (2013).
114. Torres-Barcelo, C. The disparate effects of bacteriophages on antibiotic-resistant bacteria. *Emerg. Microbes Infect.* **7**, 168 (2018).
115. Norman, J. M. et al. Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell* **160**, 447–460 (2015).
116. Majewska, J. et al. Oral application of T4 phage induces weak antibody production in the gut and in the blood. *Viruses* **7**, 4783–4799 (2015).
117. Nguyen S., et al. Bacteriophage transcytosis provides a mechanism to cross epithelial cell layers. *MBio* **2017** **8**, e01874–17 (2017).
118. Barfoot, R. et al. Some properties of dendritic macrophages from peripheral lymph. *Immunology* **68**, 233–239 (1989).
119. Wenger, S. L., Turner, J. H. & Petricciani, J. C. The cytogenetic, proliferative and viability effects of four bacteriophages on human lymphocytes. *Vitro* **14**, 543–549 (1978).
120. Aronow, R., Danon, D., Shahar, A. & Aronson, M. Electron microscopy of in vitro endocytosis of T2 phage by cells from rabbit peritoneal exudate. *J. Exp. Med.* **120**, 943–954 (1964).
121. Mabbott, N. A., Donaldson, D. S., Ohno, H., Williams, I. R. & Mahajan, A. Microfold (M) cells: important immunosurveillance posts in the intestinal epithelium. *Mucosal Immunol.* **6**, 666–677 (2013).

122. Yang, J. Y. et al. Enteric viruses ameliorate gut inflammation via Toll-like receptor 3 and Toll-like receptor 7-mediated interferon-beta production. *Immunity* **44**, 889–900 (2016).
123. Van Bellegem, J. D., Clement, F., Merabishvili, M., Lavigne, R. & Vaneechoutte, M. Pro- and anti-inflammatory responses of peripheral blood mononuclear cells induced by *Staphylococcus aureus* and *Pseudomonas aeruginosa* phages. *Sci. Rep.* **7**, 8004 (2017).
124. Sweere, J. M. et al. Bacteriophage trigger antiviral immunity and prevent clearance of bacterial infection. *Science* **363**, eaat9691 (2019).
125. Brussow, H. Hurdles for phage therapy to become a reality—an editorial comment. *Viruses* **11**, E557 (2019).
126. Zheng, Q., Chen, Q., Xu, Y., Suttle, C. A. & Jiao, N. A virus infecting marine photoheterotrophic Alphaproteobacteria (*Citromicrobium* spp.) defines a new lineage of ssDNA viruses. *Front. Microbiol.* **9**, 1418 (2018).



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) 2020

Manuscrit de l'article 2

Hundreds of viral families in the healthy infant gut

Au cours de ma thèse, j'ai eu la chance de modestement collaborer dans le cadre d'un projet international visant entre autres à résoudre le problème de la « matière noire » présente dans la plus grande étude sur le virome humain publiée à ce jour.

Il est admis que le microbiote intestinal est façonné dès la petite enfance et joue un rôle majeur dans la détermination de la susceptibilité aux maladies inflammatoires chroniques au cours de la vie. Il est également connu que les phages modulent les populations bactériennes dans de nombreux écosystèmes, dont l'intestin. Cependant, les données du virome sont difficiles à analyser car elles sont principalement constituées de virus inconnus, plus communément dénommés « matière noire virale ». Cet article présente les travaux réalisés afin de résoudre ce problème dans les viromes fécaux d'une cohorte de 647 enfants âgés d'un an.

Ces viromes ont été séquencés à une profondeur de 10 millions de reads par échantillon, analysés et organisés en matrice de distances puis en arbre, sur la base de leur fraction de gènes partagés. Ainsi, plus de 10 000 espèces virales, réparties en 248 familles virales et appartenant à 17 ordres viraux ont été découvertes. La plupart des familles et des ordres viraux définis sont nouveaux et appartiennent à la classe virale des *Caudoviricetes*. Des hôtes bactériens ont été prédits pour 79% des espèces virales. Comme attendu, diverses familles de Crassphages, phages apparentés au phage intestinal Φ CrAss001, prévalents chez les adultes et infectant majoritairement *Bacteroides*, ont été identifiées. De plus, de nouvelles familles virales, plus prédominantes que les Crassphages, ont été découvertes, en particulier des phages infectant les *Clostridiales* et les *Bifidobacterium*. Ma contribution à ce projet s'inscrit dans une optique de quantification des phages au sein de ces viromes et plus particulièrement des VLPs dans les fèces. Sur 18 échantillons analysés, la concentration médiane était de $3,31 \times 10^8$ VLPs/g, soit 24 fois inférieure à celle des adultes sains ($7,87 \times 10^9$ VLPs/g) étudiés pendant ma thèse (**Tableau V**). Cet écart peut s'expliquer par la différence d'âge au sein des populations étudiées, à savoir des enfants d'un an dans le cadre de ce projet, contre des adultes dans le cadre de ma thèse.

Les styles de vie des phages ont été déterminés pour plus de trois mille espèces de caudophages et se sont révélés être homogènes au niveau des familles pour 149 familles de *Caudoviricetes* : 32 de ces familles comportaient uniquement des phages virulents, et 117 des phages tempérés. En abondance relative, les phages des familles de phages virulents étaient les plus abondants mais ceux des familles de phages tempérés étaient les plus divers et les plus répandus. Ensemble, les familles virales découvertes dans cette étude représentent une expansion majeure de la taxonomie existante des phages.

Échantillon	VLPs/g
4_4	3,10E+08
11_5	2,08E+08
25_10	1,76E+08
33_6	3,20E+08
29_4	3,76E+08
37_2	3,41E+08
30_3	1,40E+08
14_1	1,56E+09
24_7	2,72E+08
19_8	1,15E+09
21_14	1,69E+09
24_6	1,23E+09
4_2	2,54E+08
15_4	2,17E+08
29_18	6,50E+08
32_8	8,87E+08
34_9	2,19E+08
38_16	5,27E+08
Médiane	3,31E+08
Moyenne	8,04E+08
SD	4,98E+08
Valeur min	1,40E+08
Valeur max	1,69E+09

Tableau V. Concentrations en VLPs/g des échantillons analysés

Les concentrations en VLPs/g ont été déterminées par microscopie par épifluorescence.

Hundreds of viral families in the healthy infant gut

Shiraz A. Shah^{1*}, Ling Deng^{2*}, Jonathan Thorsen^{1,3}, Anders G. Pedersen⁴, Moïra B. Dion^{5,6}, Josué L. Castro-Mejía², Ronalds Silins², Fie O. Romme², Romain Sausset⁷, Leon E. Jessen¹, Eric Olo Ndela⁸, Mathis Hjemlsø¹, Morten A. Rasmussen^{1,2}, Tamsin A. Redgwell¹, Cristina Leal-Rodríguez¹, Gisle Vestergaard⁴, Yichang Zhang², Bo Chawes¹, Klaus Bønnelykke¹, Søren J. Sørensen⁹, Hans Bisgaard¹, Francois Enault⁸, Jakob Stokholm^{1,2}, Sylvain Moineau^{5,6,10}, Marie-Agnès Petit⁷, Dennis S. Nielsen²

1. Copenhagen Prospective Studies on Asthma in Childhood, Copenhagen University Hospital, Herlev-Gentofte, Ledreborg Allé 34, DK-2820 Gentofte, Denmark
2. Department of Food Science, University of Copenhagen, Denmark
3. Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen
4. Department of Health Technology, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark.
5. Département de biochimie, de microbiologie, et de bio-informatique, Faculté des sciences et de génie, Université Laval, Québec City, QC, Canada
6. Groupe de recherche en écologie buccale, Faculté de médecine dentaire, Université Laval, Québec City, QC, Canada
7. Université Paris-Saclay, INRAE, Agroparistech, Micalis institute, , Jouy en Josas, France
8. Lab de Microorganismes : Génome et Environnement, Université Clermont Auvergne, 49, Bd François-Mitterrand - CS 60032, France
9. Department of Biology, University of Copenhagen, Copenhagen, Denmark
10. Félix d'Hérelle Reference Center for Bacterial Viruses, Université Laval, Québec City, QC, Canada

* : These authors contributed equally

Abstract

The gut microbiome is shaped through infancy and plays a major role in determining susceptibility to chronic inflammatory diseases later in life. Bacteriophages (phages) are known to modulate bacterial populations in numerous ecosystems, including the gut. However, virome data can be difficult to analyse because it often consists of unknown viruses. Here, we exhaustively resolved the uncharted viral diversity in a cohort of 647 infants at one year of age. Faecal viromes were deeply sequenced and characterized through successive rounds of clustering and curation. We uncovered more than ten thousand viral species distributed over 248 viral family-level clades (VFCs) falling within 17 viral order-level clusters (VOCs). Of the defined VFCs, 232 were novel and belonged to the *Caudoviricetes* viral class. Bacterial hosts were predicted for 79% of the viral species using CRISPR spacers, including those in metagenomes from the same faecal samples. While *Bacteroides*-infecting crAssphages were present, novel viral families were more predominant, including phages infecting *Clostridiales* and *Bifidobacterium*. Phage lifestyles were also determined for more than three thousand caudoviral species. Lifestyles were homogeneous at the family level for 151 *Caudoviricetes* families, with 33 families found to be virulent, while 118 were temperate.

Virulent phages were more abundant, while temperate ones were more diverse and widespread in the infant gut. Together, the viral families found in this study represent a major expansion of existing bacteriophage taxonomy, enabling clinical translational viromics studies in the infant gut.

Introduction

The establishment of the gut microbiome (GM) during the first years of life plays a pivotal role in the maturation of the infant immune system^{1,2}. Early-life GM dysbiosis has been linked to a series of chronic diseases occurring later in life, indicative of a lasting effect on immune programming³⁻⁶. GM imbalances are also linked to flares of inflammatory bowel disease⁷, rheumatoid arthritis and systemic lupus erythematosus⁸. Most existing research has focused on the bacterial component of the GM, but in recent years, it has become evident that viruses are prominent GM members. Recent studies have shown how the transfer of gut viral content from healthy donors can cure recurrent *Clostridoides difficile* infections⁹, alleviate diet induced obesity¹⁰ and even prevent necrotizing enterocolitis in preterm neonates¹¹. The mechanisms are still unclear, but are likely to involve modulation of GM composition through viral infection. While some gut viruses infect human cells, indeed most are bacteriophages (phages), *i.e.* viruses that infect bacteria¹².

Phages colonise the gut during the first months of life following a patterned, host specific, trajectory resembling the establishment of gut bacteria¹³⁻¹⁶. Virulent phages undergo the lytic cycle in which they readily multiply and kill their host cell through lysis and release new virions into the ecosystem. On the other hand, temperate phages have a second form of subsistence. Besides performing lytic cycles, they can integrate their genome into the bacterial chromosome, thereby becoming prophages. This prophage status postpones the killing of the host until particular environmental conditions are experienced, during which a prophage is induced to enter the lytic cycle. Some phages can also cause chronic infections leading to continuous shedding of viral particles¹⁷. Bacteria will defend themselves against these viruses using an extensive arsenal of defence systems¹⁸, which include among others, CRISPR-Cas systems, an adaptive immune mechanism where DNA records (spacers) of past viral infections are saved on a chromosomal CRISPR array to help combat future phage attacks¹⁹.

In addition to phages being able to alter GM composition and function^{10,14}, there are also reported interactions between phage components and the human host immune response²⁰⁻²², suggesting a tripartite interaction that may modulate host health. The first report on the viral metagenome (virome) composition in the infant gut dates back more than a decade²³, and the infant virome has recently been shown to be influenced by caesarean section²⁴. Nevertheless large-scale studies establishing the early-life virome composition and structure are sparse, and human virome studies in general have been challenged by the large proportion of uncharted viral diversity, which is sometimes referred to as the viral “dark matter” problem²⁵.

The latter refers to the phenomenon where only a small fraction of nucleic acid sequences in a virome can be traced back to any known virus. Attempts at *de novo* virus discovery have been limited by the lack of universal viral marker genes, which makes it difficult to distinguish viral sequences from contaminating DNA. In addition, *de novo* classification of novel viruses has been held back by the lack of criteria and standardised methods. However, important progress has been made in recent years²⁶⁻²⁸, and the developments have enabled the compiling of large human gut virus databases²⁹⁻³¹. Such databases are a major step forward but still early in their development. Currently they lack

viral taxonomies and there is still much viral diversity left to be discovered. In the meantime, uncharted viral diversity prevents researchers from fully harnessing their virome data, perhaps missing important biology as a result. We posit that comprehensive annotation of viromics data, including new viral taxonomies, trees and gene families is required for informative and powerful analyses against sample metadata, eventually paving the way for interventions to modulate the bacterial populations of specific niches for preventive or treatment purposes.

Traditionally, the definition of new viral taxa has required the laboratory isolation of the virus along with its host and its subsequent characterization³². However, over the past few years the International Committee for the Taxonomy of Viruses (ICTV) has opened up the possibility for defining viral taxa based on sequence information alone. This significant change is already having major implications as several new taxa are being introduced and proposed, particularly for tailed phages (or caudoviruses), which are highly diverse³³. As an example, the ICTV established the complete taxonomy of the new *Herelleviridae* caudoviral family, demonstrating the proper definition of viral families, subfamilies and genera according to this new paradigm³⁴. Subsequently, three new caudoviral families were identified in human gut metagenome data³⁵. Recently, the prominent human gut phage family *Crassviridae*³⁶ was elevated into a viral order *Crassvirales*³⁷, belonging to the viral class *Caudoviricetes* which itself is proposed now to encompass caudoviruses³⁸ in general.

Here, we present the characterization of the fecal virome at one year of age from 647 infants enrolled in the Copenhagen Prospective Studies on Asthma in Childhood 2010 (COPSAC2010) cohort³⁹. *De novo* assembly and annotation of the uncharted viral diversity enabled the identification of over ten thousand viral species, more than half of which appear to be completely sequenced. Hierarchical clustering of these viruses based on encoded protein similarity enabled the *de novo* definition of novel viral genera, subfamilies and virus family-level clades (VFCs). In total, we identified 248 viral VFCs that fell within 17 tentative virus order-level clusters (VOCs). Most of the novel VFCs belonged to *Caudoviricetes*, representing a major expansion to current caudoviral taxonomy. We also predicted the hosts for 79% of the viral species by matching spacer sequences from bacterial CRISPRs and found that members of the *Bacteroides*-infecting *Crassvirales* - otherwise abundant in adult gut viromes - were outnumbered by novel phage clades infecting numerous gut bacteria such as *Clostridiales* and *Bifidobacterium*. Temperate phages were the most widespread and diverse, while virulent phages were more abundant.

Results

Study population

COPSAC2010 is a population-based mother-child cohort study of 700 Danish children originating almost equally from rural, suburban and urban regions in and around the greater Copenhagen area. Participants were recruited in pregnancy with the aim of prospectively studying the causes for chronic inflammatory diseases³⁹. A total of 647 children had a faecal sample obtained at one year of age where the virome was characterised here. Of note, metagenomes for the same samples have been sequenced⁴⁰ previously.

Identifying the viruses and resolving their taxonomies

Virome extractions are known to contain various amounts of bacterial contaminating DNA⁴¹ and the uncharted viral diversity²⁵ makes it difficult to discern novel viruses from contaminants. We elected to resolve this issue by assembly, clustering and successive rounds of manual curation (see Methods). The manual route was chosen in order to avoid potential selection biases (see Additional Results and Table S1 for details) in existing tools and criteria such as “circular contigs”³⁵, which could have prevented the identification of truly novel viral clades.

In short, the extracted viromes from 647 infants were sequenced to an average of 3 Gbp per sample. After assembly and species-level deduplication, resulting OTUs were clustered by protein content (Figure S1), visualised (Figure S2) and then manually curated. In the end 10,021 manually confirmed viral OTUs (vOTUs) made up the final set of viral species (see Additional Results and Methods for details) found in this study. These vOTUs recruited roughly half of the sequencing reads from all samples, with the remaining half mapping mainly to OTUs of bacterial contaminating DNA (see Additional Results and figures S3, S4 and S5), which is comparable to other studies⁴². The contaminant OTUs were not analysed further.

In order to determine which vOTUs were parts of existing viral families, we pooled them with 7.7K species-level deduplicated reference phages⁴³. After gene calling, protein alignments were used for defining viral ortholog gene clusters (VOGs) *de novo* and for constructing an aggregate protein similarity (APS) tree. The tree was rooted and cut at the levels reproducing the recent taxonomy for the *Herelleviridae*³⁴ phage family, thus yielding clusters corresponding to viral families, subfamilies and genera covering all vOTUs and reference phages. An additional order-level cutoff was based on the newly proposed caudoviral *Crassvirales* order³⁷.

The 10,021 species-level vOTUs fell within 248 curated virus family-level clades (VFCs), including 16 known families (Figure 1) harbouring 2497 vOTUs and 232 novel VFCs containing 7524 vOTUs. The novel VFCs were named after the infants that delivered the faecal samples. All VFCs were additionally grouped into 17 VOCs (Table S2), five of which were already known (Figure 1). After estimating the typical complete genome size at the family-level (Figure 1), 56 % of the vOTUs were found to be complete or near-complete. vOTU DNA sequences and taxonomies along with visualisations of the VFCs (Figure S2) have been made available online via the interactive version of Figure 1 at <http://copsac.com/earlyvir/f1y/fig1.svg>

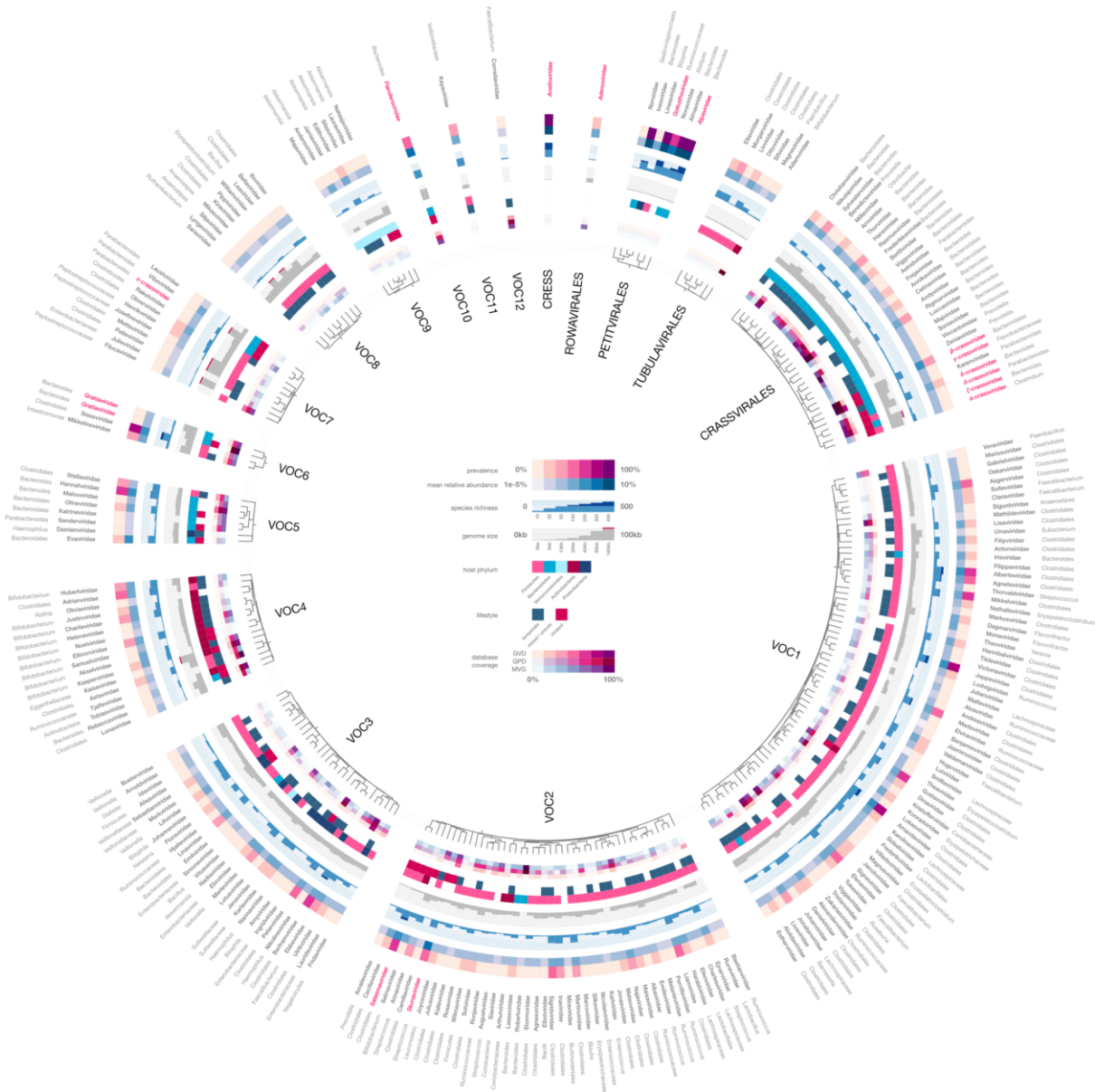


Figure 1: An atlas of infant gut DNA virus diversity. Faecal viromes from 647 infants at age one year were deeply sequenced, assembled and curated, resulting in the identification of 10,021 viral species falling within 248 viral family-level clades (VFCs). Predicted host ranges for each VFC are given, and the VFCs have been grouped into 17 virus order-level clusters (VOCs). Trees show how families are interrelated within each VOC, and heatmaps and histograms encode their lifestyle, host range, abundance and prevalence across the cohort, their genome sizes and species diversities as well as their prevalence in published gut virus databases. For the 16 previously known viral families, names are written in red. An interactive version of the figure with expandable families can be accessed online, for browsing the gene contents and downloading the genome of each virus: <http://copsac.com/earlyvir/f1y/fig1.svg>.

Infant gut vOTUs are largely absent from gut virus databases

We checked whether any of our 10,021 curated viral species were already found within three recently published gut virus databases that were built mainly on adult faecal metagenomic data. GVD³¹ contained 819 of our vOTUs, while the larger and more recent GPD³⁰ and MG²⁹ databases covered 2,307 and 2,171 vOTUs, respectively. Combined, 7,046 (70%) of the infant gut vOTUs identified here were not found in any of the three gut virus databases. At the family level, however, most of the 248 VFCs had some representatives in either database, with *Crassvirales* VFCs being particularly well represented in both GPD and MG. Importantly, the majority of our most species-rich VFCs (e.g. candidate family “Amandaviridae”) were poorly represented in all three databases, while the VFCs best covered by the databases were often minor in our data (Figure 1). In other words, most large gut phage clades in databases are only occasionally found in our infant gut viromes, and *vice versa*. This pattern suggests that the infant gut is a unique niche harbouring specialised viruses distinct from the adult gut ecosystem.

Major novel viral families dominate the infant gut virome

As already mentioned, cutting the APS tree at the family-³⁴ and order-level³⁷ yielded 248 VFCs and 17 VOCs. The family-level cutoff reproduced the expected crAssphage families recently defined³⁷ (Figure 1). The order-level cutoff reproduced the five known viral orders in our data (i.e. *Petitvirales*, *Tubulavirales*, anelloviruses (yet unclassified CRESS viruses), *Rowavirales* and *Crassvirales*) along with 12 additional VOCs all of which were caudoviral (Table S2). At the family-level, 232 out of 248 VFCs were caudoviral, further underlining the diversity of this viral class. The mean and median VFC size was 40 and 17 species-level vOTUs respectively, making the typical VFC similar in richness to currently known gut phage families such as *Flandersviridae*³⁵.

In order to identify the most predominant viral clades, three measures were calculated, namely, total species richness, prevalence across samples, and mean relative abundance (MRA) (Figure 2). Family and order-level MRA and prevalence estimates were determined by first mapping sample reads to vOTUs, then aggregating their counts based on taxonomic affiliation. All three measures were highly correlated (figures 2 and S6), meaning that the most diverse VFCs and VOCs were also the most widespread and abundant. The correlation between these measures is predicted by the neutral community model which is also known to accurately explain bacterial community structures^{44,45}.

In our data, human-infecting ssDNA anelloviruses (*Anelloviridae*) and bacterial ssDNA microviruses (*Petitvirales*) were amongst the most abundant viral clades (Figure 2A). These abundant ssDNA viral families (see next subsection for details) were followed by ten major dsDNA VFCs belonging to the *Caudoviricetes* viral class (Figure 2B). Four of these are known caudoviral families pending ICTV approval, namely *Skunaviridae*, *Salasmaviridae*, β -*crassviridae* and *Flandersviridae*, while the remaining six comprise novel candidate families. Importantly, crAss-like viruses, which are abundant in adult faecal viromes⁴⁶, were only a minor part of the infant gut virome when compared to some of the novel clades found here (Figure 2A).

Skunaviridae is a family of virulent phages infecting *Lactococcus* dairy cultures⁴⁷ that may have originated from the diet but still ended up as the caudoviral family with the highest abundance in our data (2.7% mean relative abundance, MRA). *Salasmaviridae* is small viral family harbouring around a dozen *Bacillus* podoviral species including the well known phage phi29. With the infant gut vOTUs we were able to broaden the scope of the *Salasmaviridae* family with more than two hundred diverse

gut phages spanning more than twenty viral subfamilies, infecting a wide variety of important gut *Firmicutes* and *Actinobacteria*. *β-crassviridae*, a minor crAss-like viral family in adults, was found in almost a third of the infants (n = 210 out of 647), infecting both *Bacteroides* and *Clostridiales* hosts. The major adult crAssphage variant, *α-crassviridae*^{37,48} however, was only present in 5% of the infants. *Flandersviridae* is a *Bacteroides*-infecting phage family recently defined based on 30 complete phage genomes³⁵ from public metagenome assemblies. Found in almost half of the children (n = 286), here we markedly expand this family with a set of 80 complete species-level vOTUs spanning four subfamilies.

Apart from the four known and virulent viral families above, six novel candidate families were additionally found to be highly abundant, prevalent and diverse, harbouring hundreds of viral species most of which were temperate. The prevalence and richness estimates for these candidate families indicate that they are at least as predominant in the infant gut ecosystem as crAssphage in adults⁴⁶. Candidate family “Sisseviridae” is almost universally present in the infants, at 80% prevalence. Apart from harbouring the highly prevalent *Faecalibacterium* phage Oengus⁴⁹, this candidate family encompasses a wide range of both temperate and virulent vOTUs infecting diverse *Firmicutes* and *Actinobacteria*. The temperate candidate families “Amandaviridae”, “Jeppeviridae” and “Alberteviridae” are all related by belonging to the major virus order-level cluster, VOC1. These candidate families were present in 70% of the infants, containing between 200 and 300 viral species each and identified to infect key *Clostridiales* gut bacteria such as *Ruminococcus*, *Blautia*, *Anaerostipes* and *Hungatella*. Apart from a few unclassified *Clostridium* and *Brevibacillus* reference phage species that co-cluster within them, these expansive gut phage clades are largely unexplored. Finally, “Evaviridae” and “Hannahviridae” comprise two related candidate families of *Bacteroides*-infecting phages containing around 200 species in total. The former appears strictly virulent while the latter harbours both virulent and temperate subfamilies. “Hannahviridae” includes the recently described *Bacteroides* phage “Hankyphage”⁵⁰ known for its diversity-generating retroelements, and it has been extensively described in a parallel provirome study performed on the same samples⁵¹.

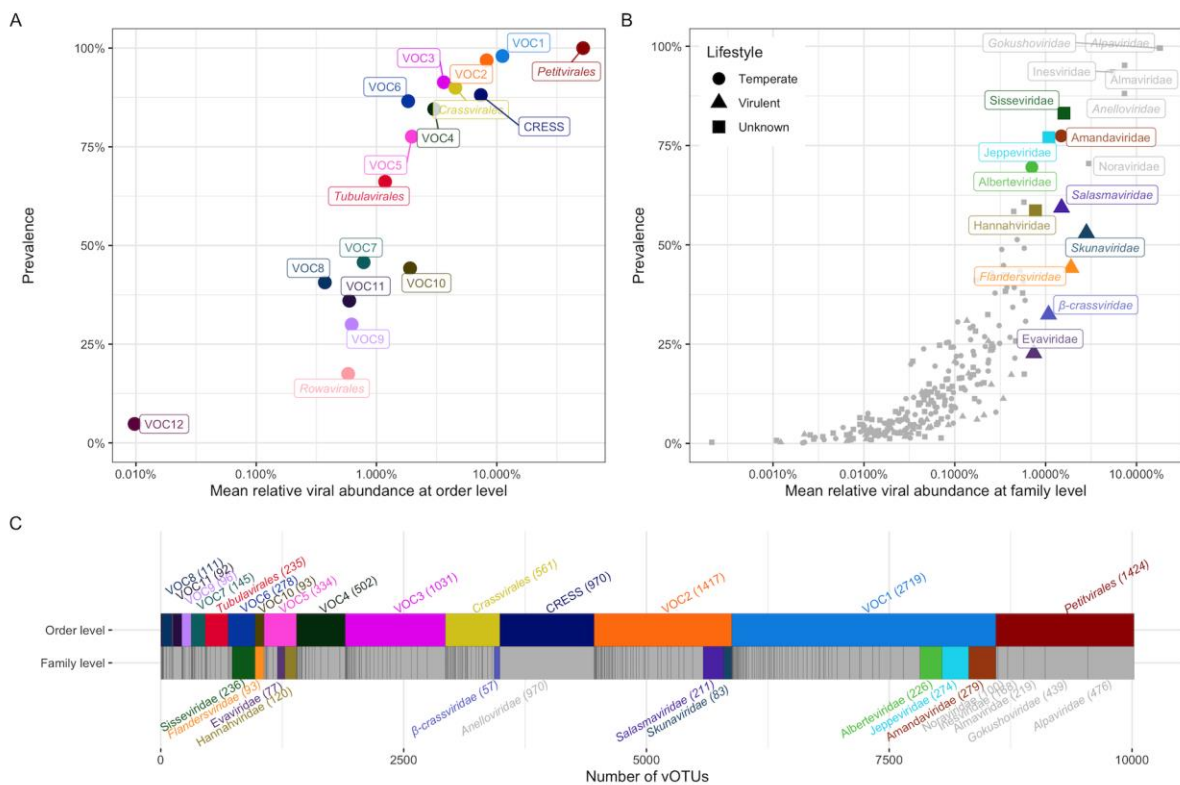


Figure 2: Abundance, prevalence and richness of the novel families in the 1-year old infant gut. Already-known viral clades are indicated in italics. A) Prevalence and mean relative abundance (MRA) of the 17 VOCs across samples. B) Prevalence and MRA of the 248 VFCs. The major VFCs were defined as the ten most

abundant caudoviral VFCs in the data, and are colored and labelled. Minor VFCs as well as ssDNA families are in grey. Predicted lifestyles for the 10 major VFCs are indicated by different shapes. C) VOCs and VFCs scaled by species richness, ordered by MRA. VOC12 and *Rowavirales* are not shown due to their small sizes. The VFCs are represented underneath the VOCs they belong to.

Clades of ssDNA viruses in the infant gut

ssDNA vOTUs collectively recruited around a third of the sequencing reads, but after normalising for their short genome sizes, they accounted for 60% of the MRA (Figure S3). Although the short multiple-displacement amplification (sMDA) protocol we used to detect the ssDNA viruses could have inflated their counts, both their prevalence and richness was in line with their high abundance (see also Additional Results and Figure S9). The ssDNA VFCs displayed canonical positioning along the neutral community model (figures 2B and 3F) indicating that any artificial inflation would have been limited. ssDNA viruses in our data fell within three separate viral classes, *Malgrandeviricetes*, CRESS viruses, and *Faserviricetes*, each harboring a single viral order.

Microviruses of the *Petitvirales* viral order (class *Malgrandeviricetes*) are ubiquitous small icosahedral ssDNA phages and were the most prevalent and abundant group of viruses found in our viromes, making up 52% of the MRA. Their importance was supported by the fact that 21% of all CRISPR spacer matches from the metagenome targeted microviruses (data not shown). vOTUs from the two major families *Gokushoviridae* and *Alpaviridae* (currently known as subfamilies *Gokushovirinae* and *Alpavirinae*) in our data are predicted to infect *Clostridiales* and *Bacteroidales* respectively, but other minor VFCs were also detected (Figure 1).

Anelloviruses from the CRESS class of ssDNA viruses are also known as Torque Teno viruses, and comprise a single family (*Anelloviridae*) of small 3-kb ssDNA viruses that infect animal cells. They are known to cause chronic asymptomatic infections in healthy humans, with elevated titers in immunocompromised patients⁵². The immature immunity of the infants may explain why *Anelloviridae* were abundant in our samples, making up 7% of the MRA and comprising by far the richest single family with 970 species-level vOTUs. On average, each infant harboured 10 species of *Anelloviridae*, and this is consistent with earlier findings¹⁵. Unsurprisingly, no CRISPR spacer matches were found targeting any *Anelloviridae* vOTUs.

Inoviruses from the *Tubulavirales* order (class *Faserviricetes*) are a ubiquitous and diverse group of filamentous phages with small ssDNA genomes⁵³. Some of them can integrate into their host genomes using encoded integrases while others cause chronic non-lethal infections which result in the continuous shedding of new viral particles¹⁷. Although they were diverse in our data, distributed among seven distinct families, like the *Petitvirales*, their species richness was much lower at 235 vOTUs, and abundances were correspondingly lower totalling 1% MRA. Most of the inoviral families found were predicted to infect *Clostridiales*, although members of the VFC “Adamviridae”, appear to specifically infect *Bifidobacterium* (Figure 1).

Virus lifestyle determines abundance and prevalence patterns

While examining the ten major caudoviral VFCs we noticed that half of them lacked any encoded integrase, otherwise commonly found throughout the less abundant VFCs in the data. Since an integrase is an indicator of a temperate lifestyle, we investigated systematically whether a virulent lifestyle was linked to higher abundances overall. First, the typical complete genome size per VFC was successfully determined for 228 VFCs by examining the size distribution of their constituent

vOTUs. The median [interquartile range (IQR)] complete genome size for the VFCs was 35kb [30 - 50 kb]. Using the determined minimum complete size limit per viral family (Figure 1), 5,608 vOTUs with complete and near-complete genomes were screened for the presence of integrases (see methods). Phage lifestyles were mostly homogeneous at the family-level and a total of 118 VFCs were deemed temperate, while only 33 were found to be virulent. The remaining 97 VFCs exhibited either a mixed lifestyle pattern or were uncertain due to an insufficient number of complete genomes.

On its own, family-level abundance was not significantly linked to phage lifestyle (two-sided Wilcoxon test, $p=0.90$; Figure 3A), but temperate VFCs were significantly more prevalent than virulent VFCs ($p=0.048$; Figure 3B). Temperate phages have been shown to be more genetically diverse than their virulent counterparts⁵⁴, so we tested this by comparing the amount of unique branch length (as a fraction of total branch length) in virulent vs. temperate family-level APS subtrees. Indeed, temperate caudoviral VFCs were more genetically diverse by this measure ($p < 2.2 \cdot 10^{-16}$; Figure 3C), compared to virulent VFCs. As for the predicted bacterial hosts, *Clostridiales* were particularly enriched in temperate VFCs, whereas most virulent VFCs were predicted to infect *Bacteroidales* (Figure 1). Using the CRISPR spacer mappings we found, in line with other studies^{30,55}, that some vOTUs appeared to infect multiple host species, genera or even families of bacteria. We decided to check whether the CRISPR-Cas system targeted virulent phages more often than temperate phages, or whether a virulent lifestyle was associated with a broader host range. This was not the case as both temperate and virulent families exhibited similar mean host ranges and numbers of targeting spacers ($p=0.2$ and $p=0.097$, respectively; figures 3D-E).

Finally, plotting the abundance and prevalence of the virulent and temperate VFCs against each other (Figure 3F) resurfaced our initial suspicion that virulent VFCs were indeed present in higher amounts despite being found in fewer children. Thus, we tested this hypothesis systematically by using the neutral community model (Figure 3G), which describes the typical relationship between abundance and prevalence, as the baseline assumption⁵⁶. After fitting the model on all of our VFC abundances, virulent VFCs had significantly lower residuals against it than temperate VFCs (two-sided Wilcoxon test, $p=2.1 \cdot 10^{-05}$; Figure 3H), confirming that they were both less prevalent and more abundant than temperate VFCs.

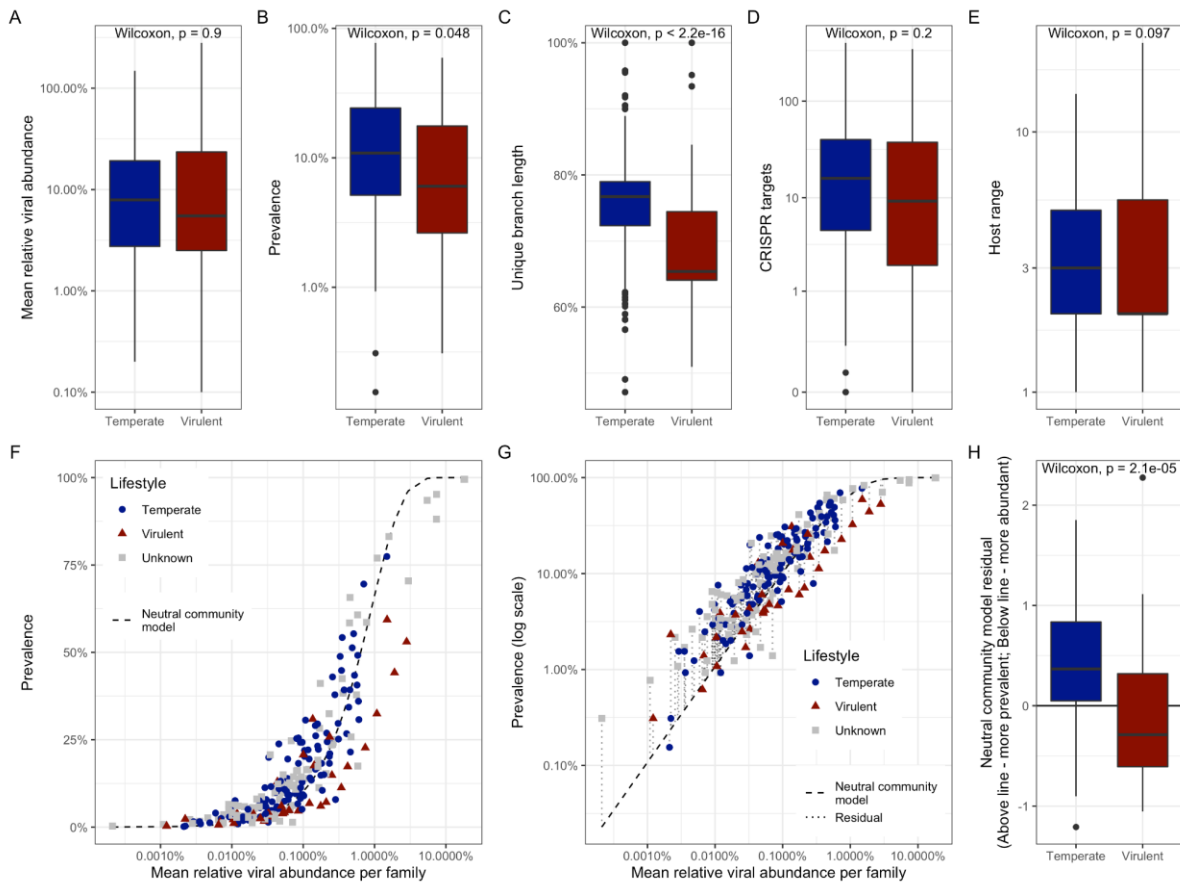


Figure 3: Characteristics of temperate vs. virulent VFCs in the data in terms of A) Mean relative abundance (MRA); B) Prevalence; C) Genetic diversity as measured by unique branch length D) Number of metagenomic CRISPR spacer matches; E) Host range (number of host species); F) Fit of the neutral community model on the VFCs from Figure 2B, G) Deriving neutral community model residuals from the log-transformed prevalences; and H) Comparison of neutral community model residuals, showing that temperate VFCs tend to have positive residuals, whereas virulent VFCs tend towards negative residuals, indicating that temperate phages are present in lower abundance despite being found in more children, as compared to virulent phages. For panels A to E and H, $n = 151$ (118 temperate + 33 virulent); Boxplot elements: center line, median; box limits, upper and lower quartiles; whiskers, 1.5x IQR; points, outliers. Two-sided Wilcoxon test p-values reported. For panels F and G, $n = 248$ (118 temperate + 33 virulent + 97 unknown).

Phage-host abundances are linked regardless of virus lifestyle

Bacterial hosts for the vOTUs were predicted using 318K CRISPR spacers from our metagenome assembled microbial genomes (MAGs)⁴⁰ and 11M spacers from the CRISPR spacer database⁵⁷ as well as by using WISH⁵⁸. The three host predictions were merged by selecting the last common ancestor (LCA). Host predictions at the bacterial genus-level were made for 63% of the vOTUs, while 77% of them were covered at the order-level (Figure 4A) and 79% at the phylum level. *Bacteroides* was by far the most commonly predicted bacterial host genus followed by *Faecalibacterium* and *Bifidobacterium*. At the order-level, approximately half of the annotated vOTUs had *Clostridiales* as hosts, with *Bacteroidales* covering just one quarter (Figure 4A). Interestingly, these differences mirror the corresponding pattern for the bacterial taxa in the metagenomes, where *Bacteroides* was the most abundant genus, while *Clostridiales* were more diverse (Figure 4B).

In order to determine whether the abundance of phages infecting a particular host was associated with the abundance of the host itself, we correlated phages and hosts at both the sample-level and study-wide. First individual vOTU relative abundances were aggregated by predicted host genus.

These were then compared to bacterial genus abundances in metagenomes for the corresponding samples using Spearman tests (Figure S7). Correlations between corresponding phage-host genus pairs were enriched for positive values across all genera (Figure S7A, $p = 2.4 \cdot 10^{-12}$), indicating positive co-abundance for phages and their hosts between children. Positive associations between phage-host pairs were expected for temperate phages actively induced by their hosts, and since such phages were more common in our data (4800 temperate vs. 1006 virulent vOTUs), they could be masking any negative signal from virulent phages killing their hosts. After splitting up the vOTU data by predicted phage lifestyle and recomputing correlations between pairs of phage-host genera, both lifestyles still yielded strong positive correlations (Figure S7C).

Next, study-wide mean relative abundances (MRAs) were calculated by averaging the relative abundances for each vOTU across all samples. These were also aggregated at the host-genus level and compared with MRAs for bacterial genera in the metagenomes, yielding a strong positive correlation (Spearman's $\rho = 0.76$, $p < 2.2 \cdot 10^{-16}$, Figure 4C). A similar result was obtained even after splitting the data into temperate and virulent phages (Figure S8).

The strong phage-host correlations support the accuracy of both the host predictions and viral abundance estimations. Overall, in the infant gut, both virulent and temperate phages correlate positively with the abundance of their hosts. This applies both across children and between hosts. In conclusion, although virulent phages kill their hosts, they still act as markers for their presence, at least when seen through the lens of a single viromics snapshot lacking the temporal resolution to capture predator-prey fluctuations.

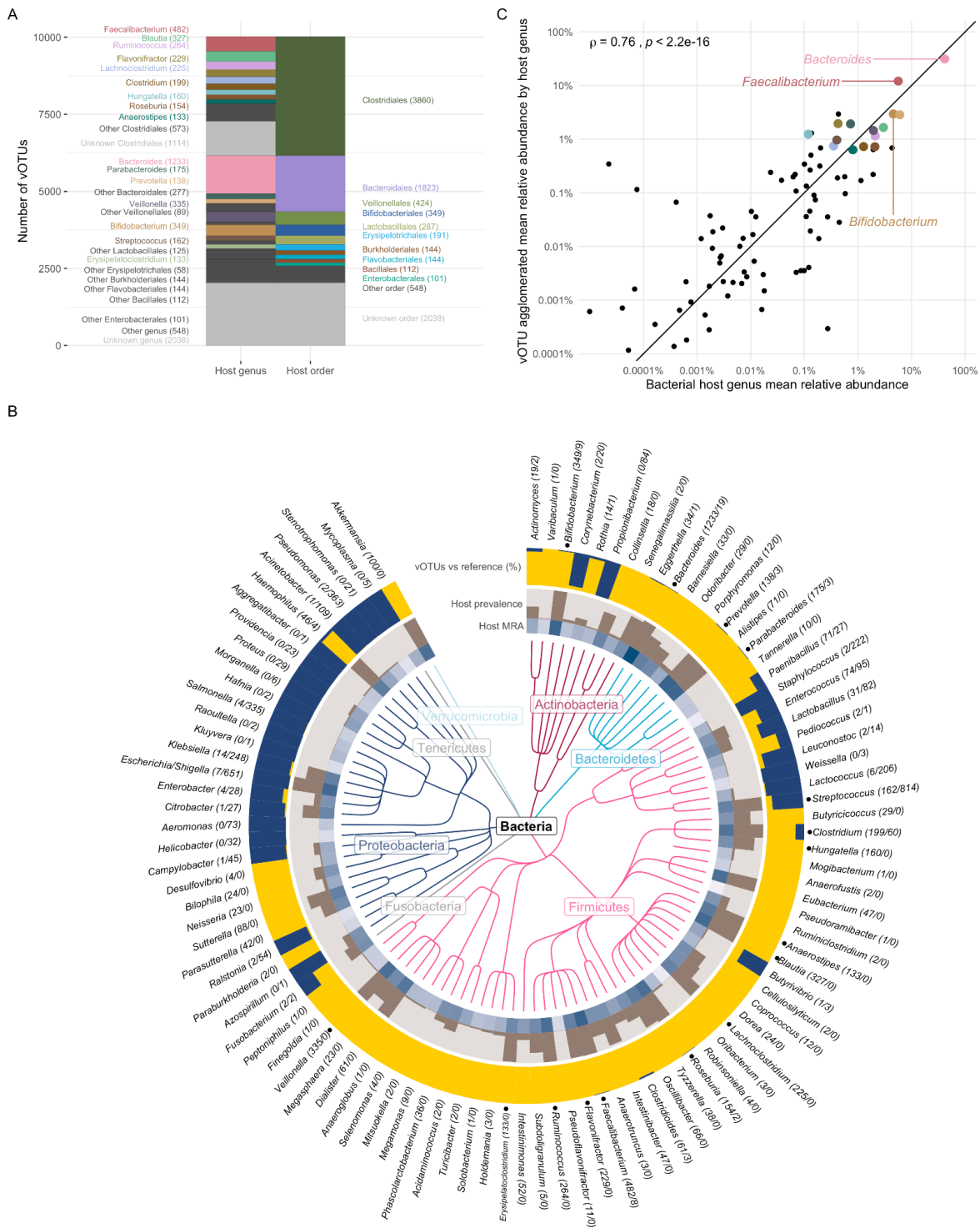


Figure 4: Prediction of bacterial hosts for the 10,021 vOTUs found in the infant gut virome shows that *Bacteroides*, *Faecalibacterium*, and *Bifidobacterium* are the three most prominent host genera. A) Distribution of virus host predictions collapsed to bacterial order and genus levels, respectively. Numbers in parentheses denote the number of vOTUs with a given host genus or order. B) The top 100 gut bacterial genera found in gut metagenomes from the same infant faecal samples, as represented by a taxonomic tree. The mean relative abundance (MRA) of each bacterial genus is shown in the blue heatmap, while the fraction of the 647 infants harbouring the host genus (i.e. its prevalence) is shown with the brown barplot. For each host bacterium, the yellow bars show the proportion of viral species found in this study relative to those in the reference phage set⁹¹ shown in dark blue. Numbers behind each genus name denote the total number of vOTU vs. reference phage species per bacterial host genus. The 16 major host genera from panel A are indicated by a dot in front of their names in panel B. C) Each dot represents a genus from panel B, by its MRA in the metagenome

against the aggregate MRA of all its vOTUs in the virome. Host abundances correlated with corresponding phage abundances as tested by a Spearman's rank test.

Discussion

Until recently, the gut virome was understudied, but it is becoming increasingly clear that it is an important player in GM community interactions. The recent publication of several large and curated gut virus databases have illustrated the massive diversity of this community^{29–31}. Yet, as we show here, significant parts of this ecological niche are still uncharacterised. A thorough description of the gut viruses is essential if one aims at modulating the GM with novel treatments or efforts involving early-life prevention of later chronic disease.

In this study we described the DNA virome in faecal samples from 647 one-year-old infants, and the study represents three major advances in descriptive viromics. Our attempt at exhaustively resolving the uncharted viral diversity within a virome data set left only 7% of the virome sequences unaccounted (Figure S3). This result shows, for DNA viromes, that what has traditionally been termed “virome dark matter” consists essentially of either novel phages or bacterial contaminating DNA. Secondly, we performed a comprehensive taxonomic assignment covering the entire data set. This led to the identification of 232 novel viral family-level clades, representing a major expansion of current phage taxonomy, along with a large number of complete or near-complete representative genomes. Lastly, we generated the first ecosystem-wide resolution of phage lifestyles. This analysis showed that the bulk of phage diversity in the 1 year old infant gut is composed of temperate phages, even if less diverse virulent phages can be more abundant. This latter finding echoes what has recently been shown for the neonate gut¹⁴, where phages appeared to be almost entirely temperate, and it is in strong contrast to the healthy adult gut where virulent phages appear to dominate³¹.

In addition to the six major candidate families described in the results, numerous additional predominant caudoviral VFCs can be browsed online (Figure 1). As a general rule, we found that *Bacteroides*-infecting VFCs were more often virulent and host specific, while VFCs infecting *Clostridiales* featured much wider host ranges and were overwhelmingly temperate. Multiple VFCs were often specialised for a single host genus, e.g. the 27 *Bacteroides*-specific VFCs, or the seven *Akkermansia*-specific VFCs found (Figure 1). *Clostridiales*-specific VFCs, on the other hand, were more agnostic in their host ranges featuring multiple host genera (browse “Amandaviridae” for examples) or even single vOTUs predicted to infect hosts belonging to multiple families within the same order. Such features underscore the rapid rate at which caudoviruses are able to evolve both horizontally across hosts, but also vertically within tight host niches, as in *Bacteroides* and *Akkermansia*. Since e.g. *Akkermansia* is phylogenetically distinct from other gut bacterial genera⁵⁹, its phages likely face greater barriers for host switching, making its phage families appear more host-specific in a human gut context. Overall we found caudoviral richness to exceed host richness by an order of magnitude, both at the species and genus levels (e.g. 2,858 caudoviral genera vs. 203 host genera in the metagenome). In fact, in our data, host genus-level richness was more comparable to caudoviral family-level richness, at 232 VFCs.

In most virome studies carried out to date, the extracted DNA is amplified by MDA prior to sequencing, a technique that has been claimed to bias sequencing composition towards ssDNA viruses^{60,61} in addition to compromising quantitative analyses overall. However, the largest meta-analysis of virome studies published³¹, did not find any significant differences between non-MDA and standard 2-hour MDA gut viromes. Furthermore, in a recent gut virome study using a different DNA library kit enabling ssDNA detection without biases⁶¹, caudoviruses outnumbered microviruses in half of the samples, while for the other half the opposite was true⁶². Here, we used a 30-minute short MDA (sMDA) step to enable ssDNA detection while limiting biases. We found that caudoviruses outnumbered microviruses in around a third of the infants, while the opposite was true for the

remaining two thirds. We also showed strong co-abundances between phages and their hosts. Moreover, we were able to make an unprecedented comparison of plaque forming units to our virome abundances (Figure S9), and found both estimates to be highly correlated. We conclude that the results of this study are relevant in quantitative terms in spite of using sMDA, and that viral abundances are quantitative, at least for dsDNA viruses.

Skunaviridae, our most abundant caudoviral family, comprised only eight complete vOTUs in our dataset. This is atypical considering the hundreds of vOTUs in most of the other abundant viral families we found. All reference phages belonging to the family infect *Lactococcus* while our vOTUs were predicted to infect *Streptococcus*, but this could be an artefact caused by the lack of CRISPR-Cas systems in *Lactococcus*⁶³, preventing this host from being predicted by matching CRISPR spacers. *Streptococcus*, although very prevalent in the children, may not have been abundant enough to support the high counts of virulent *Skunaviridae*. We also did not find any strong correlation between *Skunaviridae* and *Streptococcus* or *Lactococcus* in the data (data not shown). Thus, it remains a possibility that these strictly virulent phages were ingested as part of a diet rich in fermented dairy, as proposed earlier⁶⁴, since *Skunaviridae* phages are naturally present in several dairy products.

In a previous study on *E. coli* phages isolated from the same samples⁶⁵, virulent coliphages were found less frequently than temperate coliphages, but were more abundant and had broader host ranges. Temperate coliphages on the contrary, were frequently isolated, but had limited host ranges. Here, we found a very similar pattern albeit on a much larger scale. Virulent phage families were more abundant but less prevalent than temperate phage families. Although we could not detect a difference in host range, we did find that the temperate phage families were more genetically diverse compared to the virulent ones. The higher prevalence and lower abundance of temperate phages likely reflects frequent prophage induction from lysogenic gut bacteria, as shown in mouse models⁶⁶⁻⁶⁸, and that induced virions do not readily infect new hosts and multiply. In viromics, this would appear as a stable background of temperate phages on top of which virulent blooms would stochastically appear, originating from random phage-host encounters in the gut. For our infant samples this temperate background was intense enough to overshadow the diversity of virulent phages. Possibly, in adult viromes where the GM and host immunity has reached an equilibrium, the bacteria are less stressed in turn making the temperate virome less dominant. This notion is consistent with how a virulent phage core is linked to adult gut health⁶⁹, as well as the paucity of crAssphage in infant viromes³¹.

Conclusion

We deeply sequenced 647 infant gut viromes and tackled the uncharted viral diversity by *de novo* assembly and classification. The approach led to the uncovering of 248 virus family-level clades, 232 of which were novel, and most of which belonged to the *Caudoviricetes* viral class. We also found that temperate phages dominate the infant gut virome at age one year, while crAssphage is a minor player overshadowed by several novel viral clades. Our comprehensive annotation of the infant gut viromes provides a platform for making biologically meaningful statistical analyses against cohort clinical phenotypes for future translational viromics research efforts.

Methods

The study was embedded in the Danish population-based COPSAC2010 prospective mother-child cohort of 736 women and their children followed from week 24 of pregnancy, with the aim of studying the mechanisms underlying chronic inflammatory diseases⁷⁰. A total of 660 participants delivered a faecal sample at age 1 year. Each sample was mixed with 10% vol/vol glycerol broth and stored at -80°C until DNA extraction for metagenomes⁴⁰, and virome extraction. Extraction and sequencing of viromes were done using previously described protocols⁷¹. Briefly, DNA from faecal filtrates enriched in viral particles was extracted and subjected to short (30 minutes) MDA amplification and libraries were prepared following manufacturer's procedures for the Illumina Nextera XT kit (FC-131-1096). For epifluorescence VLP estimations, 10 µL of a virome sample was diluted 100-fold, fixed and deposited on a 0.02 µM filter, dried and stained with SYBRTM-Gold (200X), then visualised with an epifluorescence microscope using a 475 nm laser. VLPs were counted in 8 to 10 fields and multiplied over the remaining filter surface area.

Virome libraries were sequenced on the Illumina HiSeq X platform to an average depth of 3 Gb per sample with paired end 2x150 bp reads. Satisfactory sequencing results were obtained for 647 samples. Virome reads were quality filtered and trimmed using Fastq Quality Trimmer/Filter (options -Q 33 -t 13 -l 32 -p 90 -q 13), and residual Illumina adapters were removed using cutadapt. Trimmed reads were dereplicated using vsearch derep_prefix and then assembled with Spades v3.10.1 using the meta flag while disabling read error correction. Decontamination clusters were generated by reducing redundancy by deduplicating the 1.5M contigs above 1 kb in size into 267k 90% ANI representatives⁷², then calling genes⁷³ and aligning⁷⁴ proteins all-against-all for building an APS tree⁷⁵ (see Supplementary Methods). The tree was cut close to the root to obtain the decontamination clusters. Bacterial MAGs from the same samples⁴⁰ were mined for CRISPR spacers using CRISPRDetect⁷⁶, and the virome decontamination clusters were ranked by their extent of CRISPR targeting multiplied by sample prevalence. The protein alignment results were used to define orthologous gene clusters *de novo*⁷⁷, and gene orthology information was used to visualise the gene contents of contigs within each decontamination cluster. The top 400 ranked clusters were inspected visually for two viral signatures, namely conservation of contig sizes and of gene content. There were diminishing returns beyond the top 400 mark and the remaining decontamination clusters were assumed to represent contaminants.

Species-level (95% ANI) deduplication of contigs into OTUs was done using BLAT⁷⁸. Reference phages were deduplicated to the species-level using the same strategy. Comparisons of the vOTUs to the GVD, GPD and MGV were also performed similarly. Decontaminated vOTUs and reference phage species⁹³ were pooled and the APS tree and gene orthology (VOGs) were recalculated. VOGs were aligned against Pfam⁷⁹, CDD⁸⁰, COG⁸¹ and TIGRFAMs⁸² using HH-suite3⁸³ to gain functional annotations. The APS tree was cut using phylotree-lib (<https://github.com/agormp/phylotree-lib>) to reproduce existing phage taxonomy^{37,84}.

Family visualisations (Figure S2) were used to 1) further curate each individual vOTU in order to separate confirmable viruses that had structural VOGs, from vOTUs representing small fragments or various virus-related MGEs that did not harbour genes coding for any structural proteins. 2) The OTU length distribution within each family was plotted in a histogram with 5-kb steps to locate the right-most size peak. The 5-kb step immediately preceding this peak was set as the lower size bound for a complete or near-complete genome. 3) The family visualisations were inspected to manually remove families that were dominated by reference phages, so as to avoid interference with ongoing classification efforts. Weak families composed mainly of MGEs or fragments, having less than five

vOTUs or less than two complete vOTUs were also removed. For the final version of the family visualisations available online, VOG annotations were redone against PHROGs⁸⁵ because it was more informative.

MAG spacers, along with spacers from CRISPRopenDB⁵⁷ and WisH⁵⁷ were used to generate separate host predictions for each vOTU. The three predictions were integrated using the LCA of the two most closely matching predictions, as an error-correction strategy, since all three methods would occasionally mispredict. Bacterial genus abundances in the metagenome were derived by running mOTUs2⁸⁶ on the reads from each sample followed by aggregating mOTU abundances at the genus-level in phyloseq⁸⁷.

Bacterial contamination was estimated for each virome sample using ViromeQC⁴² along with a custom approach where we leveraged the metagenomes cognate to each virome: Reads were mapped from both fractions against the 16S rRNA gene⁸⁸ and *cpn60*⁸⁹ and the degree of contamination was calculated as the ratio between the two fractions. Abundances of vOTUs in each sample were determined by mapping sample reads to sample contigs using bwa mem -a⁹⁰, then using msamtools profile to determine depth and length-normalised relative abundances with iterative redistribution of ambiguously mapped reads proportionally to uniquely mapped reads (<https://github.com/arumugamlab/msamtools>). The obtained contig abundances were then aggregated at the OTU level in order to obtain vOTU abundances per sample. vOTU abundances were aggregated at the family and order levels using phyloseq⁸⁷ to obtain the statistics used for figures 2 and 3.

A list of VOGs matching to integrase and recombinase protein families was first curated, then used to predict whether complete vOTUs within viral families were temperate or virulent. Families where more than 95% of complete vOTUs did not harbour an integrase were deemed virulent, whereas for temperate families at least 50% of both complete and incomplete vOTUs were required to carry an integrase.

The versions of virus discovery tools used for benchmarking were DeepVirFinder v1.0, VIBRANT v1.2.1, VIRSorter 1.0.6, VIRSorter2 v2.0 commit 22f6a7d, Seeker commit 9ae1488, PPR-Meta v1.1, and CheckV v.0.7.0. The random prediction was created by randomly sampling the 362k OTUs 12,500 times without replacement. The number 12,500 was chosen because it was reasonably close to our own positive set and the number of positives generated by most tools.

Acknowledgments

We express our deepest gratitude to the children and families of the COPSAC2010 cohort study for all their support and commitment. We acknowledge and appreciate the unique efforts of the COPSAC research team.

Funding information

This work is supported by the Joint Programming Initiative 'Healthy Diet for a Healthy Life', specifically here, the Danish Agency for Science and Higher Education, Institut National de la Recherche Agronomique (INRA), and the Canadian Institutes of Health Research (Team grant on Intestinal Microbiomics, Institute of Nutrition, Metabolism, and Diabetes, grant number 143924). M.B.D. is recipient of graduate scholarships from the Fonds de Recherche du Québec - Nature et Technologies [259257] as well as Sentinel North and is a recipient of the Goran-Enhoring Graduate Student Research Award from the Canadian Allergy, Asthma and Immunology Foundation. JT is supported by the BRIDGE Translational Excellence Program (bridge.ku.dk) at the Faculty of Health and Medical Sciences, University of Copenhagen, funded by the Novo Nordisk Foundation (grant no. NNF18SA0034956). S.M. holds the Tier 1 Canada Research Chair in Bacteriophages [950-232136]. S.A.S is a recipient of a Novo Nordisk Foundation project grant in basic bioscience [NNF18OC0052965], along with M.A.R.

Data availability

Viral genome sequences, taxonomy and host predictions for all viruses are available through the online version of Figure 1 on <http://copsac.com/earlyvir/f1y/fig1.svg>. Sequencing FASTQ files can be accessed on ENA with the project number PRJEB46943.

Code availability

Data analyses were carried out using free and open source software as specified in the Supplementary Methods. Custom code was also used and is being made available on GitHub, and can in the meantime be provided upon request.

Availability of unique biological materials

Access to the infant faecal samples used in the study may be granted after signing a collaboration agreement with COPSAC provided that any use is compliant with Danish and EU regulations for protecting the rights of the underage human research participants.

Ethics declarations

Conflicts of interest

The authors declare no conflicts of interest

Ethical compliance

The study was conducted in accordance with the guiding principles of the Declaration of Helsinki and was approved by The National Committee on Health Research Ethics (H-B-2008-093) and the Danish Data Protection Agency (2015-41-3696). Both parents gave written informed consent before enrollment.

Supplementary information

Tables

Table S1: Benchmarking statistics from various metagenomic virus discovery methods against 10,021 vOTUs in the manually curated viral set from a total of 362,668 OTUs. The imbalanced nature of the data set, with non-vOTUs outnumbering vOTUs, inflates the specificity metric, which is 0.965 for a random prediction. The Kappa performance metric was included because it is robust against such imbalances. Prediction methods that provided a confidence score were cut according to authors' recommendations where available, or else by matching the number of positive predictions to the manual set. The performances of VIRSorter and VIBRANT were also checked in alternate modes (bottom portion).

method	predictions	cutoff	sensitivity	specificity	kappa
CheckV	9393	>= medium qual	0.71161	0.99359	0.7273
viralVerify	10921	> 15	0.69873	0.98889	0.6589
VIBRANT	5843	>= medium qual	0.44696	0.99613	0.5556
VIRSorter	14366	cat. 1 + 2	0.69335	0.97896	0.5553
Length + 20kb	13452	> 20kb	0.48548	0.97565	0.3954
VIRSorter2	18638	virus maxscore = 1	0.50983	0.96164	0.3325
DeepVirFinder	10192	q < 0.05	0.058078	0.972749	0.0306
PPR Meta	17222	score > 0.95	0.060772	0.952891	0.0101
random	12500	w/o replacement	0.0330306	0.9654924	-0.0013
Seeker	13520	>= 0.9	0.0263447	0.9624100	-0.0096
VIRSorter virome	30736	cat. 1 + 2	0.83155	0.93647	0.3832
VIRSorter db2	13298	cat. 1 + 2	0.70971	0.98246	0.5973
VIBRANT virome	6181	>= medium qual	0.45395	0.99537	0.5521

Table S2: The 10,021 viral species belonging to the 248 viral families were grouped into viral order-level clades (VOCs). VOCs are sorted here by total richness. Caudoviral mean read abundances (cMRA) are shown for the caudoviral VOCs. For each VOC the number of viral family clades (VFCs), subfamilies, and genera are also given along with their most frequent hosts.

order-level clade	% cMRA	VFCs	subfamilies	genera	vOTUs	major host
VOC1	27	67	531	1,163	2,719	<i>Clostridiales</i>
<i>Petitvirales</i>	-	7	67	154	1,424	<i>Bacteroides</i>
VOC2	19	45	315	618	1,417	<i>Clostridiales</i>
VOC3	9	30	171	358	1,031	<i>Veillonellaceae</i>
anelloviruses	-	1	21	268	970	human
<i>Crassvirales</i>	15	30	114	214	561	<i>Bacteroidales</i>
VOC4	7	17	64	149	502	<i>Bifidobacterium</i>
VOC5	6	8	47	108	334	<i>Bacteroidales</i>
VOC6	5	4	27	75	278	<i>Clostridiales</i>
<i>Tubulavirales</i>	-	7	50	78	235	<i>Clostridiales</i>
VOC7	2	11	37	61	145	<i>Clostridiales</i>
VOC8	1	10	38	61	111	<i>Clostridiales</i>
VOC9	1	7	19	31	96	<i>Akkermansia</i>
VOC10	6	1	4	10	93	<i>Bacteroides</i>
VOC11	1	1	3	4	92	<i>Veillonellaceae</i>
VOC12	0.05	1	3	6	8	<i>Faecalibacterium</i>
<i>Rowavirales</i>	-	1	1	2	5	human

Figures

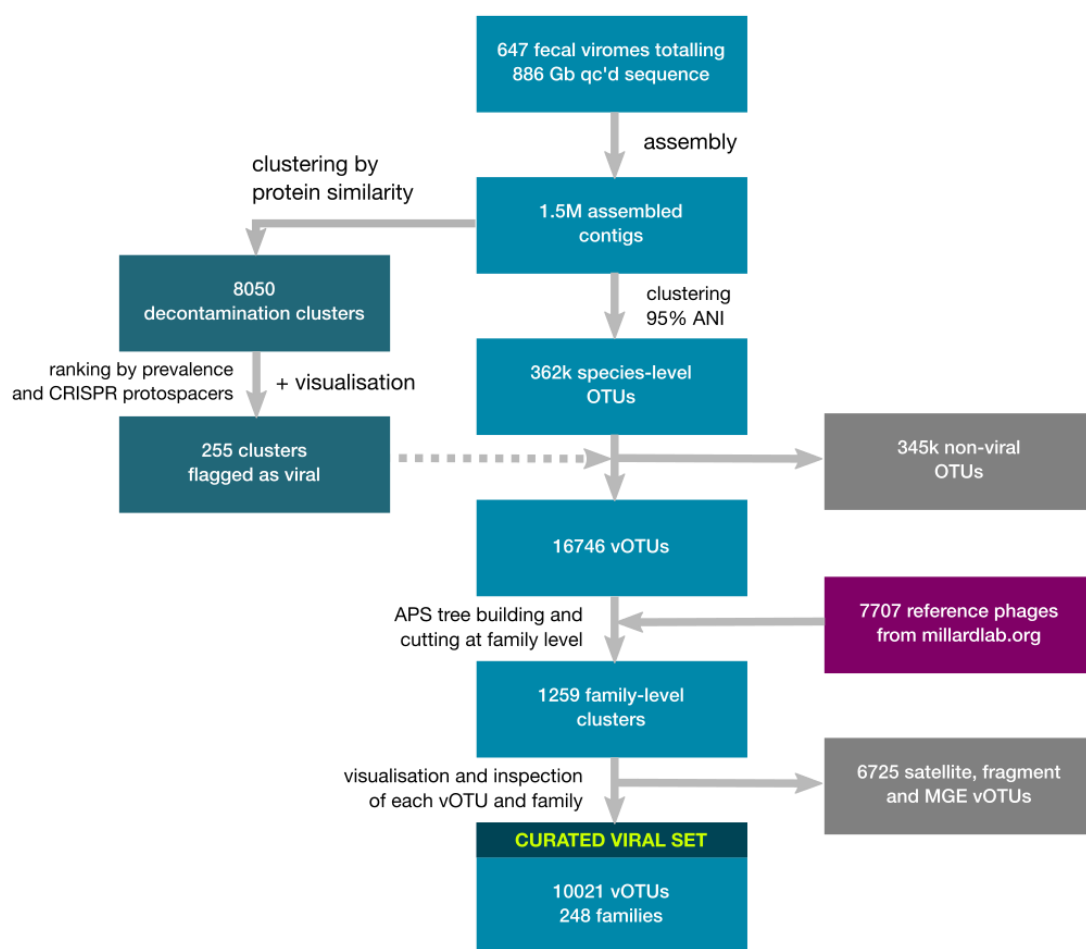


Figure S1: Overview of decontamination and curation procedure.



Figure S2: Clickable gene map of vOTUs belonging to the *Ingridviridae* family available at <http://copsac.com/earlyvir/f1y/families/Ingridviridae.svg> along with similar maps for the remaining 247 families, available via <http://copsac.com/earlyvir/f1y/fig1.svg>. Small vertical gaps between vOTUs denote genus boundaries, while large gaps denote subfamily boundaries. Ordering of the vOTUs follows the order in the APS tree and thus, related vOTUs are next to each other. ORFs are aligned vertically based on strandedness and colored by VOG affiliation. VOG definitions against the PhROGs database⁸⁵ can be looked up by clicking on each ORF. ORF gene product (GP) numbers are displayed by mouse-over hovering. GenBank files for each vOTU can be viewed along with virus and host taxonomy by clicking on the OTU name. Caudoviral maps were inverted and zeroed according to TerL gene coordinates, while the GenBank files were not. Reference phages that belong to the same family were also included in the maps and are indicated by GenBank accession numbers.

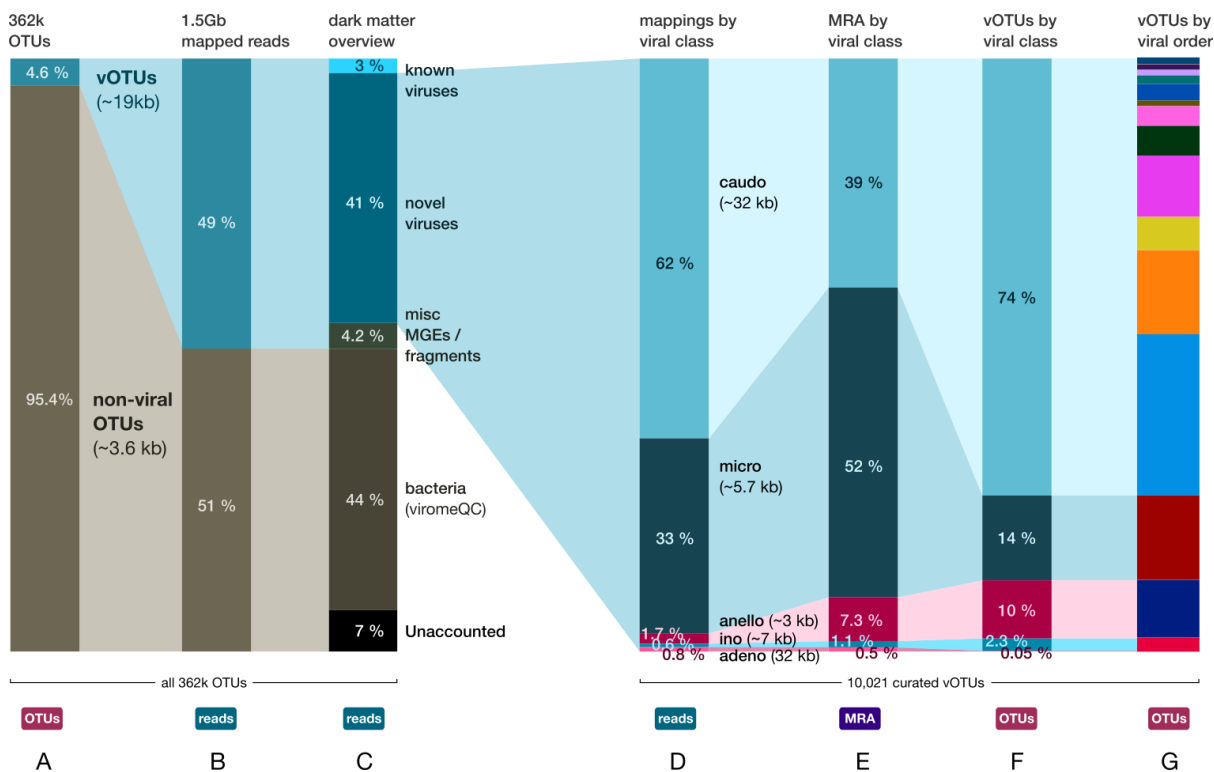


Figure S3: From assembly to curated vOTUs in numbers. After assembly, species-level deduplication and manual decontamination, most OTUs were found to be non-viral and had small sizes while viral OTUs were much fewer but longer (A). After mapping, vOTUs accounted for roughly half of the reads (B). 97% of the reads originally comprised “dark matter” but only 7% was left after resolution (C). The 10,021 curated vOTUs fell within five viral classes (caudoviruses [dsDNA], microviruses [ssDNA], anelloviruses [ssDNA], inoviruses [ssDNA] and adenoviruses [dsDNA]). Distributions of the viral classes by: mapped reads (D), mean relative abundances (MRAs), after normalising read counts for sequencing depth and genome size (E) and species richness, i.e. number of vOTUs (F) are shown. G) Same as F but at viral order-level, with orders colored as in Figure 3.

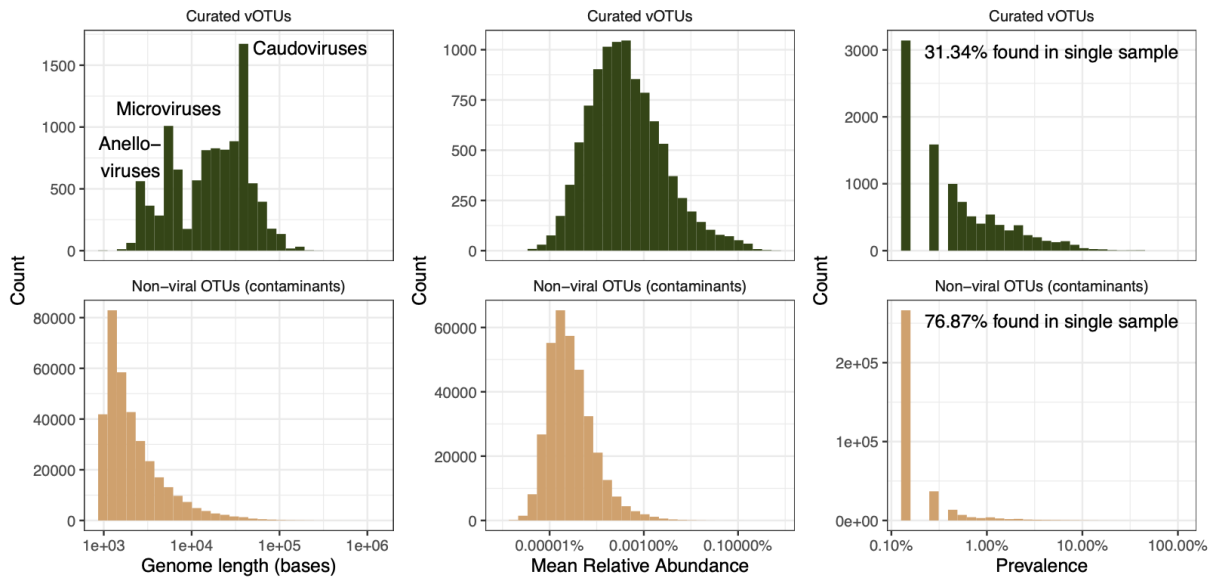


Figure S4: Distribution of size, mean relative abundance (MRA) and sample prevalence for non-vOTUs and curated vOTUs respectively. The vOTU size distribution shows peaks corresponding to genome lengths for the three major classes of viruses in the data set, namely anelloviruses, microviruses and caudoviruses (3 kb, 5.5 kb, and 40 kb). The non-vOTU size distribution peaks at the contig inclusion cutoff (1 kb) continuing with a long uniform tail, consistent with the unspecific origin expected for contaminating DNA. Curated vOTUs were more abundant and prevalent than non-viral OTUs. The majority of non-viral OTUs were sample-specific, in contrast to most curated vOTUs which were found in more than one sample. The latter is consistent with their bacterial chromosomal origin, as unspecific subsampling of the large bacterial genome space is unlikely to yield overlaps between samples.

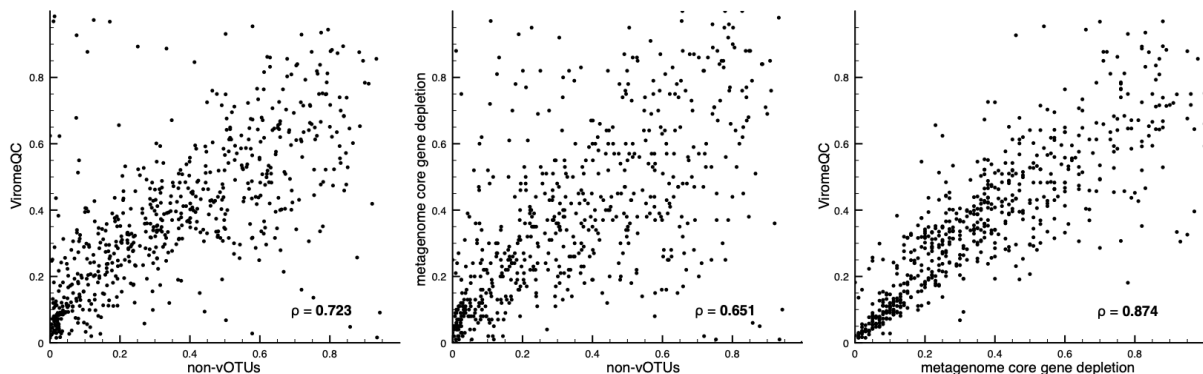


Figure S5: Comparison of three approaches for estimating the proportion of bacterial contamination. Each graph has 647 dots, one for each sample. Axes denote the proportion of bacterial contamination as estimated by the indicated method. Each graph is a pairwise comparison of two different methods. A) non-vOTU mappings vs. ViromeQC B) non-vOTU mappings vs. metagenome core gene depletion C) metagenome core gene depletion vs. ViromeQC. Spearman correlation coefficients (ρ) are given for all three comparisons.

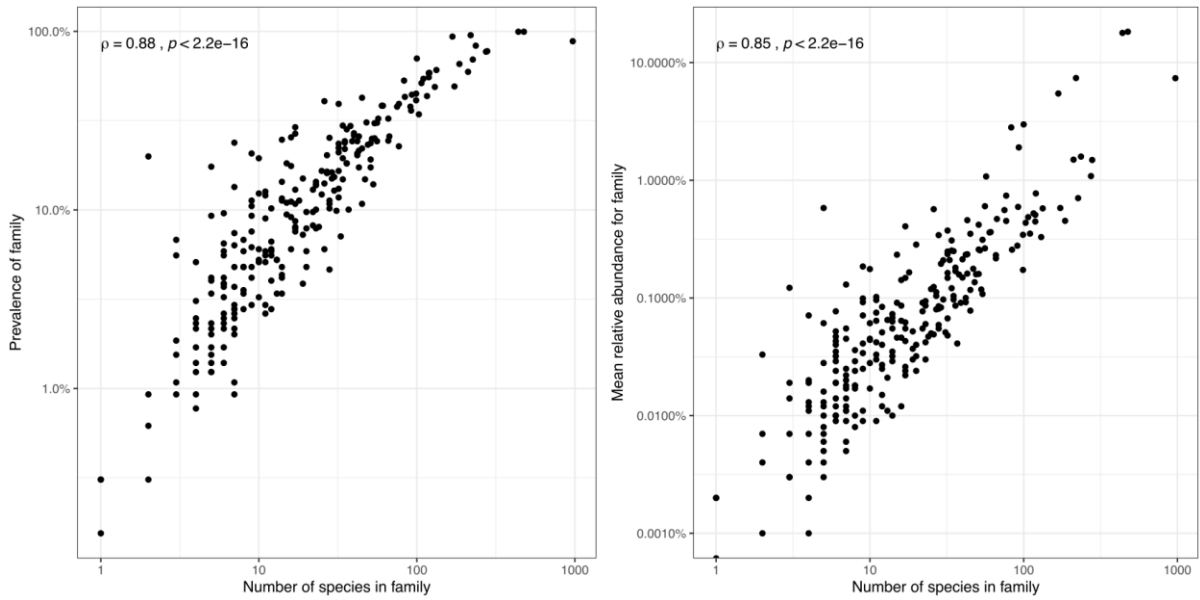


Figure S6: The species-richness within a family is highly correlated with both its prevalence and the MRA across samples, shown here with Spearman correlations. MRAs are correspondingly correlated with prevalence as already shown in Figure 3. The correlation between all three measures is in line with predictions made by the neutral community model. MRA: mean relative abundance.

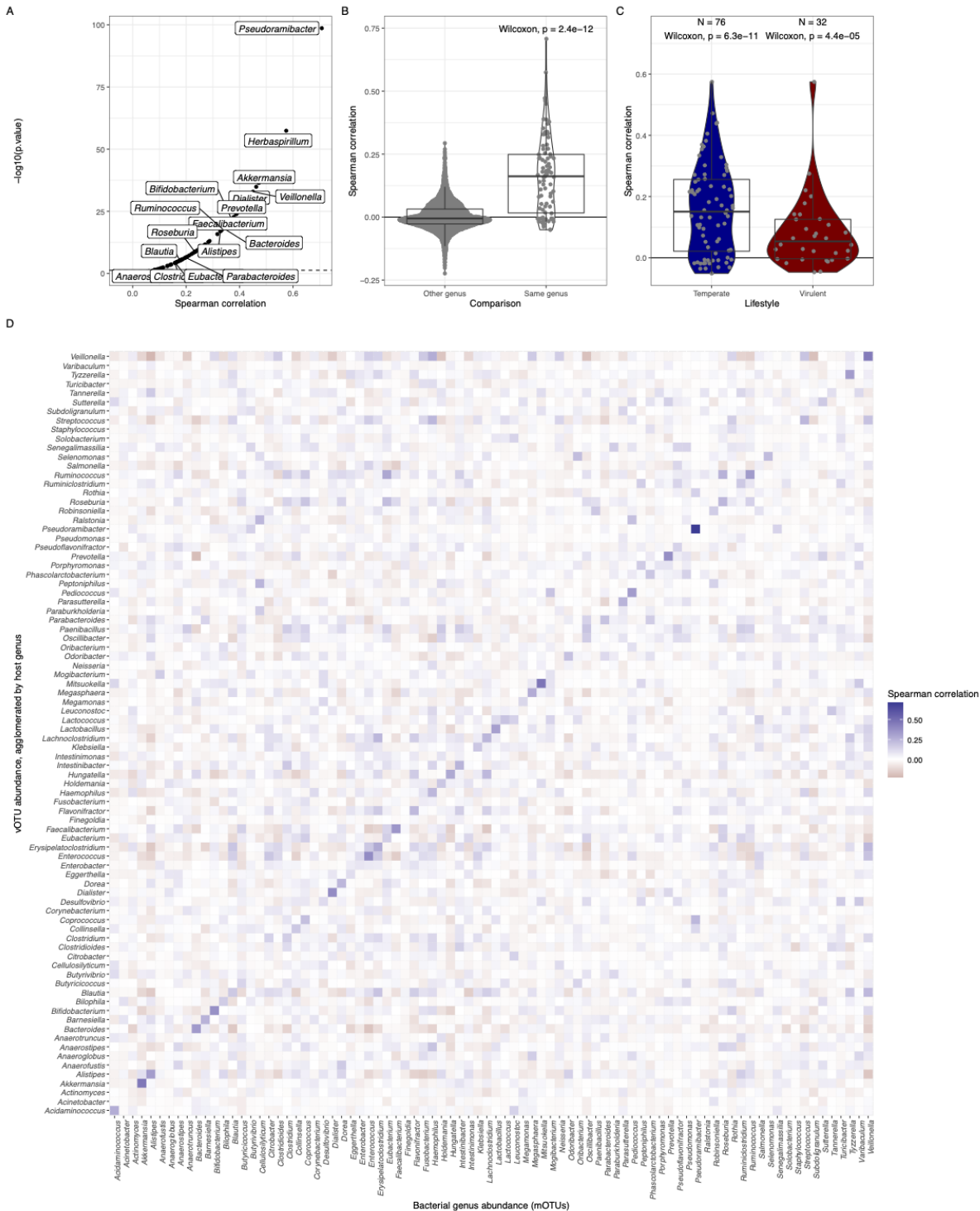


Figure S7: Correlations between host bacterial relative genus abundances in the metagenomes with aggregate relative abundances for phages predicted to infect those host genera in the virome, compared across all children. A) Volcano plot showing how all significant correlations between phage-host pairs were positive ($\rho > 0$), and that the distribution of correlation values was significantly higher than zero (Wilcoxon $p = 2.4 \cdot 10^{-12}$). B) Abundance correlations with the predicted phage-host pairs were stronger than against other genera, C) predictions were positive regardless of phage lifestyle, and D) stood out against the background of all genus combinations tested (same data shown in panel B).

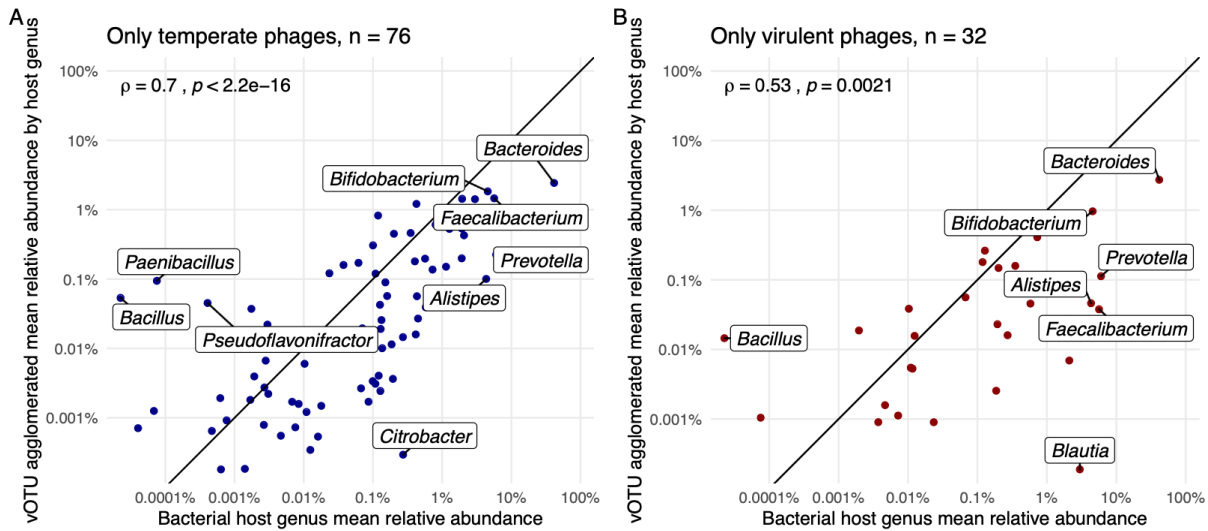


Figure S8: Correspondence between host genus abundances in the metagenome with aggregate abundances for all phages infecting those genera in the virome, as stratified by virus lifestyle. The MRA of both virulent and temperate phages correlates positively with host MRA. MRA: mean relative abundance. Correlations were tested using Spearman's rank test.

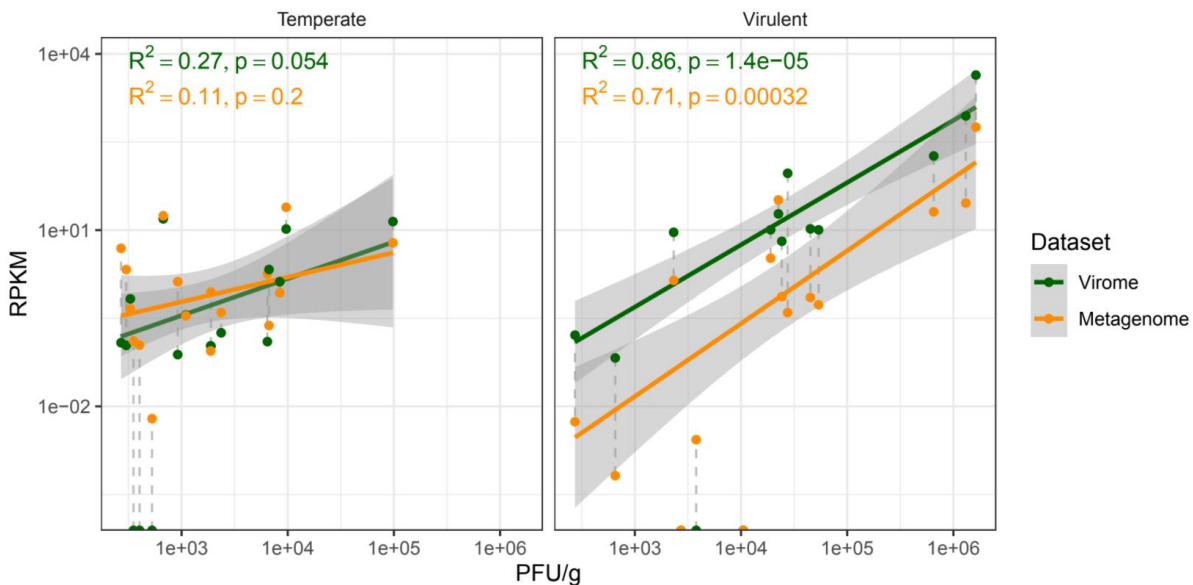


Figure S9: The sMDA amplified viromes were quantitative for dsDNA phages. The relationship between experimentally determined PFU/g of faeces for 32 coliphages⁶⁵, against mapped virome and metagenome reads per kilobase per million (RPKM), from the corresponding 32 samples. The two panels show data for temperate and virulent coliphages respectively. Axes were log-transformed to capture the dynamic range. A linear model was fit following log-transformation. Temperate coliphages show only a tendency of being associated presumably because read-mappings were shared between induced phage DNA and bacterial chromosomal DNA. For the virulent coliphages, however, the relationship was quantitative throughout the range of PFU counts (from 270 to 1.6M). The sMDA amplified virome is no less quantitative than the unamplified metagenomes for the same samples. sMDA: short multiple-displacement amplification.

```

initn: 42 init1: 42 opt: 42 Z-score: 66.6 bits: 21.5 E(2): 0.13
Smith-Waterman score: 60; 24.8% identity (51.3% similar) in 113 aa overlap (215-323:226-331)

```

```

OTU_3060_5      190      200      210      220      230      240
KHWEFNTMLSRLENGKIIIMTRWASDLAGRAIAHY-TOQGVNVRHITMKALIDKE
OTU_3175_18    200      210      220      230      240      250
QRKMLAAPKGTLYKNKILGLRGRATGLVFLDLPNRIISLGTAGQFKFERFDS--AGLDTA

```

```

initn: 2503 init1: 1900 opt: 2506 Z-score: 1933.2 bits: 367.2 E(2): 1.4e-105
Smith-Waterman score: 2506; 76.9% identity (90.4% similar) in 490 aa overlap (3-492:5-491)

```

```

OTU_3060_4      10      20      30      40      50
MKIGERFKSMIQNWLNIIVPAANQSVVLQELLPREIEVLRSLQWYRGDATERLQFFHQI
OTU_3175_17    10      20      30      40      50      60
MKFDIGKVRQKFLWLNINPASEQTFVMEKERTGLMADILIAKMLWYRGDATERLQFFHQI

```



```

250      260      270      280      290      300      OTU_3060_4      60      70      80      90      100      110
KHTMLCFEIL--SYVSYLAKTKAMGVDIASANYQQEFDLKGRLYDSKTYE-LFKGSN      OTU_3060_4      GGSGSGFMSVFMNHNIRKIHSGLPALADTLAYIVYSMDMKIKVTEKENSIFTEVSKA
... .. : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OTU_3175_18      -YSQSSPDTIAFFVGITADRKCVLDEETYNRRDRQIPLPFS--DIPKIFTEFLKNNR      OTU_3175_17      GCCTNSFWGSDVDFNEKVRKHISGLPALADTLAYIVYSMDMDIAGEGKGRVAFEDISQN
260      270      280      290      300      310      OTU_3060_4      70      80      90      100      110      120
GNSLFEGIYSYTDTADEGDDDFLCSIIIGVYMRAYVLDVYFSKANMEITEKETARRHKEF      OTU_3060_4      VDFNELVGRVVDTLVSDGAFKISVDMTENSVDVPIVEFWSADKVEYRIRGVLKEVV
. . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OTU_3175_18      LWGFANDV--IBSADQAIIELKQKFKRLSSGLYNFIPAFKKYKIIDRIHLQSANLAAGD      OTU_3175_17      TDFTLVGRVVDTLVSDGAFKISVD--DPLSLTPIVEFVGADKIEIRLGLVLEVI
260      270      280      290      300      310      OTU_3060_4      120      130      140      150      160      170
GNSLFEGIYSYTDTADEGDDDFLCSIIIGVYMRAYVLDVYFSKANMEITEKETARRHKEF      OTU_3060_4      VDFNELVGRVVDTLVSDGAFKISVDMTENSVDVPIVEFWSADKVEYRIRGVLKEVV
. . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OTU_3175_18      LWGFANDV--IBSADQAIIELKQKFKRLSSGLYNFIPAFKKYKIIDRIHLQSANLAAGD      OTU_3175_17      TDFTLVGRVVDTLVSDGAFKISVD--DPLSLTPIVEFVGADKIEIRLGLVLEVI
320      330      340      350      360      OTU_3060_4      180      190      200      210      220      230
GNSLFEGIYSYTDTADEGDDDFLCSIIIGVYMRAYVLDVYFSKANMEITEKETARRHKEF      OTU_3060_4      FRSEHNHGRLHLEECYVGGYIESRLDMSGHEVRLDLSVCLSGIERTTFDQGVIMAV
. . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OTU_3175_18      LWGFANDV--IBSADQAIIELKQKFKRLSSGLYNFIPAFKKYKIIDRIHLQSANLAAGD      OTU_3175_17      FRSAHEDGNRIYQLEEHYGRGYIESRLDMSGHEVRLDLSVCLSGIERTTFDQGVIMAV
180      190      200      210      220      230      OTU_3060_4      240      250      260      270      280      290
GNSLFEGIYSYTDTADEGDDDFLCSIIIGVYMRAYVLDVYFSKANMEITEKETARRHKEF      OTU_3060_4      PLKFYASKKYPNRGKSI FDGGKSDCFDLDLDEVISQWMDAIRAGRVTKYIPSDKI PRSPEN
. . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OTU_3175_18      LWGFANDV--IBSADQAIIELKQKFKRLSSGLYNFIPAFKKYKIIDRIHLQSANLAAGD      OTU_3175_17      PLKFYASKKYPNRGKSI FDGGKSDCFDLDLDEVISQWMDAIRAGRVTKYIPSDKI PRSPEN
240      250      260      270      280      290      OTU_3060_4      300      310      320      330      340      350
GNSLFEGIYSYTDTADEGDDDFLCSIIIGVYMRAYVLDVYFSKANMEITEKETARRHKEF      OTU_3060_4      GALQRVNSFGNEFIEIASLGDERSSQIQVQFDIKYDAFVSSYTNCILMLQLGLVSPAT
. . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OTU_3175_18      LWGFANDV--IBSADQAIIELKQKFKRLSSGLYNFIPAFKKYKIIDRIHLQSANLAAGD      OTU_3175_17      GSVGKLNQFGNNYIITISQPLQGVTEFKIEVVQFDIKYDAFVSSYTNCILMLQLGLVSPAT
300      310      320      330      340      350      OTU_3060_4      360      370      380      390      400      410
GNSLFEGIYSYTDTADEGDDDFLCSIIIGVYMRAYVLDVYFSKANMEITEKETARRHKEF      OTU_3060_4      LGIDVGRMSADAQREKVDVTGNTRNTITALEKALPELVSALVTKYDNNMQGKAPEEYEV
. . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OTU_3175_18      LWGFANDV--IBSADQAIIELKQKFKRLSSGLYNFIPAFKKYKIIDRIHLQSANLAAGD      OTU_3175_17      LGIDVGRMSADAQREKVDVTGNTRNTITALEKALPELVSALVTKYDNNMQGKAPEEYEV
360      370      380      390      400      410      OTU_3060_4      420      430      440      450      460      470
GNSLFEGIYSYTDTADEGDDDFLCSIIIGVYMRAYVLDVYFSKANMEITEKETARRHKEF      OTU_3060_4      SVDFGEYGAPDFSRVETVKGASTYGISMSVETQVEELMGSSEDEWKAGEVFRIMQENGL
. . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OTU_3175_18      LWGFANDV--IBSADQAIIELKQKFKRLSSGLYNFIPAFKKYKIIDRIHLQSANLAAGD      OTU_3175_17      SVDFGEYGAPDFSRVETVKGASTYGISMSVETQVEELMGSSEDEWKAGEVFRIMQENGL
420      430      440      450      460      470      OTU_3060_4      480      490
GNSLFEGIYSYTDTADEGDDDFLCSIIIGVYMRAYVLDVYFSKANMEITEKETARRHKEF      OTU_3060_4      ADGTTFAVGDELA*
. . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OTU_3175_18      LWGFANDV--IBSADQAIIELKQKFKRLSSGLYNFIPAFKKYKIIDRIHLQSANLAAGD      OTU_3175_17      ADGATSAVGDELA*
480      490
GNSLFEGIYSYTDTADEGDDDFLCSIIIGVYMRAYVLDVYFSKANMEITEKETARRHKEF      OTU_3060_4      ADGATSAVGDELA*
480      490
GNSLFEGIYSYTDTADEGDDDFLCSIIIGVYMRAYVLDVYFSKANMEITEKETARRHKEF

```

Figure S10: Alignments of TerL and the portal protein for two viruses belonging to the same genus within the *Jeppeviridae* family. The 75% amino acid identity for the portal protein alignment is within the expected range for phages of the same genus, whereas the terminase proteins are so disparate that gene exchange is the most likely explanation.

Additional Results

Determining viral densities

In order to gauge the quality of the virome extractions we estimated the virus particle concentration for a subset of the samples using epifluorescence microscopy. The mean virus-like particle (VLP) concentration obtained was $1.0 \cdot 10^9$ VLPs/g of faeces, ranging from $3 \cdot 10^8$ to $3 \cdot 10^9$ VLPs/g for the 18 samples tested.

Detection of ssDNA viruses while preserving quantitative abundances

Before library preparation we used multiple-displacement amplification (MDA) to capture ssDNA viruses. MDA is believed to introduce compositional biases⁹¹ and compromise quantitative analyses⁹². To control these effects we shortened the MDA step (sMDA) down to 30 minutes instead of the recommended two hours. After sequencing and assembly, vOTU abundances were estimated by read mapping and normalising for mapping depth and contig length. To investigate whether the sMDA had compromised caudoviral abundances, we compared counts of plaque forming units (PFUs) for 15 virulent coliphages⁶⁵ against mapped reads from both the sMDA amplified virome and non-amplified metagenome. Both data sets were quantitative over the entire range of PFU measurements (250 to 1M PFU/g) (Figure S9). Thus, the sMDA protocol did not compromise quantitative analysis of the dsDNA virome, allowing for valid comparisons of viral abundances across different vOTUs and samples.

Manual decontamination and OTU delineation

Assembly of all samples yielded 1.5M contigs in total and the contigs were first dereplicated at 90% nucleotide identity and coverage, resulting in 267k OTUs. Protein coding genes from all OTUs were compared in an all-against-all alignment search for two purposes: First VOGs were delineated *de novo* and annotated with profile-profile alignments against major protein family databases. Secondly, alignments were used to cluster the OTUs by remote protein similarity, yielding 8050 deep “decontamination clusters”.

Decontamination clusters were inspected manually. To avoid manual inspection of all 8050 clusters they were ranked based on aggregate prevalence across samples and extent of CRISPR targeting from the metagenome. Using the VOGs, we made genome maps of all OTUs in each cluster (similar to Figure S2). We inspected the 400 top-ranking clusters manually, and found 255 of them to appear viral (criteria below). Diminishing returns beyond the top 400 mark meant that no additional decontamination clusters were inspected. Of the clusters we did inspect and that were not viral, 64 appeared to be plasmid clusters (e.g. encoding partitioning proteins and type IV secretion systems), while the remainder appeared spurious and were assumed to comprise bacterial contamination. Viral clusters encoded viral structural proteins and were conserved in genome size and gene content and plasmid clusters were much more heterogeneous while contaminant clusters were distinctly short. The 255 viral decontamination clusters comprised 8296 vOTUs at the 90% ANI-level, or 16,746 species-level (95% ANI) vOTUs.

After viral family-level clustering of the 16,746 vOTUs, we found that some VFCs were composed of both viruses and various MGEs of viral origin. The latter appeared to be either satellite phages or other MGEs employing viral polymerases making them co-cluster within the viral VFCs. Thus all vOTUs were curated individually to remove such viral MGEs (criteria in Additional Methods). Further,

VFCs corresponding to known viruses (e.g. Lamba-like phages, Tevenvirinae, etc.) were deliberately discarded in order to avoid interference with ongoing virus classification efforts. In the end the final set of vOTUs was reduced to 10,021 manually confirmed viral species distributed over 248 VFCs.

Estimating the ratio of bacterial contamination

The 10,021 confirmed vOTUs recruited roughly half of the reads despite making up a minor fraction of all OTUs (Figure S3). The remaining half of the read mappings were spread over the 346k non-vOTUs. Since ViromeQC⁴² estimated a mean bacterial contamination rate of 44% at the read-level, we infer that most non-vOTUs must be bacterial (Figure S3). After subtracting viral and bacterial reads, only 7% of the reads were left unaccounted.

Benchmarks of virome decontamination software

A series of virus discovery and decontamination tools have been published recently^{27,28,42,93-98} and they have already seen widespread applications. Although most of the tools are meant for use on metagenome assemblies, or predicted (pro-)phage genomes, they are often applied directly to virome assemblies for decontamination purposes.

Several large gut phage databases have been built on predictions from such tools^{30,31}. Yet, little is known about their efficacy in identifying novel viral clades or their ability to weed out contaminating DNA. The manually curated nature of our virome data set made it well suited for independently testing the performance of these tools. A purely random prediction was generated for comparison. A naive length cutoff of +20 kb was also used for comparison, since non-vOTUs were distinctly short in our data (Figure S4).

CheckV²⁸, VIBRANT²⁷ and viralVerify⁹⁴ sported the best performances with our data set (Table S1) although VIRSorter⁹⁷ also worked well. With a specificity of 97.5%, the length cutoff did a better job than VIRSorter⁹³. DeepVirFinder⁹⁸ PPR-Meta⁹⁶ and Seeker⁹⁵, all of which were “alignment free”, yielded performances that were close to random. VIRSorter and VIBRANT, when run in virome decontamination mode, improved sensitivity at the cost of greatly reduced specificity. For our data, VIRSorter performed better when used in “db2” mode.

ViromeQC⁴² estimated a mean bacterial contamination rate of 44% at the read-level (Figure S3). The median [IQR] contamination rate reported by ViromeQC was 36% [22% - 54%] and this range is similar to a test of published viromes by the authors of ViromeQC. We compared the ViromeQC estimation to two other independent measures generated using the manual curation and our coupled metagenome samples, namely 1) the proportion of mapped reads to non-vOTUs (median [IQR] 42% [24% - 64%]) and 2) the depletion of bacterial core genes in the virome reads compared to cognate metagenomes (median [IQR] 37% [19% - 46%]). Although all three estimates were in strong overall agreement, there was considerable sample to sample variation (Figure S5). Importantly, our metagenome-normalised core gene depletion did not perform better than ViromeQC when comparing against the non-vOTU standard. This result illustrates that virome contaminant estimation is non-trivial even when bulk metagenomes are available for the same samples, and that ViromeQC performs well considering all this uncertainty.

Additional Discussion

Bacterial contaminating DNA made up around half of our sequenced virome reads, which is within the typical range⁴². After assembly and species-level deduplication, the number of non-viral OTUs was 20 times greater than our total number of viral OTUs (Figure S3). vOTUs were longer and more prevalent than contaminating bacterial non-vOTUs which tended to be sample-specific (Figure S4). Random segments of bacterial chromosomal DNA likely became copurified along with the viral particles, explaining why they were not generally conserved between samples. Contaminant DNA species thus made up the majority of the overall sequence diversity but were shorter and much less prevalent than the viruses. Setting cutoffs on OTU length and prevalence may thus provide an effective means for virome decontamination in future studies.

For resolving the taxonomy of our vOTUs into genera, subfamilies, families and order-level clades we used an amino acid identity (AAI) based phylogenomic approach and applied global cutoffs after rooting. Although it has been argued that global cutoffs are not suitable for virus classification³² we found they worked well, and they came with the key advantage of reproducibility at multiple taxon levels. A different phylogenomic approach based on concatenated protein phylogeny was also remarkably successful at various taxonomic ranks among *Caudoviricetes*³³. An alternative method for automatic genomic classification relies on gene-sharing networks rather than trees, and provides satisfactory results at the genus level²⁶, while at the family level, manual inspection of the network is required³⁴. The existing guideline for defining new phage families³⁴ involves manual inspection of gene-sharing networks²⁶, and a reproducible alternative would be preferred. A common concern about clustering viruses based on shared proteins is the co-clustering of phages sharing accessory rather than core genes. However, to this day phylogenomics has proven a robust method for phage classification as it resists the formidable capacity of phages to exchange genetic material.

Although the large terminase subunit (TerL) was the most conserved protein in our caudoviruses, its gene was frequently exchanged such that even members of the same viral genus would carry different TerL homologs (Figure S10). Notable examples of this phenomenon are found in VOC1. Thus, the practice^{35,37} of using TerL phylogeny to classify caudoviral phages can sometimes produce confusing results. As shown by Yutin *et al.*³⁷, *Crassvirales* is not TerL monophyletic and non-Crass phages often encode TerLs that wind up in the middle of the crAss TerL tree. The recent introduction of *ε-crassviridae* into *Crassvirales*³⁷ may illustrate this problem, as our results suggest that the family may not be a crAssphage family (Figure 1).

Finally, we found that the latest generation of metagenome virus discovery tools such as CheckV, viralVerify and VIBRANT, in conjunction with ViromeQC should account for most sequences in one's virome data (Table S1), rendering any "virome dark matter" almost non-existent, at least for our data. Although the sensitivities of the tools against our data never got close to 100%, most of the sequences missed by the best tools were simply too short to pass the imposed quality thresholds. Thus, their predictions were good and certainly easier to obtain than manual curation. On the other hand, the performances of the alignment-free methods were very close to random with our dataset, and it appears that nucleotide-level motifs do not carry the information required to reliably distinguish viruses from their abundant hosts.

Additional Methods

The COPSAC2010 birth cohort

The study was embedded in the Danish population-based COPSAC2010 prospective mother-child cohort of 736 women and their children followed from week 24 of pregnancy, with the primary clinical endpoint persistent wheeze or asthma³⁹.

Faecal sample collection, transport and storage

Of the 700 children part of the COPSAC2010 cohort, 660 delivered a faecal sample 1-year after birth. Faecal samples were collected either at the research clinic or by parents at home using detailed instructions. Upon arrival, each faecal sample was mixed with 10% vol/vol glycerol broth and stored at -80°C until virome extraction.

Virome extraction and library preparation

Preparation of faecal samples, and extraction and sequencing of virions was done using a previously described protocol⁷¹. Briefly, viral-associated DNA was subjected to brief MDA amplification and libraries prepared following manufacturer's procedures for the Illumina Nextera XT kit (FC-131-1096). Libraries were sequenced, paired-end, on the Illumina HiSeq X platform.

VLP counts by epifluorescence

A volume of 10 µL per virome sample was diluted 100-fold in SM buffer, fixed with 0.5% glutaraldehyde, and frozen in liquid nitrogen. The sample was thawed, deposited on a 0.02 µM pore size membrane (Anodisc 25, Anopore, Whatman) using a filtering device. Next, the filter was dried, incubated on a 70 µL drop of SYBRTM-Gold (200X) in the dark for 15 minutes, dried, and mounted on a microscope slide, together with fluoromount and antifade. VLP dots were visualized using an epifluorescence microscope equipped with an ORCA camera, and a 475 nm excitation laser. Eight to ten fields (1344x1024 pixels) were imaged and fluorescent dots counted. VLP counts on the total filter surface were deduced by multiplying the average field count by the number of fields over the total membrane surface (60 493).

Virome assembly and gene calling

Virome libraries were sequenced to an average depth of 3 GB per sample with paired end 2x150 bp reads. Satisfactory sequencing results were obtained for 647 samples. Virome reads were first quality filtered and trimmed using Fastq Quality Trimmer/Filter (options -Q 33 -t 13 -l 32 -p 90 -q 13), and residual Illumina adapters were subsequently removed using cutadapt. The resulting trimmed reads were dereplicated using vsearch derep_prefix and then assembled with Spades v3.10.1 using the meta flag and disabling read error correction. In order to reduce sequence redundancy for the subsequent decontamination step, all 1.5M contigs above 1 kb in length were clustered at 90% ANI and coverage using a previously described approach⁷². Longest representatives from each of the 267K clusters were subjected to gene calling⁷³ and the resulting protein sequences were used to build an APS tree⁷⁵ and for *de novo* VOG definition⁷⁷ (more details below).

The aggregate protein similarity (APS) tree

Sharing features with DICE⁹⁹ and VIPtree¹⁰⁰, the APS tree methodology was first developed⁷⁵ for systematic and automated classification of prokaryotic defence systems and mobile genetic elements that were traditionally classified by manual inspection. The tree was found to reproduce existing manual classes and sub-classes when cutting it with global cutoffs^{75,101}. The tree also predicted deeper relationships between classes that were experimentally validated later¹⁰². The method uses as input, an all-against-all protein alignment search, preferably with sensitive alignment software like FASTA⁷⁴, although BLASTP or DIAMOND also work. We used FASTA 36.3.8g. For all viruses being compared, all constituent viral proteins are searched against each other. Next, virus to virus similarity scores are tallied by aggregating the alignment scores for all their constituent proteins. Next, using the Bray-Curtis dissimilarity¹⁰³, a distance matrix is constructed between all viruses, which is then used as input for constructing a neighbour joining tree. We used RapidNJ (<https://github.com/somme89/rapidNJ>) for the latter. In our case distances between most species in the tree were 100% because they were unrelated to most other species (unlike a phylogenetic tree built on a multiple sequence alignment (MSA), where all species are required to have a common ancestor). This feature makes the tree distinctly star-shaped and any clade that emanates from the centre of the star is suitable for rooting. After rooting the tree, it can be cut at fixed distances from the root to reproduce classes at required levels. To this end we developed “treetool” described further below. Using more sensitive alignment software (like FASTA) generates more reliable distant protein alignments, making deeper cuts less noisy. In this study we found that cuts down to the viral order-level were reproducing existing and proposed taxonomy, making such cuts useful for defining novel taxa as well.

Protein annotation, clustering and visualisation

Protein coding genes for representative contigs during initial decontamination, and for vOTUs and reference phages later were predicted using Prodigal⁷³. All protein sequences were subject to an all-against-all sequence alignment using FASTA⁷⁴. Protein alignments between viruses were used to cluster viruses into taxa using the APS tree as described above and below. But the alignments were also used to define VOGs *de novo* using a previously described orthology detection pipeline involving protein alignment coverage cutoffs⁷⁷ and Markov Clustering¹⁰⁴. MSAs and phylogenetic trees were constructed¹⁰⁵ from the protein sequences corresponding to each VOG. MSAs were used for profile-profile alignments against the Pfam⁷⁹, COG⁸¹, CDD⁸⁰, TIGRFAMs⁸² and PhROGs⁸⁵ databases using HHblits⁸³ in order to annotate them and determine their functions. Gene contents of the contigs were visualised along with other contigs belonging to the same viral clusters by colouring genes by orthology and hyperlinking profile-profile results in SVG graphics as in Figure S2. The SVG files were generated using a custom bash script. The graphics were browsed manually during the manual decontamination, species curation and family curation steps (Figure S1). The curated family visualisations and VOGs can be accessed through the online version of Figure 1 (<http://copsac.com/earlyvir/f1y/fig1.svg>), and a direct link to the final VOG annotations against PhROGs can be found here: <http://copsac.com/earlyvir/f1y/vogs.html>.

Manual decontamination of assembled contigs

All proteins from the 267K 90% ANI contig cluster representatives were subject to an all-against-all FASTA⁷⁴ search that was used for defining VOGs⁷⁷. The search results were also used to construct an APS tree that was cut close to the root in order to yield 8050 decontamination clusters of contig representatives. SVG graphics were generated for each decontamination cluster by visualising each

individual representative within it. This was done by using the gene orthology information generated earlier to colour rectangles corresponding to encoded genes. The decontamination clusters were then ranked by their sample prevalence and extent of CRISPR targeting from our MAGs⁴⁰. SVG graphics for the first 400 ranked decontamination clusters were manually inspected and the 255 clusters resembling viral families were flagged. Viral clusters were recognised by having conserved genome sizes with the majority of the genes and synteny being largely conserved between most members within the cluster. Plasmid clusters, on the other hand, were often heterogeneous in size and extremely variable with regard to gene content, and would encode mobilisation proteins, relaxases or Type IV secretion systems. Bacterial contaminant clusters were made up of short contigs sharing one or two genes, but otherwise variable in size with any remaining genes being disparate. Clusters ranking lower than the top 400 were not inspected further due to diminishing returns, as most additionally inspected clusters at this point resembled bacterial contamination. For the 255 flagged viral clusters, contigs smaller than 10 kb were discarded unless the apparent complete genome size for that cluster was shorter, as for the ssDNA viruses.

Species-level OTU delineation

Species-level deduplication was necessary because very similar viruses from different samples were assembled into separate contigs, as assemblies were carried out separately per sample. These had to be merged into single OTUs and a representative for each OTU had to be chosen. This was done by comparing all assembled contigs from all samples to each other, to find clusters of very similar contigs. In order to account for incomplete assemblies, we selected the longest version as the OTU representative sequence. However, the longest version of a species was sometimes a chimeric assembly, where two closely related viruses had been merged into a long contig e.g. containing two copies of each gene. To avoid selection of chimeras as OTU representatives, all contigs with more than 110% self-similarity (judging from the self-alignment score over length) were flagged as potentially repetitive and were not selected as a representative. In such cases the next-longest sequence was selected to represent the OTU. A variety of tools were tested for comparing the contigs and generating OTUs at the 95% species-level ANI, including *cd-hit-est*¹⁰⁶, *BB dedupe*¹⁰⁷ and *nucmer*¹⁰⁸. The most consistent results were obtained using *BLAT*⁷⁸ and setting a cutoff on the alignment score. This was done by requiring a score of 90% of what the shorter sequence obtained against itself, while aligned against a longer sequence, in which case the two were merged. The 90% score would correspond on average to 95% identity at 95% coverage, but at extremes in either direction could also occasionally include 100% identity at 90% coverage or 90% identity at 100% coverage. This flexibility was a compromise offset by the high accuracy of *BLAT* vs. other tools tested. The clustering step itself was carried out with a perl one-liner applied on top of an all-against-all *BLAT* output, where all assembly contigs had been used as input.

PhyloTreeLib.py for exploring and cutting the APS tree

phylotreeLib (<https://github.com/agormp/phylotreeLib>) is a Python library developed in-house for manipulating or extracting information from phylogenetic trees. Here, we used *phylotreeLib.py* via its command line front-end *treetool.py* (<https://github.com/agormp/treetool>). After constructing the APS tree it was visually inspected in *FigTree* (<https://github.com/rambaut/figtree/>) to find a suitable outgroup. Suitable outgroups comprised branches originating directly from the stem of the tree. The stem was easily visible due to the high diversity of sequences used to construct the APS tree. Next, *treetool.py* was used with the `--root` option to root the APS tree based on the selected outgroup. Next, a list of accession numbers was made containing all members of the *Herelleviridae* family, along with separate lists for members of different subfamilies and genera within the *Herelleviridae*.

Another list was made containing all known crAssphages at the time of the analysis. Then, `treetool.py` was used along with the `--cladeinfo` option to retrieve the distances from the root to the branch encompassing the leaves in the lists. Next, `treetool.py`'s `--clustcut` option was used to cut the tree at the above distances in order to obtain clades of vOTUs and reference phages corresponding to viral orders, families, subfamilies and genera. The distances we used to cut the tree were 0.250, 0.125, 0.04 and 0.025 respectively for genus, subfamily, family, and order-level clades. Thus, the approximate minimum AAI and coverage required for two viruses to belong to the same clade could be derived from those cut distances to be on average 70%, 50%, 28%, and 22%.

Generation of viral family-level clades (VFCs) and species-level curation

To find viral species corresponding to the flagged viral decontamination clusters, all 1.5M contigs were clustered at 95% ANI, resulting in 362K OTUs. The 8327 viral sequences from the decontamination step were aligned against 95% ANI vOTUs requiring a 50% BLAT score (~75% ANI at ~75% coverage). The 16,746 95% ANI vOTUs recruited in this manner were pooled with 95% ANI dereplicated reference phages from millardlab.org⁴³. Protein coding genes were annotated for all sequences and subject to an all-against-all FASTA search. The *de novo* VOGs and APS tree were recomputed. This was done to distinguish novel viral families from known ones. The tree was rooted manually using `phylotree`, and the *Herelleviridae* family-level root distance was measured using `treetool`'s `cladeinfo` function. *Herelleviridae* was used because the ICTV has published the complete taxonomy for the family as a reference³⁴ to guide future bacteriophage taxonomy. The tree was cut at the family-level root distance yielding 1259 viral family-level clades. In order to make sure that the large number of families was not a result of accidental recruitment of additional contaminant contigs during the 75% ANI matching step from above, each vOTU and reference phage within each family-level clade was visualised using the VOG information as seen in Figure S2. Next, each individual vOTU was manually flagged as belonging to either one of the categories virus, putative satellite, putative MGE or unknown contaminant using the following guidelines. To be considered viral, a vOTU needed to encode at least one structural protein along with an additional viral protein. vOTUs that did not live up to these criteria were considered MGEs if they encoded at least one protein indicative of a mobile lifestyle (e.g. integrases or replication proteins), or if they were conserved in size and gene content across multiple samples. The rationale for the latter was that random segments of contaminating chromosomal DNA should not be conserved in size and gene content, while any mobile DNA would be expected to be so. Unknown viruses that encoded novel structural proteins could thus end up in the MGE category and this is a shortcoming of the approach. MGEs that co-clustered within viral family-level clades, or contained at least two non-structural viral proteins were classed as putative satellites. vOTUs that did not live up to any criteria were classed as unknown contaminants. These must have slipped through the previous decontamination step, but only comprised 458 out of a total 16,746 vOTUs.

Final VFC curation

All 1259 family-level clades were manually inspected to make sure that they looked like true families instead of being artefacts of the methods employed. The majority of the family-level clades consisted of singleton vOTUs or reference phages, and these were discarded. Additionally, families consisting mostly of reference phages were discarded to avoid interference with independent classification efforts. Finally, VFCs that looked unconvincing because they consisted of mostly MGEs, satellites, or genome fragments were also removed, leaving 248 curated VFCs in the end. For each curated

VFC, a histogram of vOTU length was made using a bin width of 5 kb. The minimum genome completion size for each family was determined by inspecting each histogram and choosing the 5-kb bin immediately preceding the rightmost peak. This size cutoff was used subsequently to select all complete and near-complete vOTUs within each family. Families that did not have enough vOTUs to yield a convincing genome size peak were flagged as having an unknown completion size. The complete genome size cutoff along with the curated list of integrase VOGs (1, 10, 49, 95, 261, 302, 400, 493, 613, 635, 688, 750, 765, 800, 1386, 2501, 2668, 3230, 4214, 4529 from <http://copsac.com/earlyvir/f1y/vogs.htm>) was used to predict phage lifestyles at the family-level using the rules mentioned in the methods section.

Host prediction

As part of a separate effort⁴⁰, metagenomes from the same samples had been assembled and binned into MAGs that had taxonomies predicted using GTDB-tk. CRISPR arrays were predicted on each MAG using CRISPRdetect⁷⁶, and the spacers were pooled and used as a host prediction database, alongside the CRISPR spacer database⁵⁷. The host prediction algorithm from the CRISPR spacer database was used to predict hosts for each vOTU using either spacer database. An additional host prediction was made using WISH v1.0⁵⁸, against a database of complete bacterial and archaeal genomes (downloaded from <ftp://ftp.ncbi.nlm.nih.gov/genomes> in January of 2020) with null-models calibrated against all reference phage sequences available at millardlab.org⁹¹ (also in January of 2020). The best maximum-likelihood WISH prediction per vOTU having a P-value under 0.05 was used. Host predictions from all three methods were translated into NCBI taxonomy identifiers and integrated by taking the LCA of the two closest genus-level predictions. E.g. if WISH predicted *Bacteroides*, public spacers *Clostridium* and metagenome spacers *Faecalibacterium*, then the host was set to order *Clostridiales*, being the LCA of the two most closely matching predictions. The rationale for this approach was to absorb occasional false positives stemming from either of the three methods, as all three methods had been seen to make false predictions on separate occasions. However, most predictions were unanimous at the genus-level.

Quantification of bacterial contamination

Three parallel approaches were used for quantifying the amount of bacterial contamination in the virome samples. For the first approach, ViromeQC was run on the QC'd virome reads, and the proportion of bacterial contamination per sample was determined by taking the inverse of the “virus enrichment ratio” as reported by ViromeQC.

For the second approach we wanted to take advantage of our manual curation, and simply set the ratio of bacterial contamination per sample to the proportion of reads mapping to non-vOTUs.

In the third approach, we took advantage of our coupled metagenomes for each virome sample, hoping to gain a more accurate estimate than ViromeQC, which does not utilise such additional information. We did this by first searching reads from the virome with the Rfam 16S rRNA gene model (RF00177). Then the same was done for reads from cognate metagenomes. The ratio between the proportion of reads mapping to RF00177 in the virome over the same proportion for the metagenome was counted as the estimated bacterial contamination ratio by 16S rRNA gene content. In parallel, reads from both the viromes and metagenomes were mapped against a database of the bacterial *cpn60* core gene from cpnDB⁸⁹, using bowtie2¹⁰⁹. As we did for the 16S rRNA gene-based contamination estimate, the proportion of reads mapping to this database in the virome over the metagenome for each sample was counted as the estimated bacterial contamination ratio by *cpn60*

content. The mean between the two estimates was used as the “metagenome core gene depletion” contamination-estimate used in Figure S5.

Computing vOTU and VFC abundances

vOTU counts were generated by mapping the reads of each sample against only the assembled contigs from that particular sample. Mapping was performed using the `bwa mem -a` flag⁹⁰. Contig counts were generated with `msamtools profile` (<https://github.com/arumugamlab/msamtools>), which computes relative abundances of each contig, normalising by contig length and mapping depth while iteratively redistributing non-uniquely mapping reads. The 95% ANI dereplication cluster information merged onto the contig abundance table to compute OTU abundances. The vOTU to sample relative abundance matrix, or OTU table, was loaded into a phyloseq object in R and agglomerated at the viral family and order-levels to compute the statistics used in Figures 2 and 3.

Decontamination tool benchmarking

The versions of virus prediction tools used were DeepVirFinder v1.0, VIBRANT v1.2.1, VIRSorter 1.0.6, VIRSorter2 commit 22f6a7d, Seeker commit 9ae1488, PPR-Meta v1.1, and CheckV v.0.7.0. Some of these tools were designed for virus discovery in metagenomes, while others meant for virome decontamination, but the two are related given the high amount of bacterial contamination in most viromes published to date⁴². Despite our own efforts to obtain high-quality virome samples, the vast majority of our OTUs (in absolute numbers, not sequencing reads) were bacterial contamination rather than viral. For this reason no distinction was made between the two types of tools. For the tools that required cutoffs, we used either the recommendations of the authors, default parameters, or if such information was not available, we tried to set a cutoff that would generate a number of positive predictions that reasonably matched our own positive set. In any case, the specific cutoffs used are shown in Table S1. Since our data set is so skewed, with 10,021 positive viral sequences versus over thirty times more negatives, performance measures can be biased, which is why we included the “kappa” metric, designed to address this issue, as well as a random prediction for comparison. The random prediction was created by randomly subsampling the 362k OTUs 12,500 times without replacement. The number 12,500 was chosen because it was reasonably close to our own positive set and the number of positives generated by most tools.

Figures and statistical analysis

Figure 1 was drawn by first collating data at the family-level using phyloseq⁸⁷ then using Circos v0.69-8¹¹⁰ for rendering. Figures 2, 3, 4, S4, S5, S6, S7, S8 and corresponding statistical analyses were generated using the statistical software R and the tidyverse suite, including ggplot2¹¹¹ and related add-on packages ggraph¹¹², ggforce¹¹³, ggpubr¹¹⁴, ggrepel¹¹⁵, ggstance¹¹⁶, and patchwork¹¹⁷. For deriving unique branch lengths (fig 4), we used the function `pd.calc` from the `caper` package¹¹⁸. The `neutral.fit` function from the `MicEco` R library (<https://github.com/Russel88/MicEco>) was used for fitting the family-level abundances to the neutral community model.

References

1. Moeller, A. H. *et al.* Cospeciation of gut microbiota with hominids. *Science* **353**, 380–382 (2016).
2. Milani, C. *et al.* The first microbial colonizers of the human gut: Composition, activities, and health implications of the infant gut microbiota. *Microbiol. Mol. Biol. Rev.* **81**, (2017).
3. Johnson, C. C. & Ownby, D. R. The infant gut bacterial microbiota and risk of pediatric asthma and allergic diseases. *Transl. Res.* **179**, 60–70 (2017).
4. Kalliomäki, M., Collado, M. C., Salminen, S. & Isolauri, E. Early differences in fecal microbiota composition in children may predict overweight. *Am. J. Clin. Nutr.* **87**, 534–538 (2008).
5. Stokholm, J. *et al.* Maturation of the gut microbiome and risk of asthma in childhood. *Nat. Commun.* **9**, 141 (2018).
6. Bisgaard, H. *et al.* Reduced diversity of the intestinal microbiota during infancy is associated with increased risk of allergic disease at school age. *J. Allergy Clin. Immunol.* **128**, 646–52.e1–5 (2011).
7. Pascal, V. *et al.* A microbial signature for Crohn’s disease. *Gut* **66**, 813–822 (2017).
8. Tomofuji, Y. *et al.* Whole gut virome analysis of 476 Japanese revealed a link between phage and autoimmune disease. *Ann. Rheum. Dis.* **81**, 278–288 (2022).
9. Ott, S. J. *et al.* Efficacy of Sterile Fecal Filtrate Transfer for Treating Patients With *Clostridium difficile* Infection. *Gastroenterology* **152**, 799–811.e7 (2017).
10. Rasmussen, T. S. *et al.* Faecal virome transplantation decreases symptoms of type 2 diabetes and obesity in a murine model. *Gut* **69**, 2122–2130 (2020).
11. Brunse, A. *et al.* Fecal filtrate transfer protects against necrotizing enterocolitis in preterm pigs. doi:10.1101/2020.05.25.114751.
12. Shkoporov, A. N. & Hill, C. Bacteriophages of the Human Gut: The ‘Known Unknown’ of the Microbiome. *Cell Host Microbe* **25**, 195–209 (2019).
13. Siljander, H., Honkanen, J. & Knip, M. Microbiome and type 1 diabetes. *EBioMedicine* **46**, 512–521 (2019).
14. Liang, G. *et al.* The stepwise assembly of the neonatal virome is modulated by breastfeeding. *Nature* **581**, 470–474 (2020).

15. Lim, E. S. *et al.* Early life dynamics of the human gut virome and bacterial microbiome in infants. *Nat. Med.* **21**, 1228–1234 (2015).
16. Zhao, G. *et al.* Intestinal virome changes precede autoimmunity in type I diabetes-susceptible children. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E6166–E6175 (2017).
17. Hobbs, Z. & Abedon, S. T. Diversity of phage infection types and associated terminology: the problem with ‘Lytic or lysogenic’. *FEMS Microbiology Letters* vol. 363 fnw047 (2016).
18. Bernheim, A. & Sorek, R. The pan-immune system of bacteria: antiviral defence as a community resource. *Nat. Rev. Microbiol.* **18**, 113–119 (2020).
19. Sorek, R., Kunin, V. & Hugenholtz, P. CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat. Rev. Microbiol.* **6**, 181–186 (2008).
20. Fluckiger, A. *et al.* Cross-reactivity between tumor MHC class I-restricted antigens and an enterococcal bacteriophage. *Science* **369**, 936–942 (2020).
21. Górski, A. *et al.* Perspectives of phage therapy in non-bacterial infections. *Front. Microbiol.* **9**, 3306 (2018).
22. Dufour, N., Delattre, R., Chevallereau, A., Ricard, J.-D. & Debarbieux, L. Phage therapy of pneumonia is not associated with an overstimulation of the inflammatory response compared to antibiotic treatment in mice. *Antimicrob. Agents Chemother.* **63**, (2019).
23. Breitbart, M. *et al.* Viral diversity and dynamics in an infant gut. *Res. Microbiol.* vol. 159 367–373 (2008).
24. McCann, A. *et al.* Viromes of one year old infants reveal the impact of birth mode on microbiome diversity. *PeerJ* **6**, e4694 (2018).
25. Aggarwala, V., Liang, G. & Bushman, F. D. Viral communities of the human gut: metagenomic analysis of composition and dynamics. *Mob. DNA* **8**, 12 (2017).
26. Bin Jang, H. *et al.* Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat. Biotechnol.* **37**, 632–639 (2019).
27. Kieft, K., Zhou, Z. & Anantharaman, K. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* **8**, 90 (2020).

28. Nayfach, S. *et al.* CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat. Biotechnol.* (2020) doi:10.1038/s41587-020-00774-7.
29. Nayfach, S. *et al.* Metagenomic compendium of 189,680 DNA viruses from the human gut microbiome. *Nat Microbiol* **6**, 960–970 (2021).
30. Camarillo-Guerrero, L. F., Almeida, A., Rangel-Pineros, G., Finn, R. D. & Lawley, T. D. Massive expansion of human gut bacteriophage diversity. *Cell* **184**, 1098–1109.e9 (2021).
31. Gregory, A. C. *et al.* The gut virome database reveals age-dependent patterns of virome diversity in the human gut. *Cell Host Microbe* **28**, 724–740.e8 (2020).
32. Simmonds, P. *et al.* Consensus statement: Virus taxonomy in the age of metagenomics. *Nat. Rev. Microbiol.* **15**, 161–168 (2017).
33. Low, S. J., Džunková, M., Chaumeil, P.-A., Parks, D. H. & Hugenholtz, P. Evaluation of a concatenated protein phylogeny for classification of tailed double-stranded DNA viruses belonging to the order Caudovirales. *Nat Microbiol* **4**, 1306–1315 (2019).
34. Barylski, J. *et al.* Analysis of spounaviruses as a case study for the overdue reclassification of tailed phages. *Syst. Biol.* **69**, 110–123 (2020).
35. Benler, S. *et al.* Thousands of previously unknown phages discovered in whole-community human gut metagenomes. *Microbiome* **9**, 78 (2021).
36. 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).
37. Yutin, N. *et al.* Analysis of metagenome-assembled viral genomes from the human gut reveals diverse putative CrAss-like phages with unique genomic features. *Nat. Commun.* **12**, 1044 (2021).
38. Koonin, E. V. *et al.* Global organization and proposed megataxonomy of the virus world. *Microbiol. Mol. Biol. Rev.* **84**, (2020).
39. Bisgaard, H. *et al.* Deep phenotyping of the unselected COPSAC 2010 birth cohort study. *Clin. Exp. Allergy* **43** 1384–1394 (2013).
40. Li, X. *et al.* The infant gut resistome associates with *E. coli*, environmental exposures, gut microbiome maturity, and asthma-associated bacterial composition. *Cell Host Microbe* (2021) doi:10.1016/j.chom.2021.03.017.

41. Roux, S., Krupovic, M., Debroas, D., Forterre, P. & Enault, F. Assessment of viral community functional potential from viral metagenomes may be hampered by contamination with cellular sequences. *Open Biol.* **3**, 130160 (2013).
42. Zolfo, M. *et al.* Detecting contamination in viromes using ViromeQC. *Nat. Biotechnol.* **37**, 1408–1412 (2019).
43. Cook, R. *et al.* Infrastructure for a PHAge REference Database: Identification of Large-Scale Biases in the Current Collection of Cultured Phage Genomes. *PHAGE* vol. 2 214–223 (2021).
44. Roguet, A. *et al.* Neutral community model explains the bacterial community assembly in freshwater lakes. *FEMS Microbiol. Ecol.* **91**, (2015).
45. Venkataraman, A. *et al.* Application of a neutral community model to assess structuring of the human lung microbiome. *mBio* **6**, (2015).
46. Dutilh, B. E. *et al.* A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat. Commun.* **5**, 4498 (2014).
47. Murphy, J. *et al.* Comparative genomics and functional analysis of the 936 group of lactococcal Siphoviridae phages. *Sci. Rep.* **6**, 21345 (2016).
48. 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).
49. Cornuault, J. K. *et al.* Phages infecting *Faecalibacterium prausnitzii* belong to novel viral genera that help to decipher intestinal viromes. *Microbiome* **6**, 65 (2018).
50. Benler, S. *et al.* A diversity-generating retroelement encoded by a globally ubiquitous *Bacteroides* phage. *Microbiome* **6**, 191 (2018).
51. Redgwell, T. A. *et al.* Prophages in the infant gut are largely induced, and may be functionally relevant to their hosts. *bioRxiv* 2021.06.25.449885 (2021)
doi:10.1101/2021.06.25.449885.
52. Freer, G. *et al.* The virome and its major component, Anellovirus, a convoluted system molding human immune defenses and possibly affecting the development of asthma and respiratory diseases in childhood. *Front. Microbiol.* **9**, 686 (2018).
53. Roux, S. *et al.* Cryptic inoviruses revealed as pervasive in bacteria and archaea across

- Earth's biomes. *Nat Microbiol* **4**, 1895–1906 (2019).
54. Mavrich, T. N. & Hatfull, G. F. Bacteriophage evolution differs by host, lifestyle and genome. *Nat Microbiol* **2**, 17112 (2017).
55. Paez-Espino, D. *et al.* Uncovering Earth's virome. *Nature* vol. 536 425–430 (2016).
56. Hubbell, S. P. & Borda-de-Água, L. The unified neutral theory of biodiversity and biogeography: Reply. *Ecology* vol. 85 3175–3178 (2004).
57. Dion, M. B. *et al.* Streamlining CRISPR spacer-based bacterial host predictions to decipher the viral dark matter. *Nucleic Acids Res.* **49**, 3127–3138 (2021).
58. Galiez, C., Siebert, M., Enault, F., Vincent, J. & Söding, J. WISH: who is the host? Predicting prokaryotic hosts from metagenomic phage contigs. *Bioinformatics* **33**, 3113–3114 (2017).
59. Karcher, N. *et al.* Genomic diversity and ecology of human-associated Akkermansia species in the gut microbiome revealed by extensive metagenomic assembly. *Genome Biol.* **22**, 209 (2021).
60. 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).
61. Roux, S. *et al.* Towards quantitative viromics for both double-stranded and single-stranded DNA viruses. *PeerJ* **4**, e2777 (2016).
62. Garmaeva, S. *et al.* Stability of the human gut virome and effect of gluten-free diet. *Cell Rep.* **35**, 109132 (2021).
63. Millen, A. M. *et al.* Lactococcus lactis type III-A CRISPR-Cas system cleaves bacteriophage RNA. *RNA Biol.* **16**, 461–468 (2019).
64. Waller, A. S. *et al.* Classification and quantification of bacteriophage taxa in human gut metagenomes. *ISME J.* **8**, 1391–1402 (2014).
65. Mathieu, A. *et al.* Virulent coliphages in 1-year-old children fecal samples are fewer, but more infectious than temperate coliphages. *Nat. Commun.* **11** (2020).
66. 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).

67. Cornuault, J. K. *et al.* The enemy from within: a prophage of *Roseburia intestinalis* systematically turns lytic in the mouse gut, driving bacterial adaptation by CRISPR spacer acquisition. *ISME J.* **14**, 771–787 (2020).
68. Tyler, J. S. *et al.* Prophage induction is enhanced and required for renal disease and lethality in an EHEC mouse model. *PLoS Pathog.* **9**, e1003236 (2013).
69. Clooney, A. G. *et al.* Whole-virome analysis sheds light on viral dark matter in inflammatory bowel disease. *Cell Host Microbe* **26**, 764–778.e5 (2019).
70. Bisgaard, H. *et al.* Deep phenotyping of the unselected COPSAC2010 birth cohort study. *Clin. Exp. Allergy* **43**, 1384–1394 (2013).
71. Deng, L. *et al.* A Protocol for extraction of infective viromes suitable for metagenomics sequencing from low volume fecal samples. *Viruses* **11**, (2019).
72. Moreno-Gallego, J. L. *et al.* Virome diversity correlates with intestinal microbiome diversity in adult monozygotic twins. *Cell Host Microbe* **25**, 261–272.e5 (2019).
73. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
74. Pearson, W. R. Finding protein and nucleotide similarities with FASTA. *Curr. Protoc. Bioinformatics* **53**, 3.9.1–3.9.25 (2016).
75. Vestergaard, G., Garrett, R. A. & Shah, S. A. CRISPR adaptive immune systems of Archaea. *RNA Biol.* **11**, 156–167 (2014).
76. Biswas, A., Staals, R. H. J., Morales, S. E., Fineran, P. C. & Brown, C. M. CRISPRDetect: A flexible algorithm to define CRISPR arrays. *BMC Genomics* **17**, 356 (2016).
77. Shah, S. A. *et al.* Comprehensive search for accessory proteins encoded with archaeal and bacterial type III CRISPR-cas gene cassettes reveals 39 new *cas* gene families. *RNA Biol.* **16**, 530–542 (2019).
78. Kent, W. J. BLAT--the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
79. Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Res.* **49** D412–D419 (2021).
80. Lu, S. *et al.* CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res.* **48**, D265–D268 (2020).

81. Tatusov, R. L. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 28 33–36 (2000).
82. Haft, D. H. The TIGRFAMs database of protein families. *Nucleic Acids Res.* 31 371–373 (2003).
83. Steinegger, M. *et al.* HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* **20**, 473 (2019).
84. Barylski, J. *et al.* Analysis of Spounaviruses as a Case Study for the Overdue Reclassification of Tailed Phages. *Syst. Biol.* **69**, 110–123 (2020).
85. Terzian, P. *et al.* PHROG: families of prokaryotic virus proteins clustered using remote homology. *NAR Genom Bioinform* **3**, lqab067 (2021).
86. Milanese, A. *et al.* Microbial abundance, activity and population genomic profiling with mOTUs2. *Nat. Commun.* **10**, 1014 (2019).
87. McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**, e61217 (2013).
88. Kalvari, I. *et al.* Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res.* **49**, D192–D200 (2021).
89. Hill, J. E., Penny, S. L., Crowell, K. G., Goh, S. H. & Hemmingsen, S. M. cpnDB: a chaperonin sequence database. *Genome Res.* **14**, 1669–1675 (2004).
90. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
91. Roux, S., Emerson, J. B., Elie-Fadrosh, E. A. & Sullivan, M. B. Benchmarking viromics: an evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ* **5**, e3817 (2017).
92. Yilmaz, S., Allgaier, M. & Hugenholtz, P. Multiple displacement amplification compromises quantitative analysis of metagenomes. *Nat. Methods* **7**, 943–944 (2010).
93. Guo, J. *et al.* VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* **9**, 37 (2021).
94. Antipov, D., Raiko, M., Lapidus, A. & Pevzner, P. A. Metaviral SPAdes: assembly of viruses from metagenomic data. *Bioinformatics* **36**, 4126–4129 (2020).

95. Auslander, N., Gussow, A. B., Benler, S., Wolf, Y. I. & Koonin, E. V. Seeker: alignment-free identification of bacteriophage genomes by deep learning. *Nucleic Acids Res.* **48**, e121 (2020).
96. Fang, Z. *et al.* PPR-Meta: a tool for identifying phages and plasmids from metagenomic fragments using deep learning. *Gigascience* **8**, (2019).
97. Roux, S., Enault, F., Hurwitz, B. L. & Sullivan, M. B. VirSorter: mining viral signal from microbial genomic data. *PeerJ* **3**, e985 (2015).
98. Ren, J. *et al.* Identifying viruses from metagenomic data using deep learning. *Quantitative Biology* vol. 8 64–77 (2020).
99. Mizuno, C. M., Rodriguez-Valera, F., Kimes, N. E. & Ghai, R. Expanding the marine virosphere using metagenomics. *PLoS Genet.* **9**, e1003987 (2013).
100. Nishimura, Y. *et al.* ViPTree: the viral proteomic tree server. *Bioinformatics* **33**, 2379–2380 (2017).
101. Wang, H., Peng, N., Shah, S. A., Huang, L. & She, Q. Archaeal extrachromosomal genetic elements. *Microbiol. Mol. Biol. Rev.* **79**, 117–152 (2015).
102. Lin, J. *et al.* DNA targeting by subtype I-D CRISPR-Cas shows type I and type III features. *Nucleic Acids Res.* **48**, 10470–10478 (2020).
103. Bray, J. R. & Curtis, J. T. An ordination of the upland forest communities of Southern Wisconsin. *Ecol. Monogr.* **27**, 325–349 (1957).
104. Enright, A. J., Van Dongen, S. & Ouzounis, C. A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**, 1575–1584 (2002).
105. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004).
106. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
107. Bushnell, B., Rood, J. & Singer, E. BBMerge – Accurate paired shotgun read merging via overlap. *PLOS ONE* vol. 12 e0185056 (2017).
108. Delcher, A. L., Salzberg, S. L. & Phillippy, A. M. Using MUMmer to identify similar regions in large sequence sets. *Current Protocols in Bioinformatics* Chapter 10 (2003).
109. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature*

Methods vol. 9 357–359 (2012).

110. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
111. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis.* (2016).
112. Pedersen, T. L. *ggraph: An Implementation of Grammar of Graphics for Graphs and Networks.* (2019).
113. Pedersen, T. L. *ggforce: Accelerating 'ggplot2'.* (2019).
114. Kassambara, A. *ggpubr: 'ggplot2' Based Publication Ready Plots.* (2020).
115. Slowikowski, K. *ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'.* (2019).
116. Henry, L., Wickham, H. & Chang, W. *ggstance: Horizontal 'ggplot2' Components.* (2020).
117. Pedersen, T. L. *patchwork: The Composer of Plots.* (2020).
118. Orme, D. *et al.* *caper: Comparative Analyses of Phylogenetics and Evolution in R.* (2018).

Références bibliographiques

- Acs, N., M. Gambino, and L. Brondsted. 2020. "Bacteriophage Enumeration and Detection Methods." *Front Microbiol* 11:594868. doi: 10.3389/fmicb.2020.594868.
- Agus, A., K. Clement, and H. Sokol. 2021. "Gut microbiota-derived metabolites as central regulators in metabolic disorders." *Gut* 70 (6):1174-1182. doi: 10.1136/gutjnl-2020-323071.
- Agus, A., J. Planchais, and H. Sokol. 2018. "Gut Microbiota Regulation of Tryptophan Metabolism in Health and Disease." *Cell Host Microbe* 23 (6):716-724. doi: 10.1016/j.chom.2018.05.003.
- Alberti, G., M. Mazzola, C. Gagliardo, A. Pitruzzella, A. Fucarini, M. Giammanco, G. Tomasello, and F. Carini. 2021. "Extracellular vesicles derived from gut microbiota in inflammatory bowel disease and colorectal cancer." *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub* 165 (3):233-240. doi: 10.5507/bp.2021.042.
- Alves de Matos, A., Lehours, P., Timóteo, A., Roxo-Rosa, M., & Vale, F. 2013. "Comparison of induction of B45 Helicobacter pylori prophage by acid and UV radiation." *Microscopy and Microanalysis* 19 (S4):27-28. doi: doi:10.1017/S1431927613000755.
- Amosse, J., M. Durcin, M. Mallocci, L. Vergori, A. Fleury, F. Gagnadoux, S. Dubois, G. Simard, J. Boursier, O. Hue, M. C. Martinez, R. Andriantsitohaina, and S. Le Lay. 2018. "Phenotyping of circulating extracellular vesicles (EVs) in obesity identifies large EVs as functional conveyors of Macrophage Migration Inhibitory Factor." *Mol Metab* 18:134-142. doi: 10.1016/j.molmet.2018.10.001.
- Aronow R Fau - Danon, D., A. Danon D Fau - Shahar, M. Shahar A Fau - Aronson, and M. Aronson. "ELECTRON MICROSCOPY OF IN VITRO ENDOCYTOSIS OF T2 PHAGE BY CELLS FROM RABBIT PERITONEAL EXUDATE." (0022-1007 (Print)).
- Arumugam, M., J. Raes, E. Pelletier, D. Le Paslier, T. Yamada, D. R. Mende, G. R. Fernandes, J. Tap, T. Bruls, J. M. Batto, M. Bertalan, N. Borruel, F. Casellas, L. Fernandez, L. Gautier, T. Hansen, M. Hattori, T. Hayashi, M. Kleerebezem, K. Kurokawa, M. Leclerc, F. Levenez, C. Manichanh, H. B. Nielsen, T. Nielsen, N. Pons, J. Poulain, J. Qin, T. Sicheritz-Ponten, S. Tims, D. Torrents, E. Ugarte, E. G. Zoetendal, J. Wang, F. Guarner, O. Pedersen, W. M. de Vos, S. Brunak, J. Dore, H. I. T. Consortium Meta, M. Antolin, F. Artiguenave, H. M. Blottiere, M. Almeida, C. Brechot, C. Cara, C. Chervaux, A. Cultrone, C. Delorme, G. Denariar, R. Dervyn, K. U. Foerstner, C. Friss, M. van de Guchte, E. Guedon, F. Haimet, W. Huber, J. van Hylckama-Vlieg, A. Jamet, C. Juste, G. Kaci, J. Knol, O. Lakhdari, S. Layec, K. Le Roux, E. Maguin, A. Merieux, R. Melo Minardi, C. M'Rini, J. Muller, R. Oozeer, J. Parkhill, P. Renault, M. Rescigno, N. Sanchez, S. Sunagawa, A. Torrejon, K. Turner, G. Vandemeulebrouck, E. Varela, Y. Winogradsky, G. Zeller, J. Weissenbach, S. D. Ehrlich, and P. Bork. 2011. "Enterotypes of the human gut microbiome." *Nature* 473 (7346):174-80. doi: 10.1038/nature09944.
- Babalova, E. G., K. T. Katsitadze, L. A. Sakvarelidze, NSh Imnaishvili, T. G. Sharashidze, V. A. Badashvili, G. P. Kiknadze, A. N. Meipariani, N. D. Gendzekhadze, E. V. Machavariani, K. L. Gogoberidze, E. I. Gozalov, and N. G. Dekanosidze. 1968. "[Preventive value of dried dysentery bacteriophage]." *Zh Mikrobiol Epidemiol Immunobiol* 45 (2):143-5.
- Bachurski, D., M. Schuldner, P. H. Nguyen, A. Malz, K. S. Reiners, P. C. Grenzi, F. Babatz, A. C. Schauss, H. P. Hansen, M. Hallek, and E. Pogge von Strandmann. 2019. "Extracellular vesicle measurements with nanoparticle tracking analysis - An accuracy and repeatability

- comparison between NanoSight NS300 and ZetaView." *J Extracell Vesicles* 8 (1):1596016. doi: 10.1080/20013078.2019.1596016.
- Baess, I. 1971. "Report on a pseudolysogenic mycobacterium and a review of the literature concerning pseudolysogeny." *Acta Pathol Microbiol Scand B Microbiol Immunol* 79 (3):428-34. doi: 10.1111/j.1699-0463.1971.tb00084.x.
- Bajorath, J., W. Saenger, and G. P. Pal. 1988. "Autolysis and inhibition of proteinase K, a subtilisin-related serine proteinase isolated from the fungus *Tritirachium album* Limber." *Biochim Biophys Acta* 954 (2):176-82. doi: 10.1016/0167-4838(88)90069-6.
- Barfoot, R., S. Denham, L. A. Gyure, J. G. Hall, S. M. Hobbs, L. E. Jackson, and D. Robertson. 1989. "Some properties of dendritic macrophages from peripheral lymph." *Immunology* 68 (2):233-239.
- Barr, J. J., R. Auro, M. Furlan, K. L. Whiteson, M. L. Erb, J. Pogliano, A. Stotland, R. Wolkowicz, A. S. Cutting, K. S. Doran, P. Salamon, M. Youle, and F. Rohwer. 2013. "Bacteriophage adhering to mucus provide a non-host-derived immunity." *Proc Natl Acad Sci U S A* 110 (26):10771-6. doi: 10.1073/pnas.1305923110.
- Bassignani, A., S. Plancade, M. Berland, M. Blein-Nicolas, A. Guillot, D. Chevret, C. Moritz, S. Huet, S. Rizkalla, K. Clement, J. Dore, O. Langella, and C. Juste. 2021. "Benefits of Iterative Searches of Large Databases to Interpret Large Human Gut Metaproteomic Data Sets." *J Proteome Res* 20 (3):1522-1534. doi: 10.1021/acs.jproteome.0c00669.
- Bebelman, M. P., M. J. Smit, D. M. Pegtel, and S. R. Baglio. 2018. "Biogenesis and function of extracellular vesicles in cancer." *Pharmacol Ther* 188:1-11. doi: 10.1016/j.pharmthera.2018.02.013.
- Berkowitz, L., B. M. Schultz, G. A. Salazar, C. Pardo-Roa, V. P. Sebastian, M. M. Alvarez-Lobos, and S. M. Bueno. 2018. "Impact of Cigarette Smoking on the Gastrointestinal Tract Inflammation: Opposing Effects in Crohn's Disease and Ulcerative Colitis." *Front Immunol* 9:74. doi: 10.3389/fimmu.2018.00074.
- Billaud, Maud, Quentin Lamy-Besnier, Julien Lossouarn, Elisabeth Moncaut, Moira B. Dion, Sylvain Moineau, Fatoumata Traoré, Emmanuelle Le Chatelier, Catherine Denis, Jordi Estelle, Caroline Achard, Olivier Zemb, and Marie-Agnès Petit. 2021. "Analysis of viromes and microbiomes from pig fecal samples reveals that phages and prophages rarely carry antibiotic resistance genes." *ISME Communications* 1 (1):55. doi: 10.1038/s43705-021-00054-8.
- Biller, S. J., L. D. McDaniel, M. Breitbart, E. Rogers, J. H. Paul, and S. W. Chisholm. 2017. "Membrane vesicles in sea water: heterogeneous DNA content and implications for viral abundance estimates." *ISME J* 11 (2):394-404. doi: 10.1038/ismej.2016.134.
- Biller, S. J., F. Schubotz, S. E. Roggensack, A. W. Thompson, R. E. Summons, and S. W. Chisholm. 2014. "Bacterial vesicles in marine ecosystems." *Science* 343 (6167):183-6. doi: 10.1126/science.1243457.
- Bitto, N. J., R. Chapman, S. Pidot, A. Costin, C. Lo, J. Choi, T. D'Cruze, E. C. Reynolds, S. G. Dashper, L. Turnbull, C. B. Whitchurch, T. P. Stinear, K. J. Stacey, and R. L. Ferrero. 2017. "Bacterial membrane vesicles transport their DNA cargo into host cells." *Sci Rep* 7 (1):7072. doi: 10.1038/s41598-017-07288-4.
- Bladen, H. A., and J. F. Waters. 1963. "Electron Microscopic Study of Some Strains of Bacteroides." *J Bacteriol* 86:1339-44. doi: 10.1128/jb.86.6.1339-1344.1963.

- Boccara, M., Y. Fedala, C. V. Bryan, M. Bailly-Bechet, C. Bowler, and A. C. Boccara. 2016. "Full-field interferometry for counting and differentiating aquatic biotic nanoparticles: from laboratory to Tara Oceans." *Biomed Opt Express* 7 (9):3736-3746. doi: 10.1364/BOE.7.003736.
- Bolger, A. M., M. Lohse, and B. Usadel. 2014. "Trimmomatic: a flexible trimmer for Illumina sequence data." *Bioinformatics* 30 (15):2114-20. doi: 10.1093/bioinformatics/btu170.
- Boling, L., D. A. Cuevas, J. A. Grasis, H. S. Kang, B. Knowles, K. Levi, H. Maughan, K. McNair, M. I. Rojas, S. E. Sanchez, C. Smurthwaite, and F. Rohwer. 2020. "Dietary prophage inducers and antimicrobials: toward landscaping the human gut microbiome." *Gut Microbes* 11 (4):721-734. doi: 10.1080/19490976.2019.1701353.
- Bomberger, J. M., D. P. Maceachran, B. A. Coutermarsh, S. Ye, G. A. O'Toole, and B. A. Stanton. 2009. "Long-distance delivery of bacterial virulence factors by *Pseudomonas aeruginosa* outer membrane vesicles." *PLoS Pathog* 5 (4):e1000382. doi: 10.1371/journal.ppat.1000382.
- Boudeau, J., A. L. Glasser, E. Masseret, B. Joly, and A. Darfeuille-Michaud. 1999. "Invasive ability of an *Escherichia coli* strain isolated from the ileal mucosa of a patient with Crohn's disease." *Infect Immun* 67 (9):4499-509. doi: 10.1128/IAI.67.9.4499-4509.1999.
- Breitbart, M., M. Haynes, S. Kelley, F. Angly, R. A. Edwards, B. Felts, J. M. Mahaffy, J. Mueller, J. Nulton, S. Rayhawk, B. Rodriguez-Brito, P. Salamon, and F. Rohwer. 2008. "Viral diversity and dynamics in an infant gut." *Res Microbiol* 159 (5):367-73. doi: 10.1016/j.resmic.2008.04.006.
- Breitbart, M., I. Hewson, B. Felts, J. M. Mahaffy, J. Nulton, P. Salamon, and F. Rohwer. 2003. "Metagenomic analyses of an uncultured viral community from human feces." *J Bacteriol* 185 (20):6220-3. doi: 10.1128/JB.185.20.6220-6223.2003.
- Brown, L., J. M. Wolf, R. Prados-Rosales, and A. Casadevall. 2015. "Through the wall: extracellular vesicles in Gram-positive bacteria, mycobacteria and fungi." *Nat Rev Microbiol* 13 (10):620-30. doi: 10.1038/nrmicro3480.
- Browne, H. P., S. C. Forster, B. O. Anonye, N. Kumar, B. A. Neville, M. D. Stares, D. Goulding, and T. D. Lawley. 2016. "Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation." *Nature* 533 (7604):543-546. doi: 10.1038/nature17645.
- Brussaard, C. P. 2004. "Optimization of procedures for counting viruses by flow cytometry." *Appl Environ Microbiol* 70 (3):1506-13. doi: 10.1128/AEM.70.3.1506-1513.2004.
- Brussaard, C. P., D. Marie, and G. Bratbak. 2000. "Flow cytometric detection of viruses." *J Virol Methods* 85 (1-2):175-82. doi: 10.1016/s0166-0934(99)00167-6.
- Brussow, H. 2016. "How stable is the human gut microbiota? And why this question matters." *Environ Microbiol* 18 (9):2779-83. doi: 10.1111/1462-2920.13473.
- Callanan, J., S. R. Stockdale, A. Shkoporov, L. A. Draper, R. P. Ross, and C. Hill. 2020. "Expansion of known ssRNA phage genomes: From tens to over a thousand." *Sci Adv* 6 (6):eaay5981. doi: 10.1126/sciadv.aay5981.
- Canas, M. A., R. Gimenez, M. J. Fabrega, L. Toloza, L. Baldoma, and J. Badia. 2016. "Outer Membrane Vesicles from the Probiotic *Escherichia coli* Nissle 1917 and the Commensal ECOR12 Enter Intestinal Epithelial Cells via Clathrin-Dependent Endocytosis and Elicit Differential Effects on DNA Damage." *PLoS One* 11 (8):e0160374. doi: 10.1371/journal.pone.0160374.
- Carvalho, F. A., N. Barnich, A. Sivignon, C. Darcha, C. H. Chan, C. P. Stanners, and A. Darfeuille-Michaud. 2009. "Crohn's disease adherent-invasive *Escherichia coli* colonize and induce strong gut inflammation in transgenic mice expressing human CEACAM." *J Exp Med* 206 (10):2179-89. doi: 10.1084/jem.20090741.

- Chandler, W. L., W. Yeung, and J. F. Tait. 2011. "A new microparticle size calibration standard for use in measuring smaller microparticles using a new flow cytometer." *J Thromb Haemost* 9 (6):1216-24. doi: 10.1111/j.1538-7836.2011.04283.x.
- Chapuy-Regaud, S., M. Dubois, C. Plisson-Chastang, T. Bonnefois, S. Lhomme, J. Bertrand-Michel, B. You, S. Simoneau, P. E. Gleizes, B. Flan, F. Abravanel, and J. Izopet. 2017. "Characterization of the lipid envelope of exosome encapsulated HEV particles protected from the immune response." *Biochimie* 141:70-79. doi: 10.1016/j.biochi.2017.05.003.
- Chen, C. C., L. Liu, F. Ma, C. W. Wong, X. E. Guo, J. V. Chacko, H. P. Farhoodi, S. X. Zhang, J. Zimak, A. Segaliny, M. Riazifar, V. Pham, M. A. Digman, E. J. Pone, and W. Zhao. 2016. "Elucidation of Exosome Migration across the Blood-Brain Barrier Model In Vitro." *Cell Mol Bioeng* 9 (4):509-529. doi: 10.1007/s12195-016-0458-3.
- Chen, J., N. Quiles-Puchalt, Y. N. Chiang, R. Bacigalupe, A. Fillol-Salom, M. S. J. Chee, J. R. Fitzgerald, and J. R. Penades. 2018. "Genome hypermobility by lateral transduction." *Science* 362 (6411):207-212. doi: 10.1126/science.aat5867.
- Choi, J., S. M. Kotay, and R. Goel. 2010. "Various physico-chemical stress factors cause prophage induction in *Nitrosospira multiformis* 25196--an ammonia oxidizing bacteria." *Water Res* 44 (15):4550-8. doi: 10.1016/j.watres.2010.04.040.
- Cimorelli, M., R. Nieuwland, Z. Varga, and E. van der Pol. 2021. "Standardized procedure to measure the size distribution of extracellular vesicles together with other particles in biofluids with microfluidic resistive pulse sensing." *PLoS One* 16 (4):e0249603. doi: 10.1371/journal.pone.0249603.
- Claesson, M. J., S. Cusack, O. O'Sullivan, R. Greene-Diniz, H. de Weerd, E. Flannery, J. R. Marchesi, D. Falush, T. Dinan, G. Fitzgerald, C. Stanton, D. van Sinderen, M. O'Connor, N. Harnedy, K. O'Connor, C. Henry, D. O'Mahony, A. P. Fitzgerald, F. Shanahan, C. Twomey, C. Hill, R. P. Ross, and P. W. O'Toole. 2011. "Composition, variability, and temporal stability of the intestinal microbiota of the elderly." *Proc Natl Acad Sci U S A* 108 Suppl 1:4586-91. doi: 10.1073/pnas.1000097107.
- Claesson, M. J., I. B. Jeffery, S. Conde, S. E. Power, E. M. O'Connor, S. Cusack, H. M. Harris, M. Coakley, B. Lakshminarayanan, O. O'Sullivan, G. F. Fitzgerald, J. Deane, M. O'Connor, N. Harnedy, K. O'Connor, D. O'Mahony, D. van Sinderen, M. Wallace, L. Brennan, C. Stanton, J. R. Marchesi, A. P. Fitzgerald, F. Shanahan, C. Hill, R. P. Ross, and P. W. O'Toole. 2012. "Gut microbiota composition correlates with diet and health in the elderly." *Nature* 488 (7410):178-84. doi: 10.1038/nature11319.
- Clemente, J. C., L. K. Ursell, L. W. Parfrey, and R. Knight. 2012. "The impact of the gut microbiota on human health: an integrative view." *Cell* 148 (6):1258-70. doi: 10.1016/j.cell.2012.01.035.
- Clooney, A. G., T. D. S. Sutton, A. N. Shkoporov, R. K. Holohan, K. M. Daly, O. O'Regan, F. J. Ryan, L. A. Draper, S. E. Plevy, R. P. Ross, and C. Hill. 2019. "Whole-Virome Analysis Sheds Light on Viral Dark Matter in Inflammatory Bowel Disease." *Cell Host Microbe* 26 (6):764-778 e5. doi: 10.1016/j.chom.2019.10.009.
- Codemo, M., S. Muschiol, F. Iovino, P. Nannapaneni, L. Plant, S. N. Wai, and B. Henriques-Normark. 2018. "Immunomodulatory Effects of Pneumococcal Extracellular Vesicles on Cellular and Humoral Host Defenses." *mBio* 9 (2). doi: 10.1128/mBio.00559-18.
- Cornuault, J. K., E. Moncaut, V. Loux, A. Mathieu, H. Sokol, M. A. Petit, and M. De Paepe. 2020. "The enemy from within: a prophage of *Roseburia intestinalis* systematically turns lytic in the

- mouse gut, driving bacterial adaptation by CRISPR spacer acquisition." *ISME J* 14 (3):771-787. doi: 10.1038/s41396-019-0566-x.
- Cornuault, J. K., M. A. Petit, M. Mariadassou, L. Benevides, E. Moncaut, P. Langella, H. Sokol, and M. De Paepe. 2018. "Phages infecting *Faecalibacterium prausnitzii* belong to novel viral genera that help to decipher intestinal viromes." *Microbiome* 6 (1):65. doi: 10.1186/s40168-018-0452-1.
- Cosa, G., K. S. Focsaneanu, J. R. McLean, J. P. McNamee, and J. C. Scaiano. 2001. "Photophysical properties of fluorescent DNA-dyes bound to single- and double-stranded DNA in aqueous buffered solution." *Photochem Photobiol* 73 (6):585-99. doi: 10.1562/0031-8655(2001)073<0585:PPOFDD>2.0.CO;2.
- Coumans, F. A., E. van der Pol, A. N. Boing, N. Hajji, G. Sturk, T. G. van Leeuwen, and R. Nieuwland. 2014. "Reproducible extracellular vesicle size and concentration determination with tunable resistive pulse sensing." *J Extracell Vesicles* 3:25922. doi: 10.3402/jev.v3.25922.
- Cresci, G. A., and E. Bawden. 2015. "Gut Microbiome: What We Do and Don't Know." *Nutr Clin Pract* 30 (6):734-46. doi: 10.1177/0884533615609899.
- Crost, E. H., E. H. Ajandouz, C. Villard, P. A. Geraert, A. Puigserver, and M. Fons. 2011. "Ruminococcin C, a new anti-*Clostridium perfringens* bacteriocin produced in the gut by the commensal bacterium *Ruminococcus gnavus* E1." *Biochimie* 93 (9):1487-94. doi: 10.1016/j.biochi.2011.05.001.
- D'Haens, G., M. Ferrante, S. Vermeire, F. Baert, M. Noman, L. Moortgat, P. Geens, D. Iwens, I. Aerden, G. Van Assche, G. Van Olmen, and P. Rutgeerts. 2012. "Fecal calprotectin is a surrogate marker for endoscopic lesions in inflammatory bowel disease." *Inflamm Bowel Dis* 18 (12):2218-24. doi: 10.1002/ibd.22917.
- D'Herelle, F. 1929. "Studies Upon Asiatic Cholera." *Yale J Biol Med* 1 (4):195-219.
- d'Herelle, Félix. 1917. "On an invisible microbe antagonistic to dysentery bacilli." *CR Acad Sci* 165:373-375.
- Danielsen, E. M. 2015. "Probing endocytosis from the enterocyte brush border using fluorescent lipophilic dyes: lipid sorting at the apical cell surface." *Histochem Cell Biol* 143 (5):545-56. doi: 10.1007/s00418-014-1302-2.
- Davies, A. A., N. M. Wigglesworth, D. Allan, R. J. Owens, and M. J. Crumpton. 1984. "Nonidet P-40 extraction of lymphocyte plasma membrane. Characterization of the insoluble residue." *Biochem J* 219 (1):301-8. doi: 10.1042/bj2190301.
- De Filippo, C., D. Cavalieri, M. Di Paola, M. Ramazzotti, J. B. Poullet, S. Massart, S. Collini, G. Pieraccini, and P. Lionetti. 2010. "Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa." *Proc Natl Acad Sci U S A* 107 (33):14691-6. doi: 10.1073/pnas.1005963107.
- De Sordi, L., M. Lourenco, and L. Debarbieux. 2019. "The Battle Within: Interactions of Bacteriophages and Bacteria in the Gastrointestinal Tract." *Cell Host Microbe* 25 (2):210-218. doi: 10.1016/j.chom.2019.01.018.
- De Vadder, F., P. Kovatcheva-Datchary, D. Goncalves, J. Vinera, C. Zitoun, A. Duchamp, F. Backhed, and G. Mithieux. 2014. "Microbiota-generated metabolites promote metabolic benefits via gut-brain neural circuits." *Cell* 156 (1-2):84-96. doi: 10.1016/j.cell.2013.12.016.

- Dehghani, M., R. K. Montange, M. W. Olszowy, and D. Pollard. 2021. "An Emerging Fluorescence-Based Technique for Quantification and Protein Profiling of Extracellular Vesicles." *SLAS Technol* 26 (2):189-199. doi: 10.1177/2472630320970458.
- Devoto, A. E., J. M. Santini, M. R. Olm, K. Anantharaman, P. Munk, J. Tung, E. A. Archie, P. J. Turnbaugh, K. D. Seed, R. Blekhman, F. M. Aarestrup, B. C. Thomas, and J. F. Banfield. 2019. "Megaphages infect *Prevotella* and variants are widespread in gut microbiomes." *Nat Microbiol* 4 (4):693-700. doi: 10.1038/s41564-018-0338-9.
- Diard, M., E. Bakkeren, J. K. Cornuault, K. Moor, A. Hausmann, M. E. Sellin, C. Loverdo, A. Aertsen, M. Ackermann, M. De Paepe, E. Slack, and W. D. Hardt. 2017. "Inflammation boosts bacteriophage transfer between *Salmonella* spp." *Science* 355 (6330):1211-1215. doi: 10.1126/science.aaf8451.
- Diaz-Garrido, N., J. Badia, and L. Baldoma. 2021. "Microbiota-derived extracellular vesicles in interkingdom communication in the gut." *J Extracell Vesicles* 10 (13):e12161. doi: 10.1002/jev2.12161.
- Diaz-Garrido, N., C. Cordero, Y. Olivo-Martinez, J. Badia, and L. Baldoma. 2021. "Cell-to-Cell Communication by Host-Released Extracellular Vesicles in the Gut: Implications in Health and Disease." *Int J Mol Sci* 22 (4). doi: 10.3390/ijms22042213.
- Dicksved, J., J. Halfvarson, M. Rosenquist, G. Jarnerot, C. Tysk, J. Apajalahti, L. Engstrand, and J. K. Jansson. 2008. "Molecular analysis of the gut microbiota of identical twins with Crohn's disease." *ISME J* 2 (7):716-27. doi: 10.1038/ismej.2008.37.
- Dion, M. B., F. Oechslin, and S. Moineau. 2020. "Phage diversity, genomics and phylogeny." *Nat Rev Microbiol* 18 (3):125-138. doi: 10.1038/s41579-019-0311-5.
- Dixon, L. J., A. Kabi, K. P. Nickerson, and C. McDonald. 2015. "Combinatorial effects of diet and genetics on inflammatory bowel disease pathogenesis." *Inflamm Bowel Dis* 21 (4):912-22. doi: 10.1097/MIB.0000000000000289.
- Dolan, K. T., and E. B. Chang. 2017. "Diet, gut microbes, and the pathogenesis of inflammatory bowel diseases." *Mol Nutr Food Res* 61 (1). doi: 10.1002/mnfr.201600129.
- Donaldson, G. P., M. S. Ladinsky, K. B. Yu, J. G. Sanders, B. B. Yoo, W. C. Chou, M. E. Conner, A. M. Earl, R. Knight, P. J. Bjorkman, and S. K. Mazmanian. 2018. "Gut microbiota utilize immunoglobulin A for mucosal colonization." *Science* 360 (6390):795-800. doi: 10.1126/science.aag0926.
- Dorward, D. W., and C. F. Garon. 1990. "DNA Is Packaged within Membrane-Derived Vesicles of Gram-Negative but Not Gram-Positive Bacteria." *Appl Environ Microbiol* 56 (6):1960-2. doi: 10.1128/aem.56.6.1960-1962.1990.
- Doyle, L. M., and M. Z. Wang. 2019. "Overview of Extracellular Vesicles, Their Origin, Composition, Purpose, and Methods for Exosome Isolation and Analysis." *Cells* 8 (7). doi: 10.3390/cells8070727.
- Dragovic, R. A., C. Gardiner, A. S. Brooks, D. S. Tannetta, D. J. Ferguson, P. Hole, B. Carr, C. W. Redman, A. L. Harris, P. J. Dobson, P. Harrison, and I. L. Sargent. 2011. "Sizing and phenotyping of cellular vesicles using Nanoparticle Tracking Analysis." *Nanomedicine* 7 (6):780-8. doi: 10.1016/j.nano.2011.04.003.
- Dugat-Bony, E., J. Lossouarn, M. De Paepe, A. S. Sarthou, Y. Fedala, M. A. Petit, and S. Chaillou. 2020. "Viral metagenomic analysis of the cheese surface: A comparative study of rapid procedures for extracting viral particles." *Food Microbiol* 85:103278. doi: 10.1016/j.fm.2019.103278.

- Duncan, S. H., G. L. Hold, A. Barcenilla, C. S. Stewart, and H. J. Flint. 2002. "Roseburia intestinalis sp. nov., a novel saccharolytic, butyrate-producing bacterium from human faeces." *Int J Syst Evol Microbiol* 52 (Pt 5):1615-1620. doi: 10.1099/00207713-52-5-1615.
- Dutilh, B. E., N. Cassman, K. McNair, S. E. Sanchez, G. G. Silva, L. Boling, J. J. Barr, D. R. Speth, V. Seguritan, R. K. Aziz, B. Felts, E. A. Dinsdale, J. L. Mokili, and R. A. Edwards. 2014. "A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes." *Nat Commun* 5:4498. doi: 10.1038/ncomms5498.
- Eckburg, P. B., E. M. Bik, C. N. Bernstein, E. Purdom, L. Dethlefsen, M. Sargent, S. R. Gill, K. E. Nelson, and D. A. Relman. 2005. "Diversity of the human intestinal microbial flora." *Science* 308 (5728):1635-8. doi: 10.1126/science.1110591.
- Einstein, A. 1905. "Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen [AdP 17, 549 (1905)]." *Annalen der Physik* 14 (S1):182-193. doi: <https://doi.org/10.1002/andp.200590005>.
- Elmi, A., F. Nasher, H. Jagatia, O. Gundogdu, M. Bajaj-Elliott, B. Wren, and N. Dorrell. 2016. "Campylobacter jejuni outer membrane vesicle-associated proteolytic activity promotes bacterial invasion by mediating cleavage of intestinal epithelial cell E-cadherin and occludin." *Cell Microbiol* 18 (4):561-72. doi: 10.1111/cmi.12534.
- Escrevente, C., S. Keller, P. Altevogt, and J. Costa. 2011. "Interaction and uptake of exosomes by ovarian cancer cells." *BMC Cancer* 11:108. doi: 10.1186/1471-2407-11-108.
- Fabrega, M. J., A. Rodriguez-Nogales, J. Garrido-Mesa, F. Algieri, J. Badia, R. Gimenez, J. Galvez, and L. Baldoma. 2017. "Intestinal Anti-inflammatory Effects of Outer Membrane Vesicles from Escherichia coli Nissle 1917 in DSS-Experimental Colitis in Mice." *Front Microbiol* 8:1274. doi: 10.3389/fmicb.2017.01274.
- Faith, J. J., J. L. Guruge, M. Charbonneau, S. Subramanian, H. Seedorf, A. L. Goodman, J. C. Clemente, R. Knight, A. C. Heath, R. L. Leibel, M. Rosenbaum, and J. I. Gordon. 2013. "The long-term stability of the human gut microbiota." *Science* 341 (6141):1237439. doi: 10.1126/science.1237439.
- Falke S., Betzel C. 2019. "Dynamic Light Scattering (DLS)." In *Radiation in Bioanalysis*, edited by Tavares P. Pereira A., Limão-Vieira P.: Springer, Cham.
- Falony, G., M. Joossens, S. Vieira-Silva, J. Wang, Y. Darzi, K. Faust, A. Kurilshikov, M. J. Bonder, M. Valles-Colomer, D. Vandeputte, R. Y. Tito, S. Chaffron, L. Rymenans, C. Verspecht, L. De Sutter, G. Lima-Mendez, K. D'Hoe, K. Jonckheere, D. Homola, R. Garcia, E. F. Tigchelaar, L. Eeckhaut, J. Fu, L. Henckaerts, A. Zhernakova, C. Wijmenga, and J. Raes. 2016. "Population-level analysis of gut microbiome variation." *Science* 352 (6285):560-4. doi: 10.1126/science.aad3503.
- Farooqi, A. A., N. N. Desai, M. Z. Qureshi, D. R. N. Librelotto, M. L. Gasparri, A. Bishayee, S. M. Nabavi, V. Curti, and M. Daglia. 2018. "Exosome biogenesis, bioactivities and functions as new delivery systems of natural compounds." *Biotechnol Adv* 36 (1):328-334. doi: 10.1016/j.biotechadv.2017.12.010.
- Fernandes, M. A., S. G. Verstraete, T. G. Phan, X. Deng, E. Stekol, B. LaMere, S. V. Lynch, M. B. Heyman, and E. Delwart. 2019. "Enteric Virome and Bacterial Microbiota in Children With Ulcerative Colitis and Crohn Disease." *J Pediatr Gastroenterol Nutr* 68 (1):30-36. doi: 10.1097/MPG.0000000000002140.
- Ferris, M. M., C. L. Stoffel, T. T. Maurer, and K. L. Rowlen. 2002. "Quantitative intercomparison of transmission electron microscopy, flow cytometry, and epifluorescence microscopy for nanometric particle analysis." *Anal Biochem* 304 (2):249-56. doi: 10.1006/abio.2002.5616.

- Fiers, W., R. Contreras, F. Duerinck, G. Haegeman, D. Iserentant, J. Merregaert, W. Min Jou, F. Molemans, A. Raeymaekers, A. Van den Berghe, G. Volckaert, and M. Ysebaert. 1976. "Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene." *Nature* 260 (5551):500-7. doi: 10.1038/260500a0.
- Filip, R. 2021. "An Update on the Role of Extracellular Vesicles in the Pathogenesis of Necrotizing Enterocolitis and Inflammatory Bowel Diseases." *Cells* 10 (11). doi: 10.3390/cells10113202.
- Filipe, V., A. Hawe, and W. Jiskoot. 2010. "Critical evaluation of Nanoparticle Tracking Analysis (NTA) by NanoSight for the measurement of nanoparticles and protein aggregates." *Pharm Res* 27 (5):796-810. doi: 10.1007/s11095-010-0073-2.
- Fillol-Salom, A., R. Bacigalupe, S. Humphrey, Y. N. Chiang, J. Chen, and J. R. Penades. 2021. "Lateral transduction is inherent to the life cycle of the archetypical Salmonella phage P22." *Nat Commun* 12 (1):6510. doi: 10.1038/s41467-021-26520-4.
- Flesch, F. M., J. F. Brouwers, P. F. Nievelstein, A. J. Verkleij, L. M. van Golde, B. Colenbrander, and B. M. Gadella. 2001. "Bicarbonate stimulated phospholipid scrambling induces cholesterol redistribution and enables cholesterol depletion in the sperm plasma membrane." *J Cell Sci* 114 (Pt 19):3543-55. doi: 10.1242/jcs.114.19.3543.
- Flores, G. E., J. G. Caporaso, J. B. Henley, J. R. Rideout, D. Domogala, J. Chase, J. W. Leff, Y. Vazquez-Baeza, A. Gonzalez, R. Knight, R. R. Dunn, and N. Fierer. 2014. "Temporal variability is a personalized feature of the human microbiome." *Genome Biol* 15 (12):531. doi: 10.1186/s13059-014-0531-y.
- Fogg, P. C., H. E. Allison, J. R. Saunders, and A. J. McCarthy. 2010. "Bacteriophage lambda: a paradigm revisited." *J Virol* 84 (13):6876-9. doi: 10.1128/JVI.02177-09.
- Forterre, P., N. Soler, M. Krupovic, E. Marguet, and H. W. Ackermann. 2013. "Fake virus particles generated by fluorescence microscopy." *Trends Microbiol* 21 (1):1-5. doi: 10.1016/j.tim.2012.10.005.
- Frank, D. N., A. L. St Amand, R. A. Feldman, E. C. Boedeker, N. Harpaz, and N. R. Pace. 2007. "Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases." *Proc Natl Acad Sci U S A* 104 (34):13780-5. doi: 10.1073/pnas.0706625104.
- Frost, G., M. L. Sleeth, M. Sahuri-Arisoylu, B. Lizarbe, S. Cerdan, L. Brody, J. Anastasovska, S. Ghourab, M. Hankir, S. Zhang, D. Carling, J. R. Swann, G. Gibson, A. Viardot, D. Morrison, E. Louise Thomas, and J. D. Bell. 2014. "The short-chain fatty acid acetate reduces appetite via a central homeostatic mechanism." *Nat Commun* 5:3611. doi: 10.1038/ncomms4611.
- Fruhbeis, C., D. Frohlich, W. P. Kuo, and E. M. Kramer-Albers. 2013. "Extracellular vesicles as mediators of neuron-glia communication." *Front Cell Neurosci* 7:182. doi: 10.3389/fncel.2013.00182.
- Fu, L., B. Niu, Z. Zhu, S. Wu, and W. Li. 2012. "CD-HIT: accelerated for clustering the next-generation sequencing data." *Bioinformatics* 28 (23):3150-2. doi: 10.1093/bioinformatics/bts565.
- Furuta, N., H. Takeuchi, and A. Amano. 2009. "Entry of Porphyromonas gingivalis outer membrane vesicles into epithelial cells causes cellular functional impairment." *Infect Immun* 77 (11):4761-70. doi: 10.1128/IAI.00841-09.
- Galiez, C., M. Siebert, F. Enault, J. Vincent, and J. Soding. 2017. "WIsH: who is the host? Predicting prokaryotic hosts from metagenomic phage contigs." *Bioinformatics* 33 (19):3113-3114. doi: 10.1093/bioinformatics/btx383.

- Gardiner, C., Y. J. Ferreira, R. A. Dragovic, C. W. Redman, and I. L. Sargent. 2013. "Extracellular vesicle sizing and enumeration by nanoparticle tracking analysis." *J Extracell Vesicles* 2. doi: 10.3402/jev.v2i0.19671.
- Garmaeva, S., A. Gulyaeva, T. Sinha, A. N. Shkoporov, A. G. Clooney, S. R. Stockdale, J. E. Spreckels, T. D. S. Sutton, L. A. Draper, B. E. Dutilh, C. Wijmenga, A. Kurilshikov, J. Fu, C. Hill, and A. Zhernakova. 2021. "Stability of the human gut virome and effect of gluten-free diet." *Cell Rep* 35 (7):109132. doi: 10.1016/j.celrep.2021.109132.
- Gibson, M. K., K. J. Forsberg, and G. Dantas. 2015. "Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology." *ISME J* 9 (1):207-16. doi: 10.1038/ismej.2014.106.
- Glasser, A. L., J. Boudeau, N. Barnich, M. H. Perruchot, J. F. Colombel, and A. Darfeuille-Michaud. 2001. "Adherent invasive *Escherichia coli* strains from patients with Crohn's disease survive and replicate within macrophages without inducing host cell death." *Infect Immun* 69 (9):5529-37. doi: 10.1128/IAI.69.9.5529-5537.2001.
- Gogokhia, L., K. Buhrke, R. Bell, B. Hoffman, D. G. Brown, C. Hanke-Gogokhia, N. J. Ajami, M. C. Wong, A. Ghazaryan, J. F. Valentine, N. Porter, E. Martens, R. O'Connell, V. Jacob, E. Scherl, C. Crawford, W. Z. Stephens, S. R. Casjens, R. S. Longman, and J. L. Round. 2019. "Expansion of Bacteriophages Is Linked to Aggravated Intestinal Inflammation and Colitis." *Cell Host Microbe* 25 (2):285-299 e8. doi: 10.1016/j.chom.2019.01.008.
- Guerin, E., A. Shkoporov, S. R. Stockdale, A. G. Clooney, F. J. Ryan, T. D. S. Sutton, L. A. Draper, E. Gonzalez-Tortuero, R. P. Ross, and C. Hill. 2018. "Biology and Taxonomy of crAss-like Bacteriophages, the Most Abundant Virus in the Human Gut." *Cell Host Microbe* 24 (5):653-664 e6. doi: 10.1016/j.chom.2018.10.002.
- Hatzioanou, D., M. J. Mayer, S. H. Duncan, H. J. Flint, and A. Narbad. 2013. "A representative of the dominant human colonic Firmicutes, *Roseburia faecis* M72/1, forms a novel bacteriocin-like substance." *Anaerobe* 23:5-8. doi: 10.1016/j.anaerobe.2013.07.006.
- Hayes, S., J. Mahony, A. Nauta, and D. van Sinderen. 2017. "Metagenomic Approaches to Assess Bacteriophages in Various Environmental Niches." *Viruses* 9 (6). doi: 10.3390/v9060127.
- Heider, S., and C. Metzner. 2014. "Quantitative real-time single particle analysis of virions." *Virology* 462-463:199-206. doi: 10.1016/j.virol.2014.06.005.
- Hennes, K. P., C. A. Suttle, and A. M. Chan. 1995. "Fluorescently Labeled Virus Probes Show that Natural Virus Populations Can Control the Structure of Marine Microbial Communities." *Appl Environ Microbiol* 61 (10):3623-7. doi: 10.1128/aem.61.10.3623-3627.1995.
- Hessvik, N. P., A. Overbye, A. Brech, M. L. Torgersen, I. S. Jakobsen, K. Sandvig, and A. Llorente. 2016. "PIKfyve inhibition increases exosome release and induces secretory autophagy." *Cell Mol Life Sci* 73 (24):4717-4737. doi: 10.1007/s00018-016-2309-8.
- Hirano, J., T. Yoshida, T. Sugiyama, N. Koide, I. Mori, and T. Yokochi. 2003. "The effect of *Lactobacillus rhamnosus* on enterohemorrhagic *Escherichia coli* infection of human intestinal cells in vitro." *Microbiol Immunol* 47 (6):405-9. doi: 10.1111/j.1348-0421.2003.tb03377.x.
- Hodyra-Stefaniak, K., P. Miernikiewicz, J. Drapala, M. Drab, E. Jonczyk-Matysiak, D. Lecion, Z. Kazmierczak, W. Beta, J. Majewska, M. Harhala, B. Bubak, A. Klopot, A. Gorski, and K. Dabrowska. 2015. "Mammalian Host-Versus-Phage immune response determines phage fate in vivo." *Sci Rep* 5:14802. doi: 10.1038/srep14802.

- Holmfeldt, K., D. Odic, M. B. Sullivan, M. Middelboe, and L. Riemann. 2012. "Cultivated single-stranded DNA phages that infect marine Bacteroidetes prove difficult to detect with DNA-binding stains." *Appl Environ Microbiol* 78 (3):892-4. doi: 10.1128/AEM.06580-11.
- Holzwarth, G., D. G. Gordon, J. E. McGinness, B. P. Dorman, and M. F. Maestre. 1974. "Mie scattering contributions to the optical density and circular dichroism of T2 bacteriophage." *Biochemistry* 13 (1):126-32. doi: 10.1021/bi00698a020.
- Hooper, L. V., J. Xu, P. G. Falk, T. Midtvedt, and J. I. Gordon. 1999. "A molecular sensor that allows a gut commensal to control its nutrient foundation in a competitive ecosystem." *Proc Natl Acad Sci U S A* 96 (17):9833-8. doi: 10.1073/pnas.96.17.9833.
- Hoyles, L., A. L. McCartney, H. Neve, G. R. Gibson, J. D. Sanderson, K. J. Heller, and D. van Sinderen. 2014. "Characterization of virus-like particles associated with the human faecal and caecal microbiota." *Res Microbiol* 165 (10):803-12. doi: 10.1016/j.resmic.2014.10.006.
- Huang, A. S., and D. Baltimore. 1970. "Defective viral particles and viral disease processes." *Nature* 226 (5243):325-7. doi: 10.1038/226325a0.
- Hulst, H. C. van de. 1957. *Light Scattering by Small Metal Particles*. NYC.
- Humphrey, S. B., T. B. Stanton, and N. S. Jensen. 1995. "Mitomycin C induction of bacteriophages from *Serpulina hyodysenteriae* and *Serpulina innocens*." *FEMS Microbiol Lett* 134 (1):97-101. doi: 10.1111/j.1574-6968.1995.tb07921.x.
- Jafari, Behnoush, Ramazan Ali Khavari Nejad, Farzam Vaziri, and Seyed Davar Siadat. 2019. "Evaluation of the effects of extracellular vesicles derived from *Faecalibacterium prausnitzii* on lung cancer cell line." *Biologia* 74 (7):889-898. doi: 10.2478/s11756-019-00229-8.
- Jakobsson, H. E., T. R. Abrahamsson, M. C. Jenmalm, K. Harris, C. Quince, C. Jernberg, B. Bjorksten, L. Engstrand, and A. F. Andersson. 2014. "Decreased gut microbiota diversity, delayed Bacteroidetes colonisation and reduced Th1 responses in infants delivered by caesarean section." *Gut* 63 (4):559-66. doi: 10.1136/gutjnl-2012-303249.
- Janockova, J., L. Slovinska, D. Harvanova, T. Spakova, and J. Rosocha. 2021. "New therapeutic approaches of mesenchymal stem cells-derived exosomes." *J Biomed Sci* 28 (1):39. doi: 10.1186/s12929-021-00736-4.
- Jensen, O. A., J. U. Prause, and H. Laursen. 1981. "Shrinkage in preparatory steps for SEM. A study on rabbit corneal endothelium." *Albrecht Von Graefes Arch Klin Exp Ophthalmol* 215 (4):233-42. doi: 10.1007/BF00407662.
- Johnstone, R. M., M. Adam, J. R. Hammond, L. Orr, and C. Turbide. 1987. "Vesicle formation during reticulocyte maturation. Association of plasma membrane activities with released vesicles (exosomes)." *J Biol Chem* 262 (19):9412-20.
- Jones, E. J., C. Booth, S. Fonseca, A. Parker, K. Cross, A. Miquel-Clopes, I. Hautefort, U. Mayer, T. Wileman, R. Stentz, and S. R. Carding. 2020. "The Uptake, Trafficking, and Biodistribution of *Bacteroides thetaiotaomicron* Generated Outer Membrane Vesicles." *Front Microbiol* 11:57. doi: 10.3389/fmicb.2020.00057.
- Jovel, J., J. Patterson, W. Wang, N. Hotte, S. O'Keefe, T. Mitchel, T. Perry, D. Kao, A. L. Mason, K. L. Madsen, and G. K. Wong. 2016. "Characterization of the Gut Microbiome Using 16S or Shotgun Metagenomics." *Front Microbiol* 7:459. doi: 10.3389/fmicb.2016.00459.
- Jurczak-Kurek, A., T. Gasior, B. Nejman-Falenczyk, S. Bloch, A. Dydecka, G. Topka, A. Necel, M. Jakubowska-Deredas, M. Narajczyk, M. Richert, A. Mieszkowska, B. Wrobel, G. Wegrzyn, and A. Wegrzyn. 2016. "Biodiversity of bacteriophages: morphological and biological properties

- of a large group of phages isolated from urban sewage." *Sci Rep* 6:34338. doi: 10.1038/srep34338.
- Kaletta, J., C. Pickl, C. Griebler, A. Klingl, R. Kurmayer, and L. Deng. 2020. "A rigorous assessment and comparison of enumeration methods for environmental viruses." *Sci Rep* 10 (1):18625. doi: 10.1038/s41598-020-75490-y.
- Kameli, N., R. Borman, C. Lopez-Iglesias, P. Savelkoul, and F. R. M. Stassen. 2021. "Characterization of Feces-Derived Bacterial Membrane Vesicles and the Impact of Their Origin on the Inflammatory Response." *Front Cell Infect Microbiol* 11:667987. doi: 10.3389/fcimb.2021.667987.
- Kenzaka, Takehiko, Katsuji Tani, and Masao Nasu. 2010. "High-frequency phage-mediated gene transfer in freshwater environments determined at single-cell level." *The ISME Journal* 4 (5):648-659. doi: 10.1038/ismej.2009.145.
- Kerr, J. F., A. H. Wyllie, and A. R. Currie. 1972. "Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics." *Br J Cancer* 26 (4):239-57. doi: 10.1038/bjc.1972.33.
- Kestens, V., V. Bozatzidis, P. J. De Temmerman, Y. Ramaye, and G. Roebben. 2017. "Validation of a particle tracking analysis method for the size determination of nano- and microparticles." *J Nanopart Res* 19 (8):271. doi: 10.1007/s11051-017-3966-8.
- Kieft, K., Z. Zhou, and K. Anantharaman. 2020. "VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences." *Microbiome* 8 (1):90. doi: 10.1186/s40168-020-00867-0.
- Kim, M. S., and J. W. Bae. 2016. "Spatial disturbances in altered mucosal and luminal gut viromes of diet-induced obese mice." *Environ Microbiol* 18 (5):1498-510. doi: 10.1111/1462-2920.13182.
- Kim, Y., S. H. Kim, K. Y. Whang, Y. J. Kim, and S. Oh. 2008. "Inhibition of Escherichia coli O157:H7 attachment by interactions between lactic acid bacteria and intestinal epithelial cells." *J Microbiol Biotechnol* 18 (7):1278-85.
- Kivela, H. M., R. H. Mannisto, N. Kalkkinen, and D. H. Bamford. 1999. "Purification and protein composition of PM2, the first lipid-containing bacterial virus to be isolated." *Virology* 262 (2):364-74. doi: 10.1006/viro.1999.9838.
- Kleiner, M., B. Bushnell, K. E. Sanderson, L. V. Hooper, and B. A. Duerkop. 2020. "Transductomics: sequencing-based detection and analysis of transduced DNA in pure cultures and microbial communities." *Microbiome* 8 (1):158. doi: 10.1186/s40168-020-00935-5.
- Klopot, A., A. Zakrzewska, D. Lecion, J. M. Majewska, M. A. Harhala, K. Lahutta, Z. Kazmierczak, L. Laczanski, M. Klak, and K. Dabrowska. 2017. "Real-Time qPCR as a Method for Detection of Antibody-Neutralized Phage Particles." *Front Microbiol* 8:2170. doi: 10.3389/fmicb.2017.02170.
- Koenig, J. E., A. Spor, N. Scalfone, A. D. Fricker, J. Stombaugh, R. Knight, L. T. Angenent, and R. E. Ley. 2011. "Succession of microbial consortia in the developing infant gut microbiome." *Proc Natl Acad Sci U S A* 108 Suppl 1:4578-85. doi: 10.1073/pnas.1000081107.
- Kowal, J., G. Arras, M. Colombo, M. Jouve, J. P. Morath, B. Primdal-Bengtson, F. Dingli, D. Loew, M. Tkach, and C. Thery. 2016. "Proteomic comparison defines novel markers to characterize heterogeneous populations of extracellular vesicle subtypes." *Proceedings of the National Academy of Sciences of the United States of America* 113 (8):E968-E977. doi: 10.1073/pnas.1521230113.

- Kutter, E., D. De Vos, G. Gvasalia, Z. Alavidze, L. Gogokhia, S. Kuhl, and S. T. Abedon. 2010. "Phage therapy in clinical practice: treatment of human infections." *Curr Pharm Biotechnol* 11 (1):69-86. doi: 10.2174/138920110790725401.
- Labarde, A., L. Jakutyte, C. Billaudeau, B. Fauler, M. Lopez-Sanz, P. Ponien, E. Jacquet, T. Mielke, S. Ayora, R. Carballido-Lopez, and P. Tavares. 2021. "Temporal compartmentalization of viral infection in bacterial cells." *Proc Natl Acad Sci U S A* 118 (28). doi: 10.1073/pnas.2018297118.
- Laforest-Lapointe, I., and M. C. Arrieta. 2018. "Microbial Eukaryotes: a Missing Link in Gut Microbiome Studies." *mSystems* 3 (2). doi: 10.1128/mSystems.00201-17.
- Lamas, B., M. L. Richard, V. Leducq, H. P. Pham, M. L. Michel, G. Da Costa, C. Bridonneau, S. Jegou, T. W. Hoffmann, J. M. Natividad, L. Brot, S. Taleb, A. Couturier-Maillard, I. Nion-Larmurier, F. Merabtene, P. Seksik, A. Bourrier, J. Cosnes, B. Ryffel, L. Beaugerie, J. M. Launay, P. Langella, R. J. Xavier, and H. Sokol. 2016. "CARD9 impacts colitis by altering gut microbiota metabolism of tryptophan into aryl hydrocarbon receptor ligands." *Nat Med* 22 (6):598-605. doi: 10.1038/nm.4102.
- Lambert, J. E., J. P. Myslicki, M. R. Bomhof, D. D. Belke, J. Shearer, and R. A. Reimer. 2015. "Exercise training modifies gut microbiota in normal and diabetic mice." *Appl Physiol Nutr Metab* 40 (7):749-52. doi: 10.1139/apnm-2014-0452.
- Lane, R. E., D. Korbie, M. M. Hill, and M. Trau. 2018. "Extracellular vesicles as circulating cancer biomarkers: opportunities and challenges." *Clin Transl Med* 7 (1):14. doi: 10.1186/s40169-018-0192-7.
- Langlet, J., L. Kaas, D. Croucher, and J. Hewitt. 2018. "Effect of the Shellfish Proteinase K Digestion Method on Norovirus Capsid Integrity." *Food Environ Virol* 10 (2):151-158. doi: 10.1007/s12560-018-9336-6.
- Le Huerou-Luron, I., S. Blat, and G. Boudry. 2010. "Breast- v. formula-feeding: impacts on the digestive tract and immediate and long-term health effects." *Nutr Res Rev* 23 (1):23-36. doi: 10.1017/S0954422410000065.
- Lee, E. Y., D. Y. Choi, D. K. Kim, J. W. Kim, J. O. Park, S. Kim, S. H. Kim, D. M. Desiderio, Y. K. Kim, K. P. Kim, and Y. S. Gho. 2009. "Gram-positive bacteria produce membrane vesicles: proteomics-based characterization of Staphylococcus aureus-derived membrane vesicles." *Proteomics* 9 (24):5425-36. doi: 10.1002/pmic.200900338.
- Lee, J., O. Y. Kim, and Y. S. Gho. 2016. "Proteomic profiling of Gram-negative bacterial outer membrane vesicles: Current perspectives." *Proteomics Clin Appl* 10 (9-10):897-909. doi: 10.1002/prca.201600032.
- Lee, S. A., J. Y. Lim, B. S. Kim, S. J. Cho, N. Y. Kim, O. B. Kim, and Y. Kim. 2015. "Comparison of the gut microbiota profile in breast-fed and formula-fed Korean infants using pyrosequencing." *Nutr Res Pract* 9 (3):242-8. doi: 10.4162/nrp.2015.9.3.242.
- Lee, Yuan-Kun, Kim-Yoong Puong, Arthur C. Ouwehand, and Seppo Salminen. 2003. "Displacement of bacterial pathogens from mucus and Caco-2 cell surface by lactobacilli." *Journal of Medical Microbiology* 52 (10):925-930. doi: <https://doi.org/10.1099/jmm.0.05009-0>.
- Lemaitre, J. P., A. Duroux, R. Pimpie, J. M. Duez, and M. L. Milat. 2015. "Listeria phage and phage tail induction triggered by components of bacterial growth media (phosphate, LiCl, nalidixic acid, and acriflavine)." *Appl Environ Microbiol* 81 (6):2117-24. doi: 10.1128/AEM.03235-14.
- Lenoir, M., R. Martin, E. Torres-Maravilla, S. Chadi, P. Gonzalez-Davila, H. Sokol, P. Langella, F. Chain, and L. G. Bermudez-Humaran. 2020. "Butyrate mediates anti-inflammatory effects of

- Faecalibacterium prausnitzii in intestinal epithelial cells through Dact3." *Gut Microbes* 12 (1):1-16. doi: 10.1080/19490976.2020.1826748.
- Lepage, P., J. Colombet, P. Marteau, T. Sime-Ngando, J. Dore, and M. Leclerc. 2008. "Dysbiosis in inflammatory bowel disease: a role for bacteriophages?" *Gut* 57 (3):424-5. doi: 10.1136/gut.2007.134668.
- Li, W., and A. Godzik. 2006. "Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences." *Bioinformatics* 22 (13):1658-9. doi: 10.1093/bioinformatics/btl158.
- Liang, G., A. G. Cobian-Guemes, L. Albenberg, and F. Bushman. 2021. "The gut virome in inflammatory bowel diseases." *Curr Opin Virol* 51:190-198. doi: 10.1016/j.coviro.2021.10.005.
- Liang, G., M. A. Conrad, J. R. Kelsen, L. R. Kessler, J. Breton, L. G. Albenberg, S. Marakos, A. Galgano, N. Devas, J. Erlichman, H. Zhang, L. Mattei, K. Bittinger, R. N. Baldassano, and F. D. Bushman. 2020. "Dynamics of the Stool Virome in Very Early-Onset Inflammatory Bowel Disease." *J Crohns Colitis* 14 (11):1600-1610. doi: 10.1093/ecco-jcc/jjaa094.
- Lim, E. S., Y. Zhou, G. Zhao, I. K. Bauer, L. Droit, I. M. Ndao, B. B. Warner, P. I. Tarr, D. Wang, and L. R. Holtz. 2015. "Early life dynamics of the human gut virome and bacterial microbiome in infants." *Nat Med* 21 (10):1228-34. doi: 10.1038/nm.3950.
- Lima-Mendez, G., A. Toussaint, and R. Leplae. 2011. "A modular view of the bacteriophage genomic space: identification of host and lifestyle marker modules." *Res Microbiol* 162 (8):737-46. doi: 10.1016/j.resmic.2011.06.006.
- Lippe, R. 2018. "Flow Virometry: a Powerful Tool To Functionally Characterize Viruses." *J Virol* 92 (3). doi: 10.1128/JVI.01765-17.
- Liu, Junfeng, Nicolas Soler, Aurore Gorlas, Virginija Cvirkaite-Krupovic, Mart Krupovic, and Patrick Forterre. 2021. "Extracellular membrane vesicles and nanotubes in Archaea." *microLife* 2. doi: 10.1093/femsml/uqab007.
- Liu, S., A. P. da Cunha, R. M. Rezende, R. Cialic, Z. Wei, L. Bry, L. E. Comstock, R. Gandhi, and H. L. Weiner. 2016. "The Host Shapes the Gut Microbiota via Fecal MicroRNA." *Cell Host Microbe* 19 (1):32-43. doi: 10.1016/j.chom.2015.12.005.
- Liu, Y., S. Alexeeva, H. Bachmann, J. A. Guerra Martinez, N. Yeremenko, T. Abee, and E. J. Smid. 2022. "Chronic Release of Tailless Phage Particles from Lactococcus lactis." *Appl Environ Microbiol* 88 (1):e0148321. doi: 10.1128/AEM.01483-21.
- Liu, Y., K. A. Y. Defourny, E. J. Smid, and T. Abee. 2018. "Gram-Positive Bacterial Extracellular Vesicles and Their Impact on Health and Disease." *Front Microbiol* 9:1502. doi: 10.3389/fmicb.2018.01502.
- Los, M., and G. Wegrzyn. 2012. "Pseudolysogeny." *Adv Virus Res* 82:339-49. doi: 10.1016/B978-0-12-394621-8.00019-4.
- Lourenco, M., L. Chaffringeon, Q. Lamy-Besnier, T. Pedron, P. Campagne, C. Eberl, M. Berard, B. Stecher, L. Debarbieux, and L. De Sordi. 2020. "The Spatial Heterogeneity of the Gut Limits Predation and Fosters Coexistence of Bacteria and Bacteriophages." *Cell Host Microbe* 28 (3):390-401 e5. doi: 10.1016/j.chom.2020.06.002.
- Lovat, L. B. 1996. "Age related changes in gut physiology and nutritional status." *Gut* 38 (3):306-9. doi: 10.1136/gut.38.3.306.
- Luz, Bsrđ, A. Nicolas, S. Chabelskaya, V. R. Rodovalho, Y. Le Loir, V. A. C. Azevedo, B. Felden, and E. Guedon. 2021. "Environmental Plasticity of the RNA Content of Staphylococcus aureus Extracellular Vesicles." *Front Microbiol* 12:634226. doi: 10.3389/fmicb.2021.634226.

- Maas, S. L., J. De Vrij, and M. L. Broekman. 2014. "Quantification and size-profiling of extracellular vesicles using tunable resistive pulse sensing." *J Vis Exp* (92):e51623. doi: 10.3791/51623.
- Machiels, K., M. Joossens, J. Sabino, V. De Preter, I. Arijis, V. Eeckhaut, V. Ballet, K. Claes, F. Van Immerseel, K. Verbeke, M. Ferrante, J. Verhaegen, P. Rutgeerts, and S. Vermeire. 2014. "A decrease of the butyrate-producing species *Roseburia hominis* and *Faecalibacterium prausnitzii* defines dysbiosis in patients with ulcerative colitis." *Gut* 63 (8):1275-83. doi: 10.1136/gutjnl-2013-304833.
- Mackay, I. M., K. E. Arden, and A. Nitsche. 2002. "Real-time PCR in virology." *Nucleic Acids Res* 30 (6):1292-305. doi: 10.1093/nar/30.6.1292.
- Maguire, Ciarán M., Katherine Sillence, Matthias Roesslein, Claire Hannell, Guillaume Suarez, Jean-Jacques Sauvain, Sonja Capracotta, Servane Contal, Sebastien Cambier, Naouale El Yamani, Maria Dusinska, Agnieszka Dybowska, Antje Vennemann, Laura Cooke, Andrea Haase, Andreas Luch, Martin Wiemann, Arno Gutleb, Rafi Korenstein, Michael Riediker, Peter Wick, Patrick Hole, and Adrielle Prina-Mello. 2017. "Benchmark of Nanoparticle Tracking Analysis on Measuring Nanoparticle Sizing and Concentration." *Journal of Micro and Nano-Manufacturing* 5 (4). doi: 10.1115/1.4037124.
- Majewska, J., W. Beta, D. Lecion, K. Hodyra-Stefaniak, A. Klopot, Z. Kazmierczak, P. Miernikiewicz, A. Piotrowicz, J. Ciekot, B. Owczarek, A. Kopciuch, K. Wojtyna, M. Harhala, M. Makosa, and K. Dabrowska. 2015. "Oral Application of T4 Phage Induces Weak Antibody Production in the Gut and in the Blood." *Viruses* 7 (8):4783-99. doi: 10.3390/v7082845.
- Mandal, P. K., G. Ballerin, L. M. Nolan, N. K. Petty, and C. B. Whitchurch. 2021. "Bacteriophage infection of *Escherichia coli* leads to the formation of membrane vesicles via both explosive cell lysis and membrane blebbing." *Microbiology (Reading)* 167 (4). doi: 10.1099/mic.0.001021.
- Manrique, P., B. Bolduc, S. T. Walk, J. van der Oost, W. M. de Vos, and M. J. Young. 2016. "Healthy human gut phageome." *Proc Natl Acad Sci U S A* 113 (37):10400-5. doi: 10.1073/pnas.1601060113.
- Mao, R., Y. L. Xiao, X. Gao, B. L. Chen, Y. He, L. Yang, P. J. Hu, and M. H. Chen. 2012. "Fecal calprotectin in predicting relapse of inflammatory bowel diseases: a meta-analysis of prospective studies." *Inflamm Bowel Dis* 18 (10):1894-9. doi: 10.1002/ibd.22861.
- Marbouty, M., L. Baudry, A. Cournac, and R. Koszul. 2017. "Scaffolding bacterial genomes and probing host-virus interactions in gut microbiome by proximity ligation (chromosome capture) assay." *Sci Adv* 3 (2):e1602105. doi: 10.1126/sciadv.1602105.
- Martín, Rebeca, Florian Chain, Sylvie Miquel, Jun Lu, Jean-Jacques Gratadoux, Harry Sokol, Elena F. Verdu, Premysl Bercik, Luis G. Bermúdez-Humarán, and Philippe Langella. 2014. "The Commensal Bacterium *Faecalibacterium prausnitzii* Is Protective in DNBS-induced Chronic Moderate and Severe Colitis Models." *Inflammatory Bowel Diseases* 20 (3):417-430. doi: 10.1097/01.mib.0000440815.76627.64.
- Martinez-Medina, M., X. Aldeguer, M. Lopez-Siles, F. Gonzalez-Huix, C. Lopez-Oliu, G. Dahbi, J. E. Blanco, J. Blanco, L. J. Garcia-Gil, and A. Darfeuille-Michaud. 2009. "Molecular diversity of *Escherichia coli* in the human gut: new ecological evidence supporting the role of adherent-invasive *E. coli* (AIEC) in Crohn's disease." *Inflamm Bowel Dis* 15 (6):872-82. doi: 10.1002/ibd.20860.
- Masuda, Kuniyoshi, and Tomio Kawata. 1979. "Bacteriophage-like particles induced from the Reiter treponeme by mitomycin C." *FEMS Microbiology Letters* 6 (1):29-31. doi: 10.1111/j.1574-6968.1979.tb04270.x.

- Mateescu, B., E. J. Kowal, B. W. van Balkom, S. Bartel, S. N. Bhattacharyya, E. I. Buzas, A. H. Buck, P. de Candia, F. W. Chow, S. Das, T. A. Driedonks, L. Fernandez-Messina, F. Haderk, A. F. Hill, J. C. Jones, K. R. Van Keuren-Jensen, C. P. Lai, C. Lasser, I. D. Liegro, T. R. Lunavat, M. J. Lorenowicz, S. L. Maas, I. Mager, M. Mittelbrunn, S. Momma, K. Mukherjee, M. Nawaz, D. M. Pegtel, M. W. Pfaffl, R. M. Schiffelers, H. Tahara, C. Thery, J. P. Tosar, M. H. Wauben, K. W. Witwer, and E. N. Nolte-'t Hoen. 2017. "Obstacles and opportunities in the functional analysis of extracellular vesicle RNA - an ISEV position paper." *J Extracell Vesicles* 6 (1):1286095. doi: 10.1080/20013078.2017.1286095.
- Mathieu, M., L. Martin-Jaular, G. Lavieu, and C. Thery. 2019. "Specificities of secretion and uptake of exosomes and other extracellular vesicles for cell-to-cell communication." *Nat Cell Biol* 21 (1):9-17. doi: 10.1038/s41556-018-0250-9.
- Maura, D., M. Galtier, C. Le Bouguenec, and L. Debarbieux. 2012. "Virulent bacteriophages can target O104:H4 enteroaggregative Escherichia coli in the mouse intestine." *Antimicrob Agents Chemother* 56 (12):6235-42. doi: 10.1128/AAC.00602-12.
- Mehta, R. S., G. S. Abu-Ali, D. A. Drew, J. Lloyd-Price, A. Subramanian, P. Lochhead, A. D. Joshi, K. L. Ivey, H. Khalili, G. T. Brown, C. DuLong, M. Song, L. H. Nguyen, H. Mallick, E. B. Rimm, J. Izard, C. Huttenhower, and A. T. Chan. 2018. "Stability of the human faecal microbiome in a cohort of adult men." *Nat Microbiol* 3 (3):347-355. doi: 10.1038/s41564-017-0096-0.
- Mendez, II, L. L. Hermann, P. R. Hazelton, and K. M. Coombs. 2000. "A comparative analysis of freon substitutes in the purification of reovirus and calicivirus." *J Virol Methods* 90 (1):59-67. doi: 10.1016/s0166-0934(00)00217-2.
- Miernikiewicz, P., A. Kłopot, R. Soluch, P. Szkuta, W. Keska, K. Hodyra-Stefaniak, A. Konopka, M. Nowak, D. Lecion, Z. Kazmierczak, J. Majewska, M. Harhala, A. Gorski, and K. Dabrowska. 2016. "T4 Phage Tail Adhesin Gp12 Counteracts LPS-Induced Inflammation In Vivo." *Front Microbiol* 7:1112. doi: 10.3389/fmicb.2016.01112.
- Minot, S., A. Bryson, C. Chehoud, G. D. Wu, J. D. Lewis, and F. D. Bushman. 2013. "Rapid evolution of the human gut virome." *Proc Natl Acad Sci U S A* 110 (30):12450-5. doi: 10.1073/pnas.1300833110.
- Mitsuhashi, S., L. Feldbrugge, E. Csizmadia, M. Mitsuhashi, S. C. Robson, and A. C. Moss. 2016. "Luminal Extracellular Vesicles (EVs) in Inflammatory Bowel Disease (IBD) Exhibit Proinflammatory Effects on Epithelial Cells and Macrophages." *Inflamm Bowel Dis* 22 (7):1587-95. doi: 10.1097/MIB.0000000000000840.
- Moloney, G. M., T. G. Dinan, G. Clarke, and J. F. Cryan. 2019. "Microbial regulation of microRNA expression in the brain-gut axis." *Curr Opin Pharmacol* 48:120-126. doi: 10.1016/j.coph.2019.08.005.
- Momen-Heravi, F., L. Balaj, S. Alian, J. Tigges, V. Toxavidis, M. Ericsson, R. J. Distel, A. R. Ivanov, J. Skog, and W. P. Kuo. 2012. "Alternative methods for characterization of extracellular vesicles." *Front Physiol* 3:354. doi: 10.3389/fphys.2012.00354.
- Morad, G., C. V. Carman, E. J. Hagedorn, J. R. Perlin, L. I. Zon, N. Mustafaoglu, T. E. Park, D. E. Ingber, C. C. Daisy, and M. A. Moses. 2019. "Tumor-Derived Extracellular Vesicles Breach the Intact Blood-Brain Barrier via Transcytosis." *ACS Nano* 13 (12):13853-13865. doi: 10.1021/acsnano.9b04397.
- Mosli, M. H., G. Zou, S. K. Garg, S. G. Feagan, J. K. MacDonald, N. Chande, W. J. Sandborn, and B. G. Feagan. 2015. "C-Reactive Protein, Fecal Calprotectin, and Stool Lactoferrin for Detection of Endoscopic Activity in Symptomatic Inflammatory Bowel Disease Patients: A Systematic

- Review and Meta-Analysis." *Am J Gastroenterol* 110 (6):802-19; quiz 820. doi: 10.1038/ajg.2015.120.
- Netzel, Thomas L., Kambiz Nafisi, Min Zhao, Jerome R. Lenhard, and Iain Johnson. 1995. "Base-Content Dependence of Emission Enhancements, Quantum Yields, and Lifetimes for Cyanine Dyes Bound to Double-Strand DNA: Photophysical Properties of Monomeric and Bichromophoric DNA Stains." *The Journal of Physical Chemistry* 99 (51):17936-17947. doi: 10.1021/j100051a019.
- Nguyen, S., K. Baker, B. S. Padman, R. Patwa, R. A. Dunstan, T. A. Weston, K. Schlosser, B. Bailey, T. Lithgow, M. Lazarou, A. Luque, F. Rohwer, R. S. Blumberg, and J. J. Barr. 2017. "Bacteriophage Transcytosis Provides a Mechanism To Cross Epithelial Cell Layers." *mBio* 8 (6). doi: 10.1128/mBio.01874-17.
- Ni, J., G. D. Wu, L. Albenberg, and V. T. Tomov. 2017. "Gut microbiota and IBD: causation or correlation?" *Nat Rev Gastroenterol Hepatol* 14 (10):573-584. doi: 10.1038/nrgastro.2017.88.
- Norman, J. M., S. A. Handley, M. T. Baldrige, L. Droit, C. Y. Liu, B. C. Keller, A. Kambal, C. L. Monaco, G. Zhao, P. Fleshner, T. S. Stappenbeck, D. P. McGovern, A. Keshavarzian, E. A. Mutlu, J. Sauk, D. Gevers, R. J. Xavier, D. Wang, M. Parkes, and H. W. Virgin. 2015. "Disease-specific alterations in the enteric virome in inflammatory bowel disease." *Cell* 160 (3):447-60. doi: 10.1016/j.cell.2015.01.002.
- Nurk, S., A. Bankevich, D. Antipov, A. A. Gurevich, A. Korobeynikov, A. Lapidus, A. D. Pribelski, A. Pyshkin, A. Sirotkin, Y. Sirotkin, R. Stepanauskas, S. R. Clingenpeel, T. Woyke, J. S. McLean, R. Lasken, G. Tesler, M. A. Alekseyev, and P. A. Pevzner. 2013. "Assembling single-cell genomes and mini-metagenomes from chimeric MDA products." *J Comput Biol* 20 (10):714-37. doi: 10.1089/cmb.2013.0084.
- O'Hara, A. M., and F. Shanahan. 2006. "The gut flora as a forgotten organ." *EMBO Rep* 7 (7):688-93. doi: 10.1038/sj.embor.7400731.
- Ogilvie, L. A., L. D. Bowler, J. Caplin, C. Dedi, D. Diston, E. Cheek, H. Taylor, J. E. Ebdon, and B. V. Jones. 2013. "Genome signature-based dissection of human gut metagenomes to extract subliminal viral sequences." *Nat Commun* 4:2420. doi: 10.1038/ncomms3420.
- Oh, J. H., L. M. Alexander, M. Pan, K. L. Schueler, M. P. Keller, A. D. Attie, J. Walter, and J. P. van Pijkeren. 2019. "Dietary Fructose and Microbiota-Derived Short-Chain Fatty Acids Promote Bacteriophage Production in the Gut Symbiont *Lactobacillus reuteri*." *Cell Host Microbe* 25 (2):273-284 e6. doi: 10.1016/j.chom.2018.11.016.
- Ortega Arroyo, J., D. Cole, and P. Kukura. 2016. "Interferometric scattering microscopy and its combination with single-molecule fluorescence imaging." *Nat Protoc* 11 (4):617-33. doi: 10.1038/nprot.2016.022.
- Osteikoetxea, X., B. Sodar, A. Nemeth, K. Szabo-Taylor, K. Paloczi, K. V. Vukman, V. Tamasi, A. Balogh, A. Kittel, E. Pallinger, and E. I. Buzas. 2015. "Differential detergent sensitivity of extracellular vesicle subpopulations." *Org Biomol Chem* 13 (38):9775-82. doi: 10.1039/c5ob01451d.
- Ottman, N., H. Smidt, W. M. de Vos, and C. Belzer. 2012. "The function of our microbiota: who is out there and what do they do?" *Front Cell Infect Microbiol* 2:104. doi: 10.3389/fcimb.2012.00104.
- Palmer, C., E. M. Bik, D. B. DiGiulio, D. A. Relman, and P. O. Brown. 2007. "Development of the human infant intestinal microbiota." *PLoS Biol* 5 (7):e177. doi: 10.1371/journal.pbio.0050177.
- Palmieri, Valentina, Donatella Lucchetti, Ilaria Gatto, Alessandro Maiorana, Margherita Marcantoni, Giuseppe Maulucci, Massimiliano Papi, Roberto Pola, Marco De Spirito, and Alessandro

- Sgambato. 2014. "Dynamic light scattering for the characterization and counting of extracellular vesicles: a powerful noninvasive tool." *Journal of Nanoparticle Research* 16 (9):2583. doi: 10.1007/s11051-014-2583-z.
- Pang, Y., H. Song, and W. Cheng. 2016. "Using optical trap to measure the refractive index of a single animal virus in culture fluid with high precision." *Biomed Opt Express* 7 (5):1672-89. doi: 10.1364/BOE.7.001672.
- Pasolli, E., F. Asnicar, S. Manara, M. Zolfo, N. Karcher, F. Armanini, F. Beghini, P. Manghi, A. Tett, P. Ghensi, M. C. Collado, B. L. Rice, C. DuLong, X. C. Morgan, C. D. Golden, C. Quince, C. Huttenhower, and N. Segata. 2019. "Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle." *Cell* 176 (3):649-+. doi: 10.1016/j.cell.2019.01.001.
- Peng, X., A. Nguyen, and D. Ghosh. 2018. "Quantification of M13 and T7 bacteriophages by TaqMan and SYBR green qPCR." *J Virol Methods* 252:100-107. doi: 10.1016/j.jviromet.2017.11.012.
- Qin, J., R. Li, J. Raes, M. Arumugam, K. S. Burgdorf, C. Manichanh, T. Nielsen, N. Pons, F. Levenez, T. Yamada, D. R. Mende, J. Li, J. Xu, S. Li, D. Li, J. Cao, B. Wang, H. Liang, H. Zheng, Y. Xie, J. Tap, P. Lepage, M. Bertalan, J. M. Batto, T. Hansen, D. Le Paslier, A. Linneberg, H. B. Nielsen, E. Pelletier, P. Renault, T. Sicheritz-Ponten, K. Turner, H. Zhu, C. Yu, S. Li, M. Jian, Y. Zhou, Y. Li, X. Zhang, S. Li, N. Qin, H. Yang, J. Wang, S. Brunak, J. Dore, F. Guarner, K. Kristiansen, O. Pedersen, J. Parkhill, J. Weissenbach, H. I. T. Consortium Meta, P. Bork, S. D. Ehrlich, and J. Wang. 2010. "A human gut microbial gene catalogue established by metagenomic sequencing." *Nature* 464 (7285):59-65. doi: 10.1038/nature08821.
- Quevrain, E., M. A. Maubert, H. Sokol, B. Devreese, and P. Seksik. 2016. "The presence of the anti-inflammatory protein MAM, from *Faecalibacterium prausnitzii*, in the intestinal ecosystem." *Gut* 65 (5):882. doi: 10.1136/gutjnl-2015-311094.
- Rajilic-Stojanovic, M., and W. M. de Vos. 2014. "The first 1000 cultured species of the human gastrointestinal microbiota." *FEMS Microbiol Rev* 38 (5):996-1047. doi: 10.1111/1574-6976.12075.
- Ratajczak, J., M. Wysoczynski, F. Hayek, A. Janowska-Wieczorek, and M. Z. Ratajczak. 2006. "Membrane-derived microvesicles: important and underappreciated mediators of cell-to-cell communication." *Leukemia* 20 (9):1487-95. doi: 10.1038/sj.leu.2404296.
- Ravantti, J. J., A. Gaidelyte, D. H. Bamford, and J. K. Bamford. 2003. "Comparative analysis of bacterial viruses Bam35, infecting a gram-positive host, and PRD1, infecting gram-negative hosts, demonstrates a viral lineage." *Virology* 313 (2):401-14. doi: 10.1016/s0042-6822(03)00295-2.
- Reddavid, R., O. Rotolo, M. G. Caruso, E. Stasi, M. Notarnicola, C. Miraglia, A. Nouvenne, T. Meschi, G. L. De' Angelis, F. Di Mario, and G. Leandro. 2018. "The role of diet in the prevention and treatment of Inflammatory Bowel Diseases." *Acta Biomed* 89 (9-S):60-75. doi: 10.23750/abm.v89i9-S.7952.
- Reeves, A. E., M. J. Koenigsnecht, I. L. Bergin, and V. B. Young. 2012. "Suppression of *Clostridium difficile* in the gastrointestinal tracts of germfree mice inoculated with a murine isolate from the family Lachnospiraceae." *Infect Immun* 80 (11):3786-94. doi: 10.1128/IAI.00647-12.
- Reyes, A., L. V. Blanton, S. Cao, G. Zhao, M. Manary, I. Trehan, M. I. Smith, D. Wang, H. W. Virgin, F. Rohwer, and J. I. Gordon. 2015. "Gut DNA viromes of Malawian twins discordant for severe acute malnutrition." *Proc Natl Acad Sci U S A* 112 (38):11941-6. doi: 10.1073/pnas.1514285112.

- Reyes, A., M. Haynes, N. Hanson, F. E. Angly, A. C. Heath, F. Rohwer, and J. I. Gordon. 2010. "Viruses in the faecal microbiota of monozygotic twins and their mothers." *Nature* 466 (7304):334-8. doi: 10.1038/nature09199.
- Rigottier-Gois, L. 2013. "Dysbiosis in inflammatory bowel diseases: the oxygen hypothesis." *ISME J* 7 (7):1256-61. doi: 10.1038/ismej.2013.80.
- Ripp, S., and R. V. Miller. 1997. "The role of pseudolysogeny in bacteriophage-host interactions in a natural freshwater environment." *Microbiology (Reading)* 143 (6):2065-2070. doi: 10.1099/00221287-143-6-2065.
- Robbins, P. D., A. Dorronsoro, and C. N. Booker. 2016. "Regulation of chronic inflammatory and immune processes by extracellular vesicles." *J Clin Invest* 126 (4):1173-80. doi: 10.1172/JCI81131.
- Rodvalho, V. R., B. S. R. da Luz, H. Rabah, F. L. R. do Carmo, E. L. Folador, A. Nicolas, J. Jardin, V. Briard-Bion, H. Blottiere, N. Lapaque, G. Jan, Y. Le Loir, V. A. de Carvalho Azevedo, and E. Guedon. 2020. "Extracellular Vesicles Produced by the Probiotic *Propionibacterium freudenreichii* CIRM-BIA 129 Mitigate Inflammation by Modulating the NF-kappaB Pathway." *Front Microbiol* 11:1544. doi: 10.3389/fmicb.2020.01544.
- Roose-Amsaleg, C., Y. Fedala, C. Venien-Bryan, J. Garnier, A. C. Boccara, and M. Boccara. 2017. "Utilization of interferometric light microscopy for the rapid analysis of virus abundance in a river." *Res Microbiol* 168 (5):413-418. doi: 10.1016/j.resmic.2017.02.004.
- Rossi, C. A., B. J. Kearney, S. P. Olschner, P. L. Williams, C. G. Robinson, M. L. Heinrich, A. M. Zovanyi, M. F. Ingram, D. A. Norwood, and R. J. Schoepp. 2015. "Evaluation of ViroCyt(R) Virus Counter for rapid filovirus quantitation." *Viruses* 7 (3):857-72. doi: 10.3390/v7030857.
- Ruska, H. 1940. "Die Sichtbarmachung der bakteriophagen Lyse im Übermikroskop." *Naturwissenschaften* 28:45-46. doi: doi.org/10.1007/BF01486931.
- Sabbagh, Q., G. Andre-Gregoire, C. Alves-Nicolau, A. Dupont, N. Bidere, E. Jouglar, L. Guevel, J. S. Frenel, and J. Gavard. 2021. "The von Willebrand factor stamps plasmatic extracellular vesicles from glioblastoma patients." *Sci Rep* 11 (1):22792. doi: 10.1038/s41598-021-02254-7.
- Sabino, João, Sara Vieira-Silva, Kathleen Machiels, Marie Joossens, Gwen Falony, Vera Ballet, Marc Ferrante, Gert Van Assche, Schalk Van der Merwe, Severine Vermeire, and Jeroen Raes. 2016. "Primary sclerosing cholangitis is characterised by intestinal dysbiosis independent from IBD." *Gut* 65 (10):1681-1689. doi: 10.1136/gutjnl-2015-311004.
- Salazar, N., S. Arboleya, L. Valdes, C. Stanton, P. Ross, L. Ruiz, M. Gueimonde, and C. G. de Los Reyes-Gavilan. 2014. "The human intestinal microbiome at extreme ages of life. Dietary intervention as a way to counteract alterations." *Front Genet* 5:406. doi: 10.3389/fgene.2014.00406.
- Salton, M. R., and A. Netschey. 1965. "Physical chemistry of isolated bacterial membranes." *Biochim Biophys Acta* 107 (3):539-45. doi: 10.1016/0304-4165(65)90198-4.
- Sausset, R., M. A. Petit, V. Gaboriau-Routhiau, and M. De Paepe. 2020. "New insights into intestinal phages." *Mucosal Immunol* 13 (2):205-215. doi: 10.1038/s41385-019-0250-5.
- Schwechheimer, C., and M. J. Kuehn. 2015. "Outer-membrane vesicles from Gram-negative bacteria: biogenesis and functions." *Nat Rev Microbiol* 13 (10):605-19. doi: 10.1038/nrmicro3525.
- Seemann, T. 2014. "Prokka: rapid prokaryotic genome annotation." *Bioinformatics* 30 (14):2068-9. doi: 10.1093/bioinformatics/btu153.
- Sekirov, I., S. L. Russell, L. C. Antunes, and B. B. Finlay. 2010. "Gut microbiota in health and disease." *Physiol Rev* 90 (3):859-904. doi: 10.1152/physrev.00045.2009.

- Shah, Shiraz A., Ling Deng, Jonathan Thorsen, Anders G. Pedersen, Moira B. Dion, Josué L. Castro-Mejía, Ronalds Silins, Fie O. Romme, Romain Sausset, Eric Olo Ndela, Mathis Hjemlsø, Morten A. Rasmussen, Tamsin A. Redgwell, Gisle Vestergaard, Yichang Zhang, Søren J. Sørensen, Hans Bisgaard, Francois Enault, Jakob Stokholm, Sylvain Moineau, Marie-Agnès Petit, and Dennis S. Nielsen. 2021. "Hundreds of viral families in the healthy infant gut." *bioRxiv*:2021.07.02.450849. doi: 10.1101/2021.07.02.450849.
- Shelke, G. V., C. Lasser, Y. S. Gho, and J. Lotvall. 2014. "Importance of exosome depletion protocols to eliminate functional and RNA-containing extracellular vesicles from fetal bovine serum." *J Extracell Vesicles* 3. doi: 10.3402/jev.v3.24783.
- Sherman, P., R. Soni, and H. Yeger. 1988. "Characterization of flagella purified from enterohemorrhagic, vero-cytotoxin-producing *Escherichia coli* serotype O157:H7." *J Clin Microbiol* 26 (7):1367-72. doi: 10.1128/jcm.26.7.1367-1372.1988.
- Shkoporov, A. N., A. G. Clooney, T. D. S. Sutton, F. J. Ryan, K. M. Daly, J. A. Nolan, S. A. McDonnell, E. V. Khokhlova, L. A. Draper, A. Forde, E. Guerin, V. Velayudhan, R. P. Ross, and C. Hill. 2019. "The Human Gut Virome Is Highly Diverse, Stable, and Individual Specific." *Cell Host Microbe* 26 (4):527-541 e5. doi: 10.1016/j.chom.2019.09.009.
- Shkoporov, A. N., and C. Hill. 2019. "Bacteriophages of the Human Gut: The "Known Unknown" of the Microbiome." *Cell Host Microbe* 25 (2):195-209. doi: 10.1016/j.chom.2019.01.017.
- Shkoporov, A. N., E. V. Khokhlova, C. B. Fitzgerald, S. R. Stockdale, L. A. Draper, R. P. Ross, and C. Hill. 2018. "PhiCrAss001 represents the most abundant bacteriophage family in the human gut and infects *Bacteroides intestinalis*." *Nat Commun* 9 (1):4781. doi: 10.1038/s41467-018-07225-7.
- Shkoporov, A. N., F. J. Ryan, L. A. Draper, A. Forde, S. R. Stockdale, K. M. Daly, S. A. McDonnell, J. A. Nolan, T. D. S. Sutton, M. Dalmasso, A. McCann, R. P. Ross, and C. Hill. 2018. "Reproducible protocols for metagenomic analysis of human faecal phageomes." *Microbiome* 6 (1):68. doi: 10.1186/s40168-018-0446-z.
- Shkoporov, Andrey, Stephen Stockdale, Aonghus Lavelle, Ivanela Kondova, Cara Hueston, Aditya Upadrasta, Ekaterina Khokhlova, Imme van der Kamp, Boudewijn Ouwerling, Lorraine Draper, Jan Langermans, Paul Ross, and Colin Hill. 2021. Viral biogeography of gastrointestinal tract and parenchymal organs in two representative species of mammals. Research Square.
- Sieber, V., A. Pluckthun, and F. X. Schmid. 1998. "Selecting proteins with improved stability by a phage-based method." *Nat Biotechnol* 16 (10):955-60. doi: 10.1038/nbt1098-955.
- Sieradzki, E. T., J. C. Ignacio-Espinoza, D. M. Needham, E. B. Fichot, and J. A. Fuhrman. 2019. "Dynamic marine viral infections and major contribution to photosynthetic processes shown by spatiotemporal picoplankton metatranscriptomes." *Nat Commun* 10 (1):1169. doi: 10.1038/s41467-019-09106-z.
- Silveira, C. B., and F. L. Rohwer. 2016. "Piggyback-the-Winner in host-associated microbial communities." *NPJ Biofilms Microbiomes* 2:16010. doi: 10.1038/npjbiofilms.2016.10.
- Simons, M., and G. Raposo. 2009. "Exosomes--vesicular carriers for intercellular communication." *Curr Opin Cell Biol* 21 (4):575-81. doi: 10.1016/j.ceb.2009.03.007.
- Singh, R. K., H. W. Chang, D. Yan, K. M. Lee, D. Ucmak, K. Wong, M. Abrouk, B. Farahnik, M. Nakamura, T. H. Zhu, T. Bhutani, and W. Liao. 2017. "Influence of diet on the gut microbiome and implications for human health." *J Transl Med* 15 (1):73. doi: 10.1186/s12967-017-1175-y.

- Sipponen, T., E. Savilahti, K. L. Kolho, H. Nuutinen, U. Turunen, and M. Farkkila. 2008. "Crohn's disease activity assessed by fecal calprotectin and lactoferrin: correlation with Crohn's disease activity index and endoscopic findings." *Inflamm Bowel Dis* 14 (1):40-6. doi: 10.1002/ibd.20312.
- Slomka, A., S. K. Urban, V. Lukacs-Kornek, E. Zekanowska, and M. Kornek. 2018. "Large Extracellular Vesicles: Have We Found the Holy Grail of Inflammation?" *Front Immunol* 9:2723. doi: 10.3389/fimmu.2018.02723.
- Sokol, H., V. Leducq, H. Aschard, H. P. Pham, S. Jegou, C. Landman, D. Cohen, G. Liguori, A. Bourrier, I. Nion-Larmurier, J. Cosnes, P. Seksik, P. Langella, D. Skurnik, M. L. Richard, and L. Beaugerie. 2017. "Fungal microbiota dysbiosis in IBD." *Gut* 66 (6):1039-1048. doi: 10.1136/gutjnl-2015-310746.
- Sokol, H., P. Seksik, J. P. Furet, O. Firmesse, I. Nion-Larmurier, L. Beaugerie, J. Cosnes, G. Corthier, P. Marteau, and J. Dore. 2009. "Low counts of *Faecalibacterium prausnitzii* in colitis microbiota." *Inflamm Bowel Dis* 15 (8):1183-9. doi: 10.1002/ibd.20903.
- Sokol, H., P. Seksik, L. Rigottier-Gois, C. Lay, P. Lepage, I. Podglajen, P. Marteau, and J. Dore. 2006. "Specificities of the fecal microbiota in inflammatory bowel disease." *Inflamm Bowel Dis* 12 (2):106-11. doi: 10.1097/01.MIB.0000200323.38139.c6.
- Soret, R., J. Chevalier, P. De Coppet, G. Poupeau, P. Derkinderen, J. P. Segain, and M. Neunlist. 2010. "Short-chain fatty acids regulate the enteric neurons and control gastrointestinal motility in rats." *Gastroenterology* 138 (5):1772-82. doi: 10.1053/j.gastro.2010.01.053.
- Spor, A., O. Koren, and R. Ley. 2011. "Unravelling the effects of the environment and host genotype on the gut microbiome." *Nat Rev Microbiol* 9 (4):279-90. doi: 10.1038/nrmicro2540.
- Stark, P. L., and A. Lee. 1982. "The microbial ecology of the large bowel of breast-fed and formula-fed infants during the first year of life." *J Med Microbiol* 15 (2):189-203. doi: 10.1099/00222615-15-2-189.
- Stern, A., E. Mick, I. Tirosh, O. Sagy, and R. Sorek. 2012. "CRISPR targeting reveals a reservoir of common phages associated with the human gut microbiome." *Genome Res* 22 (10):1985-94. doi: 10.1101/gr.138297.112.
- Suttle, Curtis & Fuhrman, Jed. 2010. "Enumeration of virus particles in aquatic or sediment samples by epifluorescence microscopy." In *Manual of Aquatic Viral Ecology*, 145-153.
- Sutton, T. D. S., and C. Hill. 2019. "Gut Bacteriophage: Current Understanding and Challenges." *Front Endocrinol (Lausanne)* 10:784. doi: 10.3389/fendo.2019.00784.
- Sweere, J. M., J. D. Van Belleghem, H. Ishak, M. S. Bach, M. Popescu, V. Sunkari, G. Kaber, R. Manasherob, G. A. Suh, X. Cao, C. R. de Vries, D. N. Lam, P. L. Marshall, M. Birukova, E. Katznelson, D. V. Lazzareschi, S. Balaji, S. G. Keswani, T. R. Hawn, P. R. Secor, and P. L. Bollyky. 2019. "Bacteriophage trigger antiviral immunity and prevent clearance of bacterial infection." *Science* 363 (6434). doi: 10.1126/science.aat9691.
- Swidsinski, A., J. Weber, V. Loening-Baucke, L. P. Hale, and H. Lochs. 2005. "Spatial organization and composition of the mucosal flora in patients with inflammatory bowel disease." *J Clin Microbiol* 43 (7):3380-9. doi: 10.1128/JCM.43.7.3380-3389.2005.
- Swietnicki, W., M. Morillas, S. G. Chen, P. Gambetti, and W. K. Surewicz. 2000. "Aggregation and fibrillization of the recombinant human prion protein huPrP90-231." *Biochemistry* 39 (2):424-31. doi: 10.1021/bi991967m.

- Szatanek, R., M. Baj-Krzyworzeka, J. Zimoch, M. Lekka, M. Siedlar, and J. Baran. 2017. "The Methods of Choice for Extracellular Vesicles (EVs) Characterization." *Int J Mol Sci* 18 (6). doi: 10.3390/ijms18061153.
- Tabak, S., U. Hadad, S. Schreiber-Avissar, and E. Beit-Yannai. 2021. "Non-pigmented ciliary epithelium derived extracellular vesicles uptake mechanism by the trabecular meshwork." *FASEB J* 35 (2):e21188. doi: 10.1096/fj.202002040R.
- Tamanai-Shacoori, Z., I. Smida, L. Bousarghin, O. Loreal, V. Meuric, S. B. Fong, M. Bonnaure-Mallet, and A. Jolivet-Gougeon. 2017. "Roseburia spp.: a marker of health?" *Future Microbiol* 12:157-170. doi: 10.2217/fmb-2016-0130.
- Tetz, G. V., K. V. Ruggles, H. Zhou, A. Heguy, A. Tsirigos, and V. Tetz. 2017. "Bacteriophages as potential new mammalian pathogens." *Sci Rep* 7 (1):7043. doi: 10.1038/s41598-017-07278-6.
- Thery, C., S. Amigorena, G. Raposo, and A. Clayton. 2006. "Isolation and characterization of exosomes from cell culture supernatants and biological fluids." *Curr Protoc Cell Biol* Chapter 3:Unit 3 22. doi: 10.1002/0471143030.cb0322s30.
- Thery, C., M. Boussac, P. Veron, P. Ricciardi-Castagnoli, G. Raposo, J. Garin, and S. Amigorena. 2001. "Proteomic analysis of dendritic cell-derived exosomes: a secreted subcellular compartment distinct from apoptotic vesicles." *J Immunol* 166 (12):7309-18. doi: 10.4049/jimmunol.166.12.7309.
- Thery, C., M. Ostrowski, and E. Segura. 2009. "Membrane vesicles as conveyors of immune responses." *Nat Rev Immunol* 9 (8):581-93. doi: 10.1038/nri2567.
- Thery, C., K. W. Witwer, E. Aikawa, M. J. Alcaraz, J. D. Anderson, R. Andriantsitohaina, A. Antoniou, T. Arab, F. Archer, G. K. Atkin-Smith, D. C. Ayre, J. M. Bach, D. Bachurski, H. Baharvand, L. Balaj, S. Baldacchino, N. N. Bauer, A. A. Baxter, M. Bebawy, C. Beckham, A. Bedina Zavec, A. Benmoussa, A. C. Berardi, P. Bergese, E. Bielska, C. Blenkiron, S. Bobis-Wozowicz, E. Boilard, W. Boireau, A. Bongiovanni, F. E. Borrás, S. Bosch, C. M. Boulanger, X. Breakefield, A. M. Breglio, M. A. Brennan, D. R. Brigstock, A. Brisson, M. L. Broekman, J. F. Bromberg, P. Bryl-Gorecka, S. Buch, A. H. Buck, D. Burger, S. Busatto, D. Buschmann, B. Bussolati, E. I. Buzas, J. B. Byrd, G. Camussi, D. R. Carter, S. Caruso, L. W. Chamley, Y. T. Chang, C. Chen, S. Chen, L. Cheng, A. R. Chin, A. Clayton, S. P. Clerici, A. Cocks, E. Cocucci, R. J. Coffey, A. Cordeiro-da-Silva, Y. Couch, F. A. Coumans, B. Coyle, R. Crescitelli, M. F. Criado, C. D'Souza-Schorey, S. Das, A. Datta Chaudhuri, P. de Candia, E. F. De Santana, O. De Wever, H. A. Del Portillo, T. Demaret, S. Deville, A. Devitt, B. Dhondt, D. Di Vizio, L. C. Dieterich, V. Dolo, A. P. Dominguez Rubio, M. Dominici, M. R. Dourado, T. A. Driedonks, F. V. Duarte, H. M. Duncan, R. M. Eichenberger, K. Ekstrom, S. El Andaloussi, C. Elie-Caille, U. Erdbrugger, J. M. Falcon-Perez, F. Fatima, J. E. Fish, M. Flores-Bellver, A. Forsonits, A. Frelet-Barrand, F. Fricke, G. Fuhrmann, S. Gabrielsson, A. Gamez-Valero, C. Gardiner, K. Gartner, R. Gaudin, Y. S. Gho, B. Giebel, C. Gilbert, M. Gimona, I. Giusti, D. C. Goberdhan, A. Gorgens, S. M. Gorski, D. W. Greening, J. C. Gross, A. Gualerzi, G. N. Gupta, D. Gustafson, A. Handberg, R. A. Haraszti, P. Harrison, H. Hegyesi, A. Hendrix, A. F. Hill, F. H. Hochberg, K. F. Hoffmann, B. Holder, H. Holthofer, B. Hosseinkhani, G. Hu, Y. Huang, V. Huber, S. Hunt, A. G. Ibrahim, T. Ikezu, J. M. Inal, M. Isin, A. Ivanova, H. K. Jackson, S. Jacobsen, S. M. Jay, M. Jayachandran, G. Jenster, L. Jiang, S. M. Johnson, J. C. Jones, A. Jong, T. Jovanovic-Talisman, S. Jung, R. Kalluri, S. I. Kano, S. Kaur, Y. Kawamura, E. T. Keller, D. Khamari, E. Khomyakova, A. Khvorova, P. Kierulf, K. P. Kim, T. Kislinger, M. Klingeborn, D. J. Klinke, 2nd, M. Kornek, M. M. Kosanovic, A. F. Kovacs, E. M. Kramer-Albers, S. Krasemann, M. Krause, I. V. Kurochkin, G. D. Kusuma, S. Kuypers, S. Laitinen, S. M. Langevin, L. R. Languino, J. Lannigan, C. Lasser, L. C. Laurent, G. Lavieu, E. Lazaro-Ibanez, S. Le Lay, M. S. Lee, Y. X. F. Lee,

D. S. Lemos, M. Lenassi, A. Leszczynska, I. T. Li, K. Liao, S. F. Libregts, E. Ligeti, R. Lim, S. K. Lim, A. Line, K. Linnemannstons, A. Llorente, C. A. Lombard, M. J. Lorenowicz, A. M. Lorincz, J. Lotvall, J. Lovett, M. C. Lowry, X. Loyer, Q. Lu, B. Lukomska, T. R. Lunavat, S. L. Maas, H. Malhi, A. Marcilla, J. Mariani, J. Mariscal, E. S. Martens-Uzunova, L. Martin-Jaular, M. C. Martinez, V. R. Martins, M. Mathieu, S. Mathivanan, M. Maugeri, L. K. McGinnis, M. J. McVey, D. G. Meckes, Jr., K. L. Meehan, I. Mertens, V. R. Minciocchi, A. Moller, M. Moller Jorgensen, A. Morales-Kastresana, J. Morhayim, F. Mullier, M. Muraca, L. Musante, V. Mussack, D. C. Muth, K. H. Myburgh, T. Najrana, M. Nawaz, I. Nazarenko, P. Nejsun, C. Neri, T. Neri, R. Nieuwland, L. Nimrichter, J. P. Nolan, E. N. Nolte-'t Hoen, N. Noren Hooten, L. O'Driscoll, T. O'Grady, A. O'Loghlen, T. Ochiya, M. Olivier, A. Ortiz, L. A. Ortiz, X. Osteikoetxea, O. Ostergaard, M. Ostrowski, J. Park, D. M. Pegtel, H. Peinado, F. Perut, M. W. Pfaffl, D. G. Phinney, B. C. Pieters, R. C. Pink, D. S. Pisetsky, E. Pogge von Strandmann, I. Polakovicova, I. K. Poon, B. H. Powell, I. Prada, L. Pulliam, P. Quesenberry, A. Radeghieri, R. L. Raffai, S. Raimondo, J. Rak, M. I. Ramirez, G. Raposo, M. S. Rayyan, N. Regev-Rudzki, F. L. Ricklefs, P. D. Robbins, D. D. Roberts, S. C. Rodrigues, E. Rohde, S. Rome, K. M. Rouschop, A. Rughetti, A. E. Russell, P. Saa, S. Sahoo, E. Salas-Huenuleo, C. Sanchez, J. A. Saugstad, M. J. Saul, R. M. Schiffelers, R. Schneider, T. H. Schoyen, A. Scott, E. Shahaj, S. Sharma, O. Shatnyeva, F. Shekari, G. V. Shelke, A. K. Shetty, K. Shiba, P. R. Siljander, A. M. Silva, A. Skowronek, O. L. Snyder, 2nd, R. P. Soares, B. W. Sodar, C. Soekmadji, J. Sotillo, P. D. Stahl, W. Stoorvogel, S. L. Stott, E. F. Strasser, S. Swift, H. Tahara, M. Tewari, K. Timms, S. Tiwari, R. Tixeira, M. Tkach, W. S. Toh, R. Tomasini, A. C. Torrecilhas, J. P. Tosar, V. Toxavidis, L. Urbanelli, P. Vader, B. W. van Balkom, S. G. van der Grein, J. Van Deun, M. J. van Herwijnen, K. Van Keuren-Jensen, G. van Niel, M. E. van Royen, A. J. van Wijnen, M. H. Vasconcelos, I. J. Vechetti, Jr., T. D. Veit, L. J. Vella, E. Velot, F. J. Verweij, B. Vestad, J. L. Vinas, T. Visnovitz, K. V. Vukman, J. Wahlgren, D. C. Watson, M. H. Wauben, A. Weaver, J. P. Webber, V. Weber, A. M. Wehman, D. J. Weiss, J. A. Welsh, S. Wendt, A. M. Wheelock, Z. Wiener, L. Witte, J. Wolfram, A. Xagorari, P. Xander, J. Xu, X. Yan, M. Yanez-Mo, H. Yin, Y. Yuana, V. Zappulli, J. Zarubova, V. Zekas, J. Y. Zhang, Z. Zhao, L. Zheng, A. R. Zheutlin, A. M. Zickler, P. Zimmermann, A. M. Zivkovic, D. Zocco, and E. K. Zuba-Surma. 2018. "Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines." *J Extracell Vesicles* 7 (1):1535750. doi: 10.1080/20013078.2018.1535750.

Thompson, A. G., E. Gray, S. M. Heman-Ackah, I. Mager, K. Talbot, S. E. Andaloussi, M. J. Wood, and M. R. Turner. 2016. "Extracellular vesicles in neurodegenerative disease - pathogenesis to biomarkers." *Nat Rev Neurol* 12 (6):346-57. doi: 10.1038/nrneurol.2016.68.

Townsend, E. M., L. Kelly, G. Muscatt, J. D. Box, N. Hargraves, D. Lilley, and E. Jameson. 2021. "The Human Gut Phageome: Origins and Roles in the Human Gut Microbiome." *Front Cell Infect Microbiol* 11:643214. doi: 10.3389/fcimb.2021.643214.

Toyofuku, M., N. Nomura, and L. Eberl. 2019. "Types and origins of bacterial membrane vesicles." *Nat Rev Microbiol* 17 (1):13-24. doi: 10.1038/s41579-018-0112-2.

Tulkens, J., O. De Wever, and A. Hendrix. 2020. "Analyzing bacterial extracellular vesicles in human body fluids by orthogonal biophysical separation and biochemical characterization." *Nat Protoc* 15 (1):40-67. doi: 10.1038/s41596-019-0236-5.

Turkki, V., E. Alppila, S. Yla-Herttuala, and H. P. Lesch. 2021. "Experimental Evaluation of an Interferometric Light Microscopy Particle Counter for Titering and Characterization of Virus Preparations." *Viruses* 13 (5). doi: 10.3390/v13050939.

Turner, D., A. M. Kropinski, and E. M. Adriaenssens. 2021. "A Roadmap for Genome-Based Phage Taxonomy." *Viruses* 13 (3). doi: 10.3390/v13030506.

- Twort, F.W. 1915. "AN INVESTIGATION ON THE NATURE OF ULTRA-MICROSCOPIC VIRUSES." *The Lancet* 186 (4814):1241-1243. doi: doi.org/10.1016/S0140-6736(01)20383-3.
- Tzaban, S., G. Friedlander, O. Schonberger, L. Horonchik, Y. Yedidia, G. Shaked, R. Gabizon, and A. Taraboulos. 2002. "Protease-sensitive scrapie prion protein in aggregates of heterogeneous sizes." *Biochemistry* 41 (42):12868-75. doi: 10.1021/bi025958g.
- Valadi, H., K. Ekstrom, A. Bossios, M. Sjostrand, J. J. Lee, and J. O. Lotvall. 2007. "Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells." *Nat Cell Biol* 9 (6):654-9. doi: 10.1038/ncb1596.
- Valguarnera, E., N. E. Scott, P. Azimzadeh, and M. F. Feldman. 2018. "Surface Exposure and Packing of Lipoproteins into Outer Membrane Vesicles Are Coupled Processes in Bacteroides." *mSphere* 3 (6). doi: 10.1128/mSphere.00559-18.
- Vallabhaneni, K. C., P. Penforinis, S. Dhule, F. Guillonneau, K. V. Adams, Y. Y. Mo, R. Xu, Y. Liu, K. Watabe, M. C. Vemuri, and R. Pochampally. 2015. "Extracellular vesicles from bone marrow mesenchymal stem/stromal cells transport tumor regulatory microRNA, proteins, and metabolites." *Oncotarget* 6 (7):4953-67. doi: 10.18632/oncotarget.3211.
- Van Belleghem, J. D., F. Clement, M. Merabishvili, R. Lavigne, and M. Vaneechoutte. 2017. "Pro- and anti-inflammatory responses of peripheral blood mononuclear cells induced by Staphylococcus aureus and Pseudomonas aeruginosa phages." *Sci Rep* 7 (1):8004. doi: 10.1038/s41598-017-08336-9.
- Van de Guchte, M., S. D. Burz, J. Cadiou, J. Wu, S. Mondot, H. M. Blottiere, and J. Dore. 2020. "Alternative stable states in the intestinal ecosystem: proof of concept in a rat model and a perspective of therapeutic implications." *Microbiome* 8 (1):153. doi: 10.1186/s40168-020-00933-7.
- van der Pol, E., A. N. Boing, P. Harrison, A. Sturk, and R. Nieuwland. 2012. "Classification, functions, and clinical relevance of extracellular vesicles." *Pharmacol Rev* 64 (3):676-705. doi: 10.1124/pr.112.005983.
- van der Pol, E., F. A. Coumans, A. E. Grootemaat, C. Gardiner, I. L. Sargent, P. Harrison, A. Sturk, T. G. van Leeuwen, and R. Nieuwland. 2014. "Particle size distribution of exosomes and microvesicles determined by transmission electron microscopy, flow cytometry, nanoparticle tracking analysis, and resistive pulse sensing." *J Thromb Haemost* 12 (7):1182-92. doi: 10.1111/jth.12602.
- van Niel, G., G. D'Angelo, and G. Raposo. 2018. "Shedding light on the cell biology of extracellular vesicles." *Nat Rev Mol Cell Biol* 19 (4):213-228. doi: 10.1038/nrm.2017.125.
- van Tongeren, S. P., J. P. Slaets, H. J. Harmsen, and G. W. Welling. 2005. "Fecal microbiota composition and frailty." *Appl Environ Microbiol* 71 (10):6438-42. doi: 10.1128/AEM.71.10.6438-6442.2005.
- Vandeputte, D., G. Kathagen, K. D'Hoe, S. Vieira-Silva, M. Valles-Colomer, J. Sabino, J. Wang, R. Y. Tito, L. De Commer, Y. Darzi, S. Vermeire, G. Falony, and J. Raes. 2017. "Quantitative microbiome profiling links gut community variation to microbial load." *Nature* 551 (7681):507-511. doi: 10.1038/nature24460.
- VIDAL. 2021. "Maladie de Crohn." <https://www.vidal.fr/maladies/estomac-intestins/maladie-crohn.html>.
- VIDAL. 2022. "<https://www.vidal.fr/maladies/estomac-intestins/rectocolite-hemorragique.html>." <https://www.vidal.fr/maladies/estomac-intestins/rectocolite-hemorragique.html>.

- Viennois, E., B. Chassaing, A. Tahsin, A. Pujada, L. Wang, A. T. Gewirtz, and D. Merlin. 2019. "Host-derived fecal microRNAs can indicate gut microbiota healthiness and ability to induce inflammation." *Theranostics* 9 (15):4542-4557. doi: 10.7150/thno.35282.
- Wagner, J., J. Maksimovic, G. Farries, W. H. Sim, R. F. Bishop, D. J. Cameron, A. G. Catto-Smith, and C. D. Kirkwood. 2013. "Bacteriophages in gut samples from pediatric Crohn's disease patients: metagenomic analysis using 454 pyrosequencing." *Inflamm Bowel Dis* 19 (8):1598-608. doi: 10.1097/MIB.0b013e318292477c.
- Walker, A. W., J. D. Sanderson, C. Churcher, G. C. Parkes, B. N. Hudspith, N. Rayment, J. Brostoff, J. Parkhill, G. Dougan, and L. Petrovska. 2011. "High-throughput clone library analysis of the mucosa-associated microbiota reveals dysbiosis and differences between inflamed and non-inflamed regions of the intestine in inflammatory bowel disease." *BMC Microbiol* 11:7. doi: 10.1186/1471-2180-11-7.
- Wampach, L., A. Heintz-Buschart, A. Hogan, E. E. L. Muller, S. Narayanasamy, C. C. Laczny, L. W. Hugerth, L. Bindl, J. Bottu, A. F. Andersson, C. de Beaufort, and P. Wilmes. 2017. "Colonization and Succession within the Human Gut Microbiome by Archaea, Bacteria, and Microeukaryotes during the First Year of Life." *Front Microbiol* 8:738. doi: 10.3389/fmicb.2017.00738.
- Wang, X., Y. Kim, Q. Ma, S. H. Hong, K. Pokusaeva, J. M. Sturino, and T. K. Wood. 2010. "Cryptic prophages help bacteria cope with adverse environments." *Nat Commun* 1:147. doi: 10.1038/ncomms1146.
- Weinbauer, M. G., and F. Rassoulzadegan. 2004. "Are viruses driving microbial diversification and diversity?" *Environ Microbiol* 6 (1):1-11. doi: 10.1046/j.1462-2920.2003.00539.x.
- Wenger, S. L., J. H. Turner, and J. C. Petricciani. 1978. "The cytogenetic, proliferative and viability effects of four bacteriophages on human lymphocytes." *In Vitro* 14 (6):543-9. doi: 10.1007/BF02616097.
- Wilhelm, Léa P., Laetitia Voilquin, Toshihide Kobayashi, Catherine Tomasetto, and Fabien Alpy. 2019. "Intracellular and Plasma Membrane Cholesterol Labeling and Quantification Using Filipin and GFP-D4." In *Intracellular Lipid Transport: Methods and Protocols*, edited by Guillaume Drin, 137-152. New York, NY: Springer New York.
- William, Sutherland. 1904. "LXXV. A dynamical theory of diffusion for non-electrolytes and the molecular mass of albumin." *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 9 (54):781-785. doi: 10.1080/14786440509463331.
- Willing, B. P., J. Dicksved, J. Halfvarson, A. F. Andersson, M. Lucio, Z. Zheng, G. Jarnerot, C. Tysk, J. K. Jansson, and L. Engstrand. 2010. "A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes." *Gastroenterology* 139 (6):1844-1854 e1. doi: 10.1053/j.gastro.2010.08.049.
- Winter, S., and G. Loeber. 1997. "DNA-binding and fluorescence properties of the DNA bis-intercalating purple oxazole dimer POPO-1." *J Biomed Opt* 2 (1):125-30. doi: 10.1117/12.260019.
- Woith, E., G. Fuhrmann, and M. F. Melzig. 2019. "Extracellular Vesicles-Connecting Kingdoms." *Int J Mol Sci* 20 (22). doi: 10.3390/ijms20225695.
- Wong, J. M., R. de Souza, C. W. Kendall, A. Emam, and D. J. Jenkins. 2006. "Colonic health: fermentation and short chain fatty acids." *J Clin Gastroenterol* 40 (3):235-43. doi: 10.1097/00004836-200603000-00015.

- Yanez-Mo, M., P. R. Siljander, Z. Andreu, A. B. Zavec, F. E. Borrás, E. I. Buzas, K. Buzas, E. Casal, F. Cappello, J. Carvalho, E. Colas, A. Cordeiro-da Silva, S. Fais, J. M. Falcon-Perez, I. M. Ghobrial, B. Giebel, M. Gimona, M. Graner, I. Gursel, M. Gursel, N. H. Heegaard, A. Hendrix, P. Kierulf, K. Kokubun, M. Kosanovic, V. Kralj-Iglic, E. M. Kramer-Albers, S. Laitinen, C. Lasser, T. Lener, E. Ligeti, A. Line, G. Lipps, A. Llorente, J. Lotvall, M. Mancek-Keber, A. Marcilla, M. Mittelbrunn, I. Nazarenko, E. N. Nolte-'t Hoen, T. A. Nyman, L. O'Driscoll, M. Olivan, C. Oliveira, E. Pallinger, H. A. Del Portillo, J. Reventos, M. Rigau, E. Rohde, M. Sammar, F. Sanchez-Madrid, N. Santarem, K. Schallmoser, M. S. Ostenfeld, W. Stoorvogel, R. Stukelj, S. G. Van der Grein, M. H. Vasconcelos, M. H. Wauben, and O. De Wever. 2015. "Biological properties of extracellular vesicles and their physiological functions." *J Extracell Vesicles* 4:27066. doi: 10.3402/jev.v4.27066.
- Yang, J., E. K. Kim, A. McDowell, and Y. K. Kim. 2018. "Microbe-derived extracellular vesicles as a smart drug delivery system." *Transl Clin Pharmacol* 26 (3):103-110. doi: 10.12793/tcp.2018.26.3.103.
- Yeshi, K., R. Ruscher, L. Hunter, N. L. Daly, A. Loukas, and P. Wangchuk. 2020. "Revisiting Inflammatory Bowel Disease: Pathology, Treatments, Challenges and Emerging Therapeutics Including Drug Leads from Natural Products." *J Clin Med* 9 (5). doi: 10.3390/jcm9051273.
- Yoon, M. Y., K. Lee, and S. S. Yoon. 2014. "Protective role of gut commensal microbes against intestinal infections." *J Microbiol* 52 (12):983-9. doi: 10.1007/s12275-014-4655-2.
- Yu, Z., C. Schwarz, L. Zhu, L. Chen, Y. Shen, and P. Yu. 2021. "Hitchhiking Behavior in Bacteriophages Facilitates Phage Infection and Enhances Carrier Bacteria Colonization." *Environ Sci Technol* 55 (4):2462-2472. doi: 10.1021/acs.est.0c06969.
- Zaborowski, M. P., L. Balaj, X. O. Breakefield, and C. P. Lai. 2015. "Extracellular Vesicles: Composition, Biological Relevance, and Methods of Study." *Bioscience* 65 (8):783-797. doi: 10.1093/biosci/biv084.
- Zamora, J. L. R., and H. C. Aguilar. 2018. "Flow virometry as a tool to study viruses." *Methods* 134-135:87-97. doi: 10.1016/j.ymeth.2017.12.011.
- Zhang, X., S. A. Deeke, Z. Ning, A. E. Starr, J. Butcher, J. Li, J. Mayne, K. Cheng, B. Liao, L. Li, R. Singleton, D. Mack, A. Stintzi, and D. Figeys. 2018. "Metaproteomics reveals associations between microbiome and intestinal extracellular vesicle proteins in pediatric inflammatory bowel disease." *Nat Commun* 9 (1):2873. doi: 10.1038/s41467-018-05357-4.
- Zolfo, M., F. Pinto, F. Asnicar, P. Manghi, A. Tett, F. D. Bushman, and N. Segata. 2019. "Detecting contamination in viromes using ViromeQC." *Nat Biotechnol* 37 (12):1408-1412. doi: 10.1038/s41587-019-0334-5.
- Zuo, T., X. J. Lu, Y. Zhang, C. P. Cheung, S. Lam, F. Zhang, W. Tang, J. Y. L. Ching, R. Zhao, P. K. S. Chan, J. J. Y. Sung, J. Yu, F. K. L. Chan, Q. Cao, J. Q. Sheng, and S. C. Ng. 2019. "Gut mucosal virome alterations in ulcerative colitis." *Gut* 68 (7):1169-1179. doi: 10.1136/gutjnl-2018-318131.
- Zwiehler, J., K. Liszt, M. Handschur, C. Lassl, A. Lapin, and A. G. Haslberger. 2009. "Combined PCR-DGGE fingerprinting and quantitative-PCR indicates shifts in fecal population sizes and diversity of Bacteroides, bifidobacteria and Clostridium cluster IV in institutionalized elderly." *Exp Gerontol* 44 (6-7):440-6. doi: 10.1016/j.exger.2009.04.002.

Titre : Bactériophages et vésicules extracellulaires du microbiote intestinal : développement de nouveaux outils d'étude et rôles dans les Maladies Inflammatoires Chroniques de l'Intestin (MICI)

Mots clés : phages, vésicules, Crohn, microbiote, ILM

Bien que répandues dans les pays développés, les maladies inflammatoires chroniques de l'intestin (MICI), qui incluent la maladie de Crohn, ne sont que partiellement comprises. Si leurs symptômes sont désormais bien connus, ces pathologies sont difficiles à diagnostiquer chez les patients, du fait de l'absence de biomarqueur et de notre méconnaissance concernant l'étiologie de celles-ci. Il est cependant communément admis que ces pathologies s'accompagnent de déséquilibres du microbiote intestinal, appelés dysbioses. Qu'elles soient à l'origine ou la conséquence des MICI, ces dysbioses, majoritairement étudiées par métagénomique, ne concernent pas uniquement les bactéries du microbiote intestinal. En effet, les phages, prédateurs naturels des bactéries, pourraient, comme quelques publications en font l'hypothèse, jouer un rôle dans la maladie de Crohn.

L'objectif principal de ma thèse aura été de dénombrer les phages, ainsi que les vésicules extracellulaires (VE), également potentiellement impactées au cours des MICI, dans les selles d'individus sains et de patients atteints de la maladie de Crohn, afin de mieux comprendre leurs rôles dans cette maladie.

Dans un premier temps, j'ai optimisé un protocole basé sur l'utilisation de la microscopie par épifluorescence pour le dénombrement des Virus Like Particles (VLPs), qui, dans le microbiote intestinal, correspondent en majorité à des phages. Les selles d'individus sains et de patients atteints de la maladie de Crohn en période de poussée ou de rémission ont ainsi été analysées. Le ratio VLPs/bactérie s'est ainsi révélé être 100 plus élevé chez les patients en poussée par rapport aux individus sains, suggérant que les phages intestinaux sont beaucoup plus nombreux chez les patients en poussée.

J'ai également mis au point un protocole de dénombrement des vésicules extracellulaires, en utilisant le Videodrop (Myriade), un nouveau dispositif optique basé sur l'Interferometric Light Microscopy (ILM). La mise au point de ce protocole de dénombrement aura été l'occasion de comparer les performances du Videodrop à deux appareils reposant sur l'utilisation d'une technologie proche, le Nanoparticle Tracking Analysis (NTA). Bien qu'ayant un seuil de détection plus élevé que les appareils NTA, ce qui fait qu'une partie des vésicules ne peut pas être détectée, le Videodrop s'est révélé être le plus pratique et le plus rapide d'utilisation des trois appareils testés. Une augmentation d'un facteur trois de la médiane des ratio VE/bactérie a été observé dans le groupe des patients en phase de poussée par rapport aux individus sains.

Les résultats obtenus au cours de cette thèse mettent en lumière l'augmentation du nombre de particules virales dans les selles des patients atteints de la maladie de Crohn au cours des phases de poussée. Plusieurs hypothèses, qui restent à être vérifiées, pourraient expliquer ces résultats. Premièrement, il est possible que l'inflammation intestinale soit inductrice de prophages initialement présents au sein du microbiote. D'autre part, une soudaine multiplication de virus dans le microbiote intestinal pourrait conduire à une poussée inflammatoire. Si cette augmentation du ratio VLPs/bactérie est spécifique de la maladie de Crohn, ce qui reste à investiguer, cela pourrait permettre, à terme, la mise au point d'un nouveau marqueur de cette pathologie.

Title : Bacteriophages and extracellular vesicles of the gut microbiota: development of new study tools and roles in Inflammatory Bowel Disease (IBD)

Keywords : phages, vesicles, Crohn, microbiota, ILM

Although widespread in developed countries, chronic inflammatory bowel diseases (IBD), which include Crohn's disease, are only partially understood. Although their symptoms are now well known, these pathologies are difficult to diagnose in patients, due to the absence of biomarkers and our lack of knowledge regarding their etiology. However, it is commonly accepted that these pathologies are accompanied by imbalances in the intestinal microbiota, called dysbiosis. Whether they are the origin or the consequence of IBD, these dysbiosis, mostly studied by metagenomics, do not only concern the bacteria of the intestinal microbiota. Indeed, phages, natural predators of bacteria, could, as some publications hypothesize, play a role in Crohn's disease.

The main objective of my thesis was to enumerate phages, as well as extracellular vesicles (EVs), also potentially impacted during IBD, in the stools of healthy individuals and patients with Crohn's disease, in order to better understand their roles in this disease.

First, I optimized a protocol based on the use of epifluorescence microscopy for the enumeration of Virus Like Particles (VLPs), which, in the intestinal microbiota, correspond in majority to phages. The stools of healthy individuals and of patients with Crohn's disease in relapse or remission were analyzed. The VLPs/bacteria ratio was found to be 100 times higher in the stools of patients facing a flare than in healthy individuals, suggesting that intestinal phages are much more numerous during the crises.

I also developed a protocol for enumerating extracellular vesicles, using the Videodrop (Myriade), a novel optical device based on Interferometric Light Microscopy (ILM). The development of this enumeration protocol was an opportunity to compare the performance of the Videodrop to two devices based on a similar technology, the Nanoparticle Tracking Analysis (NTA). Although it has a higher detection threshold than the NTA devices, which means that some of the vesicles cannot be detected, the Videodrop proved to be the most practical and fastest to use of the three devices tested. A three-fold increase in the median EVs/bacteria ratio was observed in the group of patients in the flare phase compared to healthy individuals.

The results obtained during this thesis highlight the increase in the number of viral particles in the stools of patients with Crohn's disease during the flare phases. Several hypotheses, which remain to be verified, could explain these results. First, it is possible that the intestinal inflammation is an inducer of prophages initially present in the microbiota. Second, a sudden increase in viruses in the gut microbiota could lead to an inflammatory phase. If this increase in the VLPs/bacteria ratio is specific to Crohn's disease, which remains to be investigated, this could eventually lead to the development of a new marker for this pathology.