



**HAL**  
open science

# Caractérisation génomique de virus d'archées et de bactéries au sein de procédés de digestion anaérobie par le couplage d'approches isotopiques et métagénomiques

Vuong Quoc Hoang Ngo

► **To cite this version:**

Vuong Quoc Hoang Ngo. Caractérisation génomique de virus d'archées et de bactéries au sein de procédés de digestion anaérobie par le couplage d'approches isotopiques et métagénomiques. *Ecologie, Environnement*. Université Paris-Saclay, 2022. Français. NNT : 2022UPASB045 . tel-03851180

**HAL Id: tel-03851180**

**<https://pastel.hal.science/tel-03851180v1>**

Submitted on 14 Nov 2022

**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.

Caractérisation génomique de virus d'archées et de bactéries au sein de procédés de digestion anaérobie par le couplage d'approches isotopiques et métagénomiques

*Genomic characterization of archaeal and bacterial viruses in anaerobic digestion processes, through the coupling of isotopic labeling and metagenomics*

**Thèse de doctorat de l'université Paris-Saclay**

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

Spécialité de doctorat : Sciences de l'environnement

Graduate School : Biosphera. Référent : •AgroParisTech

Thèse préparée dans l'unité de recherche **PROSE** (Université Paris-Saclay, INRAE), sous la direction de **Théodore BOUCHEZ** (ICPEF, HDR) et le co-encadrement d'**Ariane BIZE** (IPEF)

**Thèse soutenue à Paris-Saclay, le 4 Juillet 2022, par**

**Vuong Quoc Hoang NGO**

**Composition du Jury**

**Ludwig JARDILLIER**

Professeur, Université Paris-Saclay

Président

**Claire GESLIN**

Maîtresse de conférences (HDR), Université de Bretagne Occidentale

Rapporteur & Examinatrice

**Emilie MULLER**

Chargée de recherche (HDR), CNRS (Université de Strasbourg)

Rapporteur & Examinatrice

**Jean-Jacques GODON**

Directeur de recherche, INRAE (centre Occitanie-Montpellier)

Examineur

**Patrick DABERT**

Directeur de recherche, INRAE (centre Bretagne-Normandie)

Examineur

**Théodore BOUCHEZ**

ICPEF (HDR), INRAE (Université Paris-Saclay)

Directeur de thèse

**Titre :** Caractérisation génomique de virus d'archées et de bactéries au sein de procédés de digestion anaérobie par le couplage d'approches isotopiques et métagénomiques

**Mots clés :** Ecologie microbienne, virus, digestion anaérobie, marquage isotopique, métagénomique

**Résumé :** Les virus sont ubiquitaires et connus comme des régulateurs majeurs dans plusieurs écosystèmes. Cependant, les virus présents au sein des procédés de digestion anaérobie (DA) sont encore peu connus. Une première étape importante serait de mieux connaître leur diversité. L'objectif principal de la thèse a été de caractériser *in situ* le contenu génomique de virus au sein des bioprocédés de DA et de déployer des approches expérimentales et bioinformatiques pour prédire les hôtes cellulaires de ces virus. Cette étape de prédiction d'hôte reste actuellement un défi, en raison de la diversité et de la complexité de ces écosystèmes. Pour contribuer à lever ce verrou, une approche originale, basée sur du marquage isotopique (*Stable isotope probing*) et du séquençage haut-débit, a été développée. En appliquant cette approche avec du formate marqué au  $^{13}\text{C}$  employé comme substrat, les séries d'expériences menées à l'échelle du laboratoire ont abouti à la découverte de nouvelles familles de virus infectant les bactéries et les archées méthanogènes qui sont impliquées dans la méthanisation du formate. Un virus infectant un membre de la famille Synergistaceae a notamment été décrit pour la première fois, ainsi que deux virus fusiformes distincts infectant des membres des genres *Methanobacterium* et *Methanosarcina*. Ces travaux ont ainsi permis de développer significativement les connaissances de la diversité virale au sein de ces bioprocédés, et également de proposer des stratégies pour l'identification des interactions d'hôte-virus au sein de communautés microbiennes complexes.

**Title:** Genomic characterization of archaeal and bacterial viruses in anaerobic digestion processes, through the coupling of isotopic labeling and metagenomics

**Keywords:** Microbial ecology, viruses, anaerobic digestion, stable isotope probing, metagenomics

**Abstract:** Viruses are ubiquitous and known as major drivers in several ecosystems. However, the viruses from anaerobic digestion (AD) processes are still poorly characterized. A first important step would be to better study their diversity. The main objective of the PhD work was to characterize *in situ* the genomic content of viruses in AD processes and to deploy experimental and bioinformatic approaches to predict their cellular hosts. Predicting the host of viruses remains a challenge, due to the diversity and complexity of these ecosystems. To contribute to overcome these obstacles, an original approach based on isotopic labeling (*Stable isotope probing*) and high-throughput sequencing was implemented. By applying this approach with  $^{13}\text{C}$ -labeled formate, used as substrate, laboratory-scale experimental series have led to the discovery of new families of viruses infecting bacteria and methanogenic archaea which are involved in formate AD. In particular, a virus infecting a member of the family Synergistaceae was described for the first time, as well as two spindle-shaped viruses infecting *Methanobacterium* and *Methanosarcina* members. This work has thus significantly developed the knowledge of the viral diversity in AD processes, and it has also enabled to propose strategies for the identification of host-virus interactions within complex microbial communities.

# Remerciements

Je voudrais remercier tous ceux qui m'ont accompagné durant ces trois années de thèse, ainsi que durant ces dix années en France.

Je tiens tout d'abord à exprimer toute ma reconnaissance à mes encadrants de thèse Ariane Bize et Théodore Bouchez, qui, m'ont donné la chance de faire cette thèse, m'ont accueilli avec bienveillance et m'ont donné beaucoup de conseils non seulement pour l'avancement de la thèse mais aussi pour mon avenir.

Je tiens à remercier tous les membres du comité de suivi de ma thèse : François Enault, Mart Krupovic, Marie-Agnès Petit, Jérôme Hamelin, Olivier Chapleur, Laurent Mazéas, qui m'ont beaucoup aidé pour mener au succès mes travaux de thèse. Je voudrais également remercier tous les collaborateurs, les services analytiques avec qui j'ai passé de belles expériences.

Merci à tous les collègues de l'ancien Irstea, qui sont trop nombreuses pour être tous listés ici. Merci pour cette incroyable ambiance. Grâce à vous j'ai passé mes années de thèse sans stress !

Grand merci à ma famille d'accueil, Brigitte et Jean-Yves Plassard, les Français inconnus qui ont hébergé un gamin de 18 ans perdu en France lors de son premier jour en France. Merci de m'avoir toujours soutenu.

Merci à mes parents, mes amis d'être toujours à mon côté !

Ces 10 ans en France sont les meilleurs souvenirs de ma jeunesse !

# Table des matières

<b>REMERCIEMENTS.....</b>	<b>3</b>
<b>LISTES DES FIGURES .....</b>	<b>10</b>
<b>LISTES DES TABLEAUX.....</b>	<b>14</b>
<b>LISTE DES ABREVIATIONS.....</b>	<b>15</b>
<b>INTRODUCTION GENERALE .....</b>	<b>16</b>
<b>SYNTHESE BIBLIOGRAPHIQUE.....</b>	<b>18</b>
1. DIGESTION ANAEROBIE.....	18
1.1. Les différentes étapes .....	19
1.2. Influence des conditions physico-chimiques sur la digestion anaérobie.....	20
2. PROCARYOTES AU SEIN DU PROCESSUS DE DIGESTION ANAEROBIE .....	21
2.1. Hydrolyse .....	23
2.1. Acidogenèse .....	23
2.2. Acétogenèse .....	24
2.3. Méthanogenèse .....	24
a. Archaea : troisième domaine du vivant.....	24
b. Archées méthanogènes : acteurs clés de la production de méthane.....	28
c. Différentes voies de méthanogenèse.....	31
2.4. Métabolisme du formate : intérêt pour la bioraffinerie .....	33
3. VIRUS AU SEIN DU PROCESSUS DE DIGESTION ANAEROBIE .....	35
3.1. Virus : découverte et définition .....	35

3.2.	Diversité virale.....	36
a.	Diversité morphologique.....	36
b.	Diversité génomique et génétique.....	38
3.3.	Taxonomie et classification des virus .....	39
3.4.	Interactions hôte-virus.....	40
a.	Cycles viraux .....	40
b.	Mécanismes de défense de l'hôte .....	42
c.	Coévolution hôte-virus : course à l'armement.....	46
3.5.	Virus : régulateurs majeurs des écosystèmes .....	46
3.6.	Virus au sein du processus de digestion anaérobie.....	47
a.	Etat de l'art.....	47
b.	Virus d'archées méthanogènes .....	49
4.	LES METHODES MOLECULAIRES POUR L'ÉCOLOGIE MICROBIENNE .....	55
4.1.	Marquage isotopique.....	56
4.2.	Approches métagénétiques et méta-omiques .....	58
a.	Le gène d'ARN ribosomique 16S : marqueur universel des procaryotes ....	58
b.	Les approches méta-omiques.....	60
4.3.	Focus sur la prédiction d'hôte pour les virus.....	62
a.	Approches expérimentales.....	63
a.	Approches bioinformatiques.....	65
	<b>PRESENTATION GENERALE DES TRAVAUX DE THESE .....</b>	<b>70</b>
	<b><u>MATERIELS ET METHODES.....</u></b>	<b><u>71</u></b>
1.	SYSTEME EXPERIMENTAL .....	71

1.1.	Phase liquide .....	72
a.	Inoculum .....	72
b.	Substrat .....	72
c.	Solution minérale.....	73
1.2.	Phase gazeuse.....	74
1.3.	Mise en incubation des microcosmes .....	74
1.4.	Suivi, prélèvement et conservation des échantillons pendant l'incubation ..	74
2.	PREPARATION DES PARTICULES VIRALES ISSUES DES MICROCOSMES DE DIGESTION ANAEROBIE .....	76
3.	OBSERVATION DES PARTICULES VIRALES.....	77
4.	METHODES D'ANALYSES PHYSICO-CHIMIQUES.....	77
4.1.	pH.....	77
4.2.	Demande chimique en oxygène (DCO).....	77
4.3.	Matière sèche/Matière volatile (MS/MV).....	77
4.4.	Carbone organique/inorganique dissous (COD/CID).....	78
a.	Analyse du COD .....	78
b.	Analyse du CID .....	79
4.5.	Acides gras volatils (AGV) et autres acides organiques.....	79
4.6.	Analyse de composition des gaz.....	79
4.7.	Analyse de la composition isotopique des gaz .....	79
5.	METHODES DE BIOLOGIE MOLECULAIRE.....	81
5.1.	Extraction d'ADN cellulaire.....	81
5.2.	Extraction d'ADN viral .....	81
5.3.	Dosage d'ADN .....	82

5.4.	SIP-ADN .....	82
5.5.	Amplification des ADN prélabable au séquençage « shotgun ».....	83
5.6.	Métabarcoding basé sur le gène d'ARNr 16S .....	83
5.7.	Métagénomique <i>shotgun</i> .....	84
6.	ANALYSES BIOINFORMATIQUES.....	85
6.1.	Traitement bioinformatique des séquences de metabarcoding d'ARNr 16S	85
6.1.	<i>Pipeline</i> bioinformatique d'analyses de métagénomique <i>shotgun</i> .....	86
7.	ANALYSES STATISTIQUES.....	89
8.	SYSTEME MODELE SUR CULTURES DE SOUCHES PURES.....	90
8.1.	Mise en culture d' <i>E. coli</i> et infection virale par le bactériophage T4 .....	90
8.2.	Préparation des particules virales .....	91
8.3.	Titration sur tapis bactérien .....	91
8.4.	Extraction et dosage d'ADN.....	91
8.1.	SIP ADN .....	92
8.2.	Analyses IRMS.....	92
<b><u>CHAPITRE 1 : FOCUS SUR LES VIRUS D'ARCHEES METHANOGENES.....</u></b>		<b>93</b>
1.	INTRODUCTION .....	93
1.	EXPERIENCES TEST POUR GUIDER LE CHOIX D'UN SUBSTRAT D'INTERET.....	93
1.1.	Plan expérimental .....	93
1.2.	Résultats.....	95
a.	Test des paramètres opératoires .....	95
b.	Test des substrats .....	95
1.3.	Conclusions.....	97



2.	PRESENTATION DU TRAVAIL DECRIT DANS CE CHAPITRE.....	97
	<b>DIVERSITY OF NOVEL ARCHAEL VIRUSES INFECTING METHANOGENS DISCOVERED THROUGH COUPLING OF STABLE ISOTOPE PROBING AND METAGENOMICS.....</b>	<b>98</b>
	<b><u>CHAPITRE 2 : VIRUS INFECTANT DES MICROORGANISMES IMPLIQUES DANS LA METHANISATION DU FORMATE .....</u></b>	<b><u>149</u></b>
	<b>DIVERSITY OF NOVEL VIRUSES IN ANAEROBIC DIGESTION MICROCOSMS FED WITH FORMATE, DISCOVERED THROUGH COUPLING OF STABLE ISOTOPE PROBING AND METAGENOMICS.....</b>	<b>151</b>
	<b><u>CHAPITRE 3 : PREUVE DE CONCEPT DE L'UTILISATION DU SIP POUR LES ADN VIRAUX..</u></b>	<b><u>195</u></b>
1.	INTRODUCTION .....	195
2.	PLAN EXPERIMENTAL .....	195
3.	RESULTATS ET DISCUSSIONS .....	196
3.1.	Mise en évidence du marquage au <sup>13</sup> C.....	196
a.	Marquage des cellules .....	196
b.	Marquage des virus .....	197
3.2.	Cohérence entre la proportion de substrat marqué utilisée, et le taux d'enrichissement en <sup>13</sup> C des ADN cellulaires et viraux.....	199
a.	Marquage des cellules .....	199
b.	Marquage des virus .....	200
3.3.	Perspectives d'utilisation du SIP appliqué à l'ADN viral dans le contexte de la microbiologie environnementale .....	202
4.	CONCLUSIONS ET PERSPECTIVES .....	204
	<b><u>DISCUSSION ET PERSPECTIVES.....</u></b>	<b><u>205</u></b>
1.	VIRUS AU SEIN DES PROCEDES DE DA.....	205

2.	DEVELOPPEMENT METHODOLOGIQUE POUR LA METAGENOMIQUE VIRALE.....	208
3.	PERSPECTIVES .....	209
3.1.	Mise en évidence des nouveaux virus découverts grâce aux analyses bioinformatiques .....	209
3.2.	SIP virus : une approche prometteuse en écologie virale.....	210
3.3.	Approfondissement des recherches sur les virus d'archées méthanogènes : expériences avec du <sup>13</sup> C-méthanol.....	211
3.4.	Diversité virale dans les installations industrielles de DA .....	211
4.	CONCLUSION.....	<b>ERROR! BOOKMARK NOT DEFINED.</b>

**BIBLIOGRAPHIE .....** **214**

**VALORISATIONS SCIENTIFIQUES.....** **234**

1.	PUBLICATIONS .....	234
2.	COMMUNICATIONS .....	234
3.	ENSEIGNEMENT .....	235
4.	FORMATIONS.....	235

# Listes des Figures

Figure 1. Les principales étapes du processus de digestion anaérobie .....	18
Figure 2. Illustration de quelques environnements terrestres abritant des archées. ....	26
Figure 3. Nombres de génomes répertoriés dans les différents super-phyla et phyla d'archées.....	27
Figure 4. Flux globaux du méthane et distribution des méthanogènes.....	28
Figure 5. Phylogénie des archées basée sur les gènes mcr et mcr-like.....	29
Figure 6. Phylogénie des archées basée sur 36 gènes marqueurs .....	30
Figure 7. Trois voies distinctes de la méthanogenèse.....	31
Figure 8. La bioéconomie du formate .....	33
Figure 9. Différentes voies métaboliques d'utilisation du formate au cours de la DA .....	34
Figure 10. Structures de virus présentant différents types de symétrie .....	37
Figure 11. Diversité morphologique des virus .....	37
Figure 12. Dogme central de la biologie moléculaire. ....	38
Figure 13. Classification Baltimore récente des virus.....	39
Figure 14. Différents types de cycles viraux .....	42
Figure 15. Stratégies de défense des hôtes procaryotes contre les virus .....	43
Figure 16. Rôles des protéines argonautes (Ago).....	44
Figure 17. Mode d'action du système CRISPR-Cas.....	45
Figure 18. Aperçu de la diversité des particules virales au sein des méthaniseurs .....	49
Figure 19. $\Psi$ M1 observé au MET .....	51
Figure 20. Observations au MET de virions infectant des membres du genre <i>Methanobacterium</i> : $\Phi$ F1, $\Phi$ F3 et Drs3. ....	52
Figure 21. Virions de MetSV observés au MET .....	53
Figure 22. Virions de Blf4 observés au MET .....	54
Figure 23. Particule A3-VLP observée au MET. ....	54
Figure 24. Méthodes culture-indépendantes pour l'étude des communautés microbiennes complexes.....	56

Figure 25. Illustration de l'approche de marquage isotopique.....	57
Figure 26. Tubes contenant de l'ADN <sup>12</sup> C et <sup>13</sup> C ultracentrifugés dans une solution de CsCl, mettant en évidence leur différence de densité.....	58
Figure 27. L'arbre des trois domaines du vivant.....	59
Figure 28. Différents types d'approches méta-omiques pour l'étude des communautés microbiennes.....	60
Figure 29 Techniques d'analyse de métaprotéomique.....	62
Figure 30 Chronologie globale des approches bioinformatiques et expérimentales utilisées pour établir les liens hôtes-virus.....	63
Figure 31. Système expérimental pour les incubations en microcosmes.....	71
Figure 32. Illustration schématique des prélèvements d'échantillons liquides et de leurs analyses.....	76
Figure 33. Système de couplage de chromatographie en phase gazeuse-combustion-spectrométrie de masse isotopique.....	80
Figure 34. Pipeline bioinformatique de FROGS.....	85
Figure 35. Etapes d'analyses bioinformatiques communes aux métagénomés cellulaires et métaviromes.....	87
Figure 36. Etapes d'analyses bioinformatiques spécifiques aux métagénomés cellulaires.....	88
Figure 37. Etapes d'analyses bioinformatiques spécifiques aux métaviromes.....	89
Figure 38. Productions cumulées de biogaz lors des tests de différents paramètres opératoires.....	95
Figure 39. Productions cumulées de biogaz obtenues lors du test de différents substrats.....	96
Figure 40. Composition des communautés microbiennes lors des incubations de test de différents substrats (metabarcoding d'ADNr 16S).....	96
Figure 41. Profils de densité des ADNs du bactériophage T4 obtenus à partir des incubations avec du <sup>12</sup> C-glucose ou du <sup>13</sup> C-glucose.....	198
Figure 42. Profils de densité des ADNs du bactériophage T4, normalisés, obtenus à partir des incubations avec du <sup>12</sup> C-glucose ou du <sup>13</sup> C-glucose.....	199
Figure 43. Taux d'enrichissement en <sup>13</sup> C dans les cellules en fonction du pourcentage de <sup>13</sup> C-glucose utilisé.....	200

Figure 44. Profils de densité des ADNs du bactériophage T4, normalisés, obtenus à partir des incubations avec différentes proportions de <sup>13</sup> C-glucose .....	201
Figure 45. Densité observée pour les ADNs du bactériophage T4, en fonction de la densité théorique calculée à partir du contenu en GC du génome viral et du pourcentage de <sup>13</sup> C-glucose utilisé.....	202
Figure 46. Profil de densité, normalisé, d'un mélange d'ADNs non marqués et marqués à 100% au <sup>13</sup> C, du bactériophage T4.....	203
Figure 47. Intervalles de densité attendus pour les ADNs viraux, dans le cas de l'application du SIP à des échantillons environnementaux.....	204
Figure 48. Particules et structures indéfinies présentes dans les digesteurs anaérobies de laboratoire .....	206

**Article 1: DIVERSITY OF NOVEL ARCHAEAL VIRUSES INFECTING METHANOGENS DISCOVERED THROUGH COUPLING OF STABLE ISOTOPE PROBING AND METAGENOMICS**

Figure 1. Microbial dynamics during AD incubation.. .....	110
Figure 2. Morphotypic diversity of VLPs in different microcosms, observed by TEM.....	112
Figure 3. Strategy for the selection of contigs of interest, likely originating from viruses of methanogens .....	117
Figure 4. Bipartite network of known archaeal viruses and of the 39 contigs of interest, and of protein orthologous groups (OGs).. .....	120
Figure 5. Genome map for selected contigs of interest. ....	124
Figure S1. Dynamic of AD process. ....	142
Figure S2. Metabolic pathway analyses of cellular metagenomic data for the category energy metabolism.. .....	143
Figure S3. Taxonomic affiliation of the reads from the cellular metagenome dataset ....	144
Figure S4. Density profiles of the cellular DNAs collected from microcosms.....	145
Figure S5. Principal component analysis of the 39 contigs of interest obtained with VIBRANT.....	146
Figure S6. Bipartite network of known archaeal viruses and 39 contigs of interest, grouped into viral clusters (VC), and of protein orthologous groups (OGs).....	147

**Article 2: DIVERSITY OF NOVEL VIRUSES IN ANAEROBIC DIGESTION MICROCOSMS FED WITH FORMATE, DISCOVERED THROUGH COUPLING OF STABLE ISOTOPE PROBING AND METAGENOMICS**

Figure 1. Microbial dynamics during AD incubation. .... 163

Figure 2. Morphotypic diversity of VLPs in different microcosms, observed by TEM..... 167

Figure 3. Taxonomic affiliation of the various datasets. .... 171

Figure 4. Strategy for the selection of contigs of interest ..... 173

Figure 5. Heatmaps of the viral contigs of interest ..... 175

Figure 6. Map of 6 contigs selected for the reliability of their host prediction. .... 178

Figure S1. Formate and acetate concentration during the incubation ..... 194

Figure S2. Taxonomic assignation of the microorganisms involved in energy metabolism ..... 194

# Listes des tableaux

<a href="#">Tableau1 Réactions possible à différentes étapes de la DA.....</a>	20
<a href="#">Tableau 2. Groupes fonctionnels microbiens impliqués dans les différentes étapes de la DA.....</a>	21
<a href="#">Tableau 3 Microorganismes acteurs de la DA, (d'après (Harirchi et al., 2022)).....</a>	22
<a href="#">Tableau 4. Quelques caractéristiques clé des trois domaines du vivant .....</a>	26
<a href="#">Tableau 5. Virus et provirus d' archées méthanogènes .....</a>	50
<a href="#">Tableau 6. Différentes stratégies pour la prédiction d'hôte des virus.....</a>	66
<a href="#">Tableau7. Différents outils pour la prédiction d'hôte des virus .....</a>	68
<a href="#">Tableau 8. Caractéristiques de l'inoculum préparé .....</a>	72
<a href="#">Tableau 9. Milieu BMP 1X et solution mère d'éléments traces 100X</a> <b>Error! Bookmark not defined.</b>	
<a href="#">Tableau 10. Plan expérimental des incubations de test de différents paramètres opératoires .....</a>	94
<a href="#">Tableau11. Caractéristiques des méthaniseurs échantillonnés.....</a>	212
<b>Article 1: DIVERSITY OF NOVEL ARCHAEAL VIRUSES INFECTING METHANOGENS DISCOVERED THROUGH COUPLING OF STABLE ISOTOPE PROBING AND METAGENOMICS</b>	
<a href="#">Table 1. Overview of the shotgun metagenome datasets.....</a>	115
<a href="#">Table S1. Synthetic view on the viruses of methanogenic archaea isolated so far.....</a>	139
<a href="#">Table S2. Bioinformatic tool parameters and database versions .....</a>	140
<b>Article 2: DIVERSITY OF NOVEL VIRUSES IN ANAEROBIC DIGESTION MICROCOSMS FED WITH FORMATE, DISCOVERED THROUGH COUPLING OF STABLE ISOTOPE PROBING AND METAGENOMICS</b>	
<a href="#">Table 1. Selection of possible reaction for formate consumption in AD process .....</a>	162
<a href="#">Table.2. General overview of the shotgun metagenome datasets.....</a>	169
<a href="#">Table3. Abbreviations for protein annotation of the six contigs of interest .....</a>	179
<a href="#">Table S1. Bioinformatic tool parameters and database versions .....</a>	190
<a href="#">Table S2. Results of PERMANOVA analyses.....</a>	193

## Liste des abréviations

ADN (DNA) : acide désoxyribonucléique

AGV: Acide gras volatil

AMG : *auxiliary metabolic gene*

ARN (RNA): acide ribonucléique

CID: Carbone inorganique dissous

COD: Carbone organique dissous

CRISPR: *Clustered Regularly Interspaced Short Palindromic Repeats*

DA (AD): digestion anaérobie

db (ds): double brin (*double stranded*)

DCO: Demande chimique en Oxygène

FISH: *Fluorescence in situ hybridization*

HPLC : *High performance liquid chromatography* (chromatographie en phase liquide)

IRMS : *Isotope Ratio Mass Spectrometer*

MAG : *Metagenome assembled genome*

MET (TEM): Microscope électronique en transmission (*transmission electron microscopy*)

MS: Matière sèche

MV: Matière volatile

NGS: *Next generation sequencing*

OTU: *Operational Taxonomic Units*

PCR: *Polymerization chain reaction*

sb (ss): simple brin (*single stranded*)

SIP: *Stable isotope probing* (marquage isotopique)

VLP: *Virus-like particle*



# Introduction générale

Chaque année, 4,6 tonnes de déchets (ADEME, Le Mag n°125) sont produites en France par habitant. Ces derniers causent des impacts importants sur l'environnement, l'économie et aussi la vie sociale. La gestion de ces déchets est aujourd'hui une problématique environnementale majeure.

Les déchets organiques représentent une proportion non négligeable dans la quantité de déchets produits. Ces biodéchets sont donc une ressource d'intérêt pour la production de molécules bio-sourcées par des bioprocédés de digestion anaérobie (DA), ou méthanisation. Cette dernière est un processus naturel consistant à transformer la matière organique en méthane, et impliquant l'action catalytique de communautés microbiennes complexes (Moletta, 2015). Ce processus a été appliqué dans différents domaines et s'est initialement développé en France autour du traitement des effluents dans les industries agro-alimentaires, du traitement des boues d'épuration d'eaux usées et du traitement des déchets agricoles. Par ailleurs, de nombreux digesteurs de production de « biogaz » à différentes échelles ont été installés dans des fermes au Danemark, en Italie, au Royaume-Uni et notamment en Allemagne. En France, il y avait 500 installations de méthanisation en fonctionnement en 2017, 1180 aujourd'hui et 1700 sont prévues pour 2023 (ADEME, Le Mag n°107 et Chiffres clés 2021). Ces installations ont produit annuellement plus de 880 Gigawattheure (GWh) d'électricité, 1400 GWh de chaleur et 215 GWh de biométhane (ADEME, Le Mag n°107). En général, les grands pays européens ont largement développé ou développent la méthanisation comme un élément qui devient incontournable dans leur politique de production d'énergies renouvelables. Les connaissances sur ces bioprocédés de DA, notamment sur les communautés microbiennes complexes présentes, sont donc indispensables pour assurer le bon fonctionnement de ces procédés, ainsi que pour atteindre de meilleurs rendements.

La transformation de la matière organique résiduaire au cours de la digestion anaérobie est catalysée par une communauté microbienne diversifiée dont la composition et la dynamique dépendent de nombreux facteurs biotiques et abiotiques. Si l'avènement et la

montée en puissance des techniques moléculaires au cours des dernières décennies ont permis de mieux caractériser les populations de procaryotes (bactéries et archées) impliquées dans le processus, la composante virale de la communauté microbienne reste aujourd'hui très mal connue. Or les virus sont ubiquitaires et connus comme des régulateurs majeurs des équilibres écologiques dans plusieurs écosystèmes. Récemment, une grande diversité virale a été observée au sein des bioprocédés de méthanisation. Quels rôles jouent-ils dans ces bioprocédés ?

L'objectif principal de la thèse était de caractériser *in situ* le contenu génomique de virus présents dans des bioprocédés de digestion anaérobie, et d'essayer de décrypter leurs liens avec les populations procaryotes. Il est important de souligner qu'établir le lien entre un virus et l'identité de son ou de ses hôte(s) au sein d'un écosystème complexe constitue un défi majeur pour l'écologie microbienne. Pour lever ce verrou, le lien entre les voies métaboliques actives, les communautés microbiennes exprimant ces voies, et le contenu génomique des populations virales, a été établi grâce à une approche intégrée originale couplant des techniques de pointe en écologie moléculaire (métagénomique, isotopie, en particulier « *Stable isotope probing* » SIP) et des analyses *in silico* classiques (génomique comparative). L'utilisation de formate marqué au  $^{13}\text{C}$ , en microcosmes de méthanisation, a permis d'activer des sous-populations impliquées dans la méthanisation de ce substrat, notamment des archées méthanogènes, et d'enrichir leur ADN en  $^{13}\text{C}$ . Les ADN cellulaires ont été collectés et séparés selon leur densité. Ainsi, les fractions lourdes des ADN cellulaires, ainsi que les métaviromes des microcosmes correspondants, ont été séquencés par métagénomique *shotgun*. La diversité des familles virales et les gènes dont les fonctions sont liées au cycle de vie des virus ont été particulièrement examinés.

Bien que s'inscrivant dans un contexte biotechnologique, ce projet de thèse relève donc d'une recherche *amont*, qui vise à donner des éléments pour la compréhension de la communauté virale présente dans les digesteurs. Nous espérons toutefois que ces éléments permettront de contribuer à terme à une meilleure gestion des installations de digestion voire à l'élaboration de stratégies innovantes de biocontrôle.

# Synthèse bibliographique

## 1. Digestion anaérobie

La digestion anaérobie (DA), ou méthanisation, est un processus biologique naturel permettant de convertir des matières organiques complexes (glucides, lipides, protéines) en molécules simples ( $\text{CH}_4$ ,  $\text{CO}_2$ ,  $\text{H}_2\text{S}$ , etc), en absence d'oxygène, grâce à l'action de communautés microbiennes complexes. Ce processus, qui implique de nombreux métabolismes, est un bel exemple de chaîne trophique. Il comprend quatre étapes successives : hydrolyse, acidogénèse, acétogénèse, méthanogénèse (Figure 1).

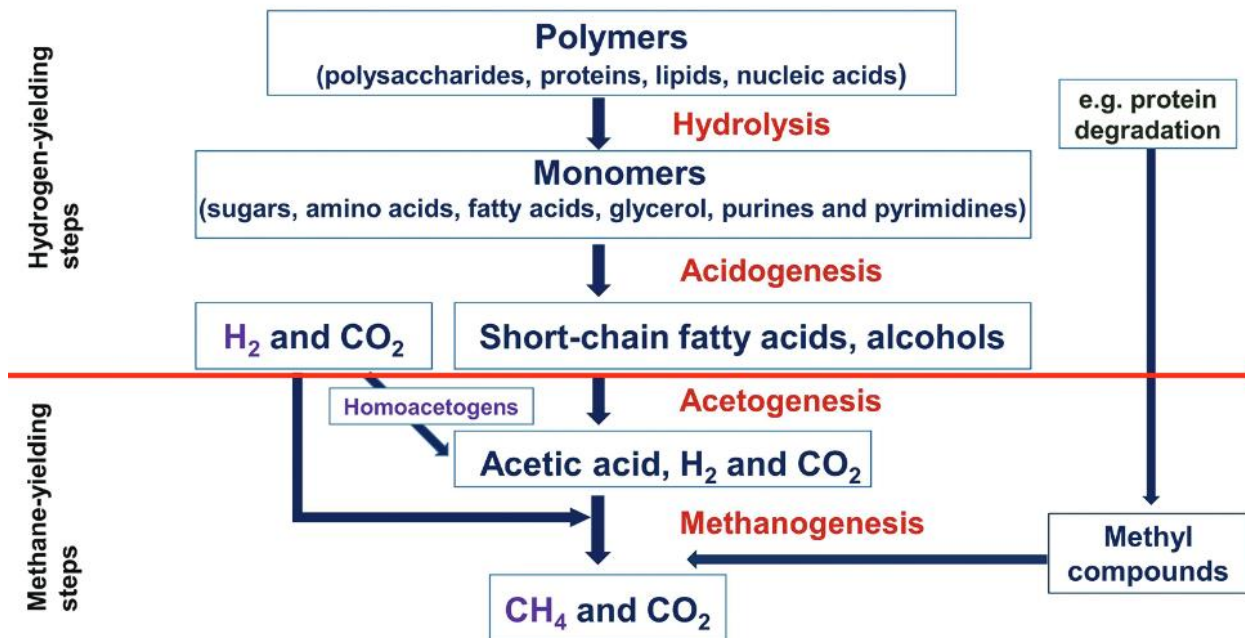


Figure1. Les principales étapes du processus de digestion anaérobie (Sikora *et al.*, 2017)

Décrit pour la première fois par Alessandro Volta en 1776 avec la découverte de « l'air inflammable des marais », ce processus a été appliqué à l'échelle industrielle depuis le siècle dernier, pour la valorisation de déchets organiques et de boues issues du traitement des eaux usées.

## 1.1. Les différentes étapes

L'étape d'hydrolyse consiste en la décomposition des matières organiques complexes (polysaccharides, lipides, protéines) en composés solubles plus simples (sucres, acides gras, acides aminés) à l'aide d'exoenzymes. Ces molécules organiques solubles sont ensuite métabolisées, pendant l'étape d'acidogénèse, en acide gras volatiles (AGV) - les acides organiques à courtes chaînes de deux à cinq carbones (formique, acétique, propionique, butyrique, pentanoïque), alcool (méthanol, éthanol), aldéhydes, dioxyde de carbone et hydrogène (Tableau 1). Suite à cette étape dite de «fermentation classique», les monomères organiques peuvent continuer à être dégradées en acétate *via* l'acétogénèse qui consiste en la production d'acétate à partir des autres acides organiques, ainsi que de H<sub>2</sub> et CO<sub>2</sub> (voie d'homoacétogénèse, aussi appelée acétogénèse chimolithoautotrophe) (Tableau 1). Dernière d'étape du processus de la digestion anaérobie, la méthanogénèse est un processus de production de méthane (CH<sub>4</sub>) à partir des différents substrats issus des étapes précédentes, par trois principales voies métaboliques : hydrogénotrophe (substrat : CO<sub>2</sub>), acétoclaste (substrat : acétate) et méthylotrophe (substrats : composé méthylés) (Lyu *et al.*, 2018) (Tableau 1).

Par ailleurs, lors de la DA, des réactions de type sulfato-réduction peuvent également se produire sous l'action de bactéries sulfato-réductrices (Zhou and Fang, 1998; Fdz-Polanco *et al.*, 2001), en présence de sulfate. Ces réactions aboutissent à la transformation d'alcools ou d'acides gras en H<sub>2</sub>S et CO<sub>2</sub> en utilisant le sulfate comme accepteur final d'électron. En particulier, les bactéries sulfato-réductrices ont la capacité d'utiliser le dihydrogène et l'acide acétique, entrant ainsi en concurrence avec les réactions de méthanogénèse. D'autres réactions peuvent également avoir lieu, faisant notamment intervenir l'azote (Fdz-Polanco *et al.*, 2001).

Tableau1. Réactions possible à différentes étapes de la DA, modifié à partir de (Moletta, 2015)

Réactions	$\Delta G_0'$ kJ/réaction
<b>Hydrolyse et acidogénèse</b>	
$C_6H_{12}O_6 + 2 H_2O \rightarrow 2 \text{ éthanol} + 2 HCO_3^- + 2 H^+$	-225,4
$C_6H_{12}O_6 \rightarrow 2 \text{ lactate}^- + 2 H^+$	-198,1
$C_6H_{12}O_6 + 2H_2O \rightarrow \text{butyrate}^- + 2 HCO_3^- + 3 H^+ + 2 H_2$	-254,4
$C_6H_{12}O_6 \rightarrow 3 \text{ acétate}^- + 3 H^+$	-310,6
$C_6H_{12}O_6 + HCO_3^- \rightarrow \text{succinate}^{2-} + \text{acetate}^- + \text{formate}^- + 3H^+ + H_2O$	-144,0
<b>Acétogénèse</b>	
$\text{Lactate}^- + 2H_2O \rightarrow \text{acétate}^- + 2 H_2 + 3 HCO_3^- + H^+$	-3,9
$\text{Ethanol} + 2 HCO_3^- \rightarrow \text{acétate}^- + 2 \text{ formate}^- + H_2O + H^+$	+7,0
$\text{Ethanol} + 2H_2O \rightarrow \text{acétate}^- + H^+ + 2 H_2$	+9,6
$\text{Butyrate}^- + 2 H_2O \rightarrow 2 \text{ acétate}^- + 3 H^+ + 2 H_2$	+48,1
$\text{Propionate}^- + 3 H_2O \rightarrow \text{acétate}^- + H^+ + 3 H_2 + HCO_3^-$	+76,1
<b>Homoacétogénèse</b>	
$H^+ + 4 H_2 + 2HCO_3^- \rightarrow \text{acétate}^- + 4 H_2O$	-104,6
<b>Méthanogénèse</b>	
$H^+ + 4 H_2 + HCO_3^- \rightarrow CH_4 + 3 H_2O$	-135,6
$\text{acétate}^- + H_2O \rightarrow CH_4 + HCO_3^-$	-31,0
$4 \text{ formate}^- + H_2O + H^+ \rightarrow CH_4 + 3 HCO_3^-$	-32,5
$\text{méthanol} + H_2 \rightarrow CH_4 + H_2O$	-112,5
$4 \text{ méthanol} \rightarrow 3CH_4 + HCO_3^- + H^+ + H_2O$	-78,6
$CO + 3 H_2 \rightarrow CH_4 + H_2O$	-142,0
$2 CO + 2 H_2 \rightarrow CH_4 + CO_2$	-171,0

## 1.2. Influence des conditions physico-chimiques sur la digestion anaérobie

Pendant plusieurs décennies, l'effet de plusieurs paramètres physico-chimiques sur la stabilité du procédé de digestion anaérobie, tels que le pH, la température, le rapport C/N de l'intrant ou la charge volumique appliquée, a été étudié afin de déterminer les plages permettant une bonne efficacité de la digestion anaérobie (Kim *et al.*, 2002, 2015; Chen *et al.*, 2008; Mao *et al.*, 2015; Poirier *et al.*, 2017, 2018). Lors de variations trop brutales et/ou trop importantes de ces paramètres, des dysfonctionnements de la digestion anaérobie peuvent apparaître. Ces variations constituent en effet des variations de biotope qui, lorsqu'elles affectent l'activité de microorganismes fonctionnels clés du processus, peuvent conduire à des pertes de rendement ou de productivité en méthane. Il est donc

essentiel de mieux comprendre ces communautés microbiennes complexes et les facteurs qui induisent leurs variations dans l'espace et le temps.

## 2. Microorganismes au sein du processus de digestion anaérobie

La digestion résulte d'une coopération de différents groupes fonctionnels microbiens, dont chacun a son rôle particulier dans le processus global de dégradation (Tableau 2). Des bactéries et des archées sont impliquées dans ce processus (Tableau 3).

Tableau 2. Groupes fonctionnels microbiens impliqués dans les différentes étapes de la DA

Etape	Microorganismes	Donneur d'électron	Accepteur d'électron	Produit(s) principaux
<b>Hydrolyse/ Acidogénèse</b>	Champignons hydrolytiques, bactéries hydrolytiques et fermentatives (acidogènes)	Carbone organique	Carbone organique	Acides gras volatils CO <sub>2</sub> + H <sub>2</sub>
<b>Acétogénèse</b>	Bactéries syntrophiques (productrices de H <sub>2</sub> )	Carbone organique	Carbone organique	H <sub>2</sub>
	Bactéries acétogènes (consommatrices de H <sub>2</sub> )	Carbone organique/H <sub>2</sub>	CO <sub>2</sub>	Acétate
<b>Méthanogénèse</b>	Archées méthanogènes	Carbone organique/H <sub>2</sub>	CO <sub>2</sub>	CH <sub>4</sub>

Tableau 3 Microorganismes acteurs de la DA, (d'après (Harirchi et al., 2022))

Phase de la DA	Microorganismes	Genres microbiens impliqués	Exemples d'espèces identifiées
Hydrolyse et Acidogénèse	Bactéries	<i>Acetivibrio, Aminobacterium, Aminomonas, Anaeromusa, Anaerosphaera</i> <i>Bacillus, Bacteroides, Bifidobacterium, Butyrivibrio</i> <i>Caldanaerobacter, Caldicellulosiruptor, Campylobacter, Cellulomonas, Clostridium</i> <i>Devosia</i> <i>Espiroquetas, Eubacterium</i> <i>Fervidobacterium, Fibrobacter, Fusobacterium</i> <i>Gelria, Gracilibacter</i> <i>Halocella</i> <i>Lactobacillus</i> <i>Paludibacter, Peptococcus, Peptoniphilus, Proteiniborus, Pseudomonas, Psychrobacter</i> <i>Ralstonia, Ruminoclostridium, Ruminococcus</i> <i>Selenomonas, Shewanella, Sporeanaerobacter, Streptococcus, Streptomyces</i> <i>Thermanaerovibrio, Thermomonas, Thermomonospora, Thermotoga, Treponema, Trichococcus</i>	<i>Pseudomonas mendocina</i> <i>Bacillus halodurans</i> <i>Clostridium hastiforme</i> <i>Gracilibacter thermotolerans</i> <i>Thermomonas haemolytica</i>
	Champignons	<i>Aspergillus</i> <i>Humicola</i> <i>Penicillium</i> <i>Trichoderma</i>	<i>Trichoderma reesei</i>
Acétogénèse	Bactéries	<i>Acetobacterium</i> <i>Clostridium</i> <i>Desulfotignum</i> <i>Eubacterium</i> <i>Holophaga</i> <i>Moorella</i> <i>Ruminococcus</i> <i>Thermoanaerobacter, Treponema</i> <i>Syntrophomonas</i> <i>Syntrophobacter</i>	<i>Moorella thermoacetica</i> <i>Desulfotignum phosphitoxidans</i> <i>Holophaga foetida</i>
Méthanogénèse	Archées	<i>Methanobacterium, Methanobrevibacter, Methanococcus, Methanoculleus, Methanosaeta,</i> <i>Methanomicrobium, Methanosarcina, Methanospirillum, Methanothermobacter</i>	<i>Methanobrevibacter smithii</i> <i>Methanobrevibacter arboriphilus</i> <i>Methanococcus vannielii</i>

## 2.1. Hydrolyse

Au cours de cette phase, les bactéries hydrolytiques dégradent la matière organique complexe (les polymères) en fragments solubles (monomères). Ces bactéries ont un métabolisme de type anaérobie strict ou facultatif et comprennent d'une part de nombreux groupes bactériens, notamment les genres *Streptococcus* et *Enterobacterium* (Bryant, 1979; Harirchi *et al.*, 2022), et d'autre part des champignons anaérobies (Kazda *et al.*, 2014; Harirchi *et al.*, 2022). La biodiversité des microorganismes est extrêmement importante, et les vitesses de croissance relativement rapides avec des temps de doublement de l'ordre de quelques heures parfois (Chapleur, 2012). Les microorganismes hydrolytiques sont généralement fermentatifs. Pendant la digestion des déchets organiques, seulement 50% des composés organiques sont métabolisés/biodégradés à ce stade en moyenne, les composés restants demeurent sous leur état primaire, par manque d'enzymes exocellulaires (Ziemiński and Frąç, 2012). Le rendement de cette étape dépend de différents facteurs : la température, le pH, la production des exo-enzymes par les microorganismes, etc. L'hydrolyse dépend aussi de la nature des déchets, notamment dans le cas de composés difficilement hydrolysables tels que la cellulose ou les graisses (Eastman and Ferguson, 1981; Pavlostathis and Giraldo-Gomez, 1991).

## 2.1. Acidogenèse

La flore microbienne de cette étape est constituée pour partie de bactéries anaérobies facultatives (genres *Bacillus*, *Pseudomonas*, *Micrococcus*, etc). Ces bactéries utilisent l'oxygène résiduel dans le processus et créent un environnement favorable pour la croissance des bactéries acidifiantes de type anaérobie stricte, comme les genres *Clostridium* et *Bacteroides*. Les produits formés à cette étape sont principalement les AGVs, incluant l'acétate, le formate, le propionate, l'isobutyrate, le butyrate, le valérate, et isovalérate. D'autres produits peuvent également être produits, tels que des alcools, H<sub>2</sub> et CO<sub>2</sub>. Parmi ces produits, certains peuvent être utilisés directement comme source d'énergie et source de carbone par les archées méthanogènes, pour la production de méthane. En général, l'acidogenèse est assez rapide, ce qui peut entraîner une



accumulation d'AGVs et une baisse du pH qui rendent les conditions défavorables pour la consommation de ces AGVs par les microorganismes de l'étape suivante, phénomène qui peut être accentué par la présence de substances toxiques ou par un changement rapide de température (Venkiteshwaran *et al.*, 2015). Ces effets conjugués peuvent inhiber ou arrêter complètement la méthanogenèse.

## 2.2. Acétogenèse

Les microorganismes impliqués dans cette étape, appelés acétogènes, sont des bactéries anaérobies strictes. Leur métabolisme peut être chimioorganotrophe, les acides organiques étant utilisés comme source d'énergie en les convertissant en acétate, ou chimiolithoautotrophe, avec l'utilisation de H<sub>2</sub>/CO<sub>2</sub>, CO/CO<sub>2</sub>, ou H<sub>2</sub>/CO comme source unique de carbone et d'énergie (Liou and Madsen, 2008). Les bactéries syntrophiques (productrices de H<sub>2</sub>) appartiennent principalement aux genres *Syntrophomonas* et *Syntrophobacter* tandis que les homoacétogènes appartiennent aux genres *Acetobacterium*, *Acetoanaerobium*, *Acetogenium*, *Butyribacterium*, *Clostridium*, *Eubacterium*, et *Pelobacter* (Borja and Rincón, 2017) (Tables 2 et 3).

Dans des conditions spécifiques, telles que l'augmentation de la température, il peut se mettre en place, au sein du processus de digestion anaérobie, la réaction d'oxydation de l'acétate par ces bactéries syntrophiques (oxydation syntrophique de l'acétate, OSA, et en Anglais *syntrophic acetate oxydation, SAO*) selon l'équation réactionnelle suivante :



## 2.3. Méthanogenèse

La production de méthane, l'étape ultime de la digestion anaérobie est un processus complexe nécessitant des enzymes et cofacteurs spécifiques (Sikora *et al.*, 2017). Les archées méthanogènes sont connues comme acteurs clés de cette étape.

### a. Archaea : troisième domaine du vivant

Auparavant, en se basant essentiellement sur les caractéristiques morphologiques, tous

les microorganismes unicellulaires ne possédant pas de noyau (comprenant les bactéries et les archées) étaient classés ensemble dans le domaine de « procaryotes ». Vers les années 1960, Zuckerkandl et Pauling ont proposé de se baser sur la comparaison de certaines molécules (ADN ou protéines) pour établir une nouvelle classification du monde du vivant (Zuckerkandl and Pauling, 1965). Suite à cette idée, Woese et Fox ont émis l'hypothèse d'un arbre du vivant à trois branches (domaines) : les Eucaryotes, les Bactéries et les Archées, en proposant une méthode comparative des ARN ribosomiques 16S ou 18S (Woese and Fox, 1977; Woese *et al.*, 1990) dont le principe sera détaillé dans la section « Approche méthodologique pour la microbiologie environnementale ». Au-delà des similarités morphologiques entre les archées et les bactéries, des études plus approfondies en biologie moléculaire ont montré plusieurs similarités génétiques et biochimiques entre les archées et les eucaryotes, telles que la présence d'introns, d'histones, ou encore la similarité de la machinerie de réplication de transcription et traduction (Baker *et al.*, 2020). De plus, l'absence de la paroi de peptidoglycane, ainsi que la liaison lipidique de type éther au lieu de type ester dans les parois cellulaires, sont aussi des caractéristiques uniques chez les archées, permettant la distinction avec les bactéries dans le monde des procaryotes. Par ailleurs, les archées montrent une forte diversité métabolique dont certains métabolismes uniques, ce qui explique leur présence ubiquitaire dans les environnements terrestres (Figure 2), notamment dans des conditions extrêmes (pH, température, pression, etc.). En effet, les archées sont observées dans des milieux hypersalins comme le lac Great Salt (États-Unis) ou la mer morte (halophiles), dans des habitats anoxiques comme les sédiments marins ou les rumens des animaux (méthanogènes), à des températures très élevées comme dans les sources hydrothermales marines profondes ou les sources chaudes terrestres comme dans le parc de Yellowstone ((hyper)thermophiles), ou inversement à basse température comme dans l'Antarctique (psychrophiles) et à des pH extrêmes (acidophiles et alkaliphiles). Elles forment donc un domaine bien distinct des bactéries et des eucaryotes. Chaque domaine possède ses propres caractéristiques morphologiques, génétiques et biochimiques (Tableau 4).



A

Great Salt Lake, Utah, USA



B

Yellowstones park



C

Ocean

Figure 2. Illustration de quelques environnements terrestres abritant des archées.

Tableau 4. Quelques caractéristiques clé des trois domaines du vivant

\* : Dans le cas général (Hinnebusch and Tilly, 1993; Hejna and Moses, 2009).

Caractéristiques	Bactéries	Archées	Eucaryotes
<b>Organisation</b>	Unicellulaire	Unicellulaire	Uni/Pluricellulaire
<b>Taille de cellules</b>	~0,4-10 µm	~0,4-10 µm	~10-100 µm
<b>Noyau</b>	Absent	Absent	Présent
<b>Organites</b>	Absent	Absent	Présent
<b>Matériel génétique</b>	Chromosome circulaire* + plasmides	Chromosome circulaire + plasmides	Plusieurs chromosomes linéaires + ADN de type procaryotes des organites
<b>Lipides membranaires</b>	Chaînes carbonées linéaires	Chaînes carbonées ramifiées	Chaînes carbonées linéaires
<b>Liaison des lipides</b>	Ester	Ether	Ester
<b>Chaînes carbonées</b>	Acides gras	Isoprènes	Acides gras
<b>ARN polymérase</b>	Simple	Complexe	Complexe
<b>Acide aminés initiateur de la traduction</b>	Formyl-méthionine	Méthionine	Méthionine
<b>Histones associées à l'ADN</b>	Non	Oui	Oui
<b>Système CRISPR</b>	Oui	Oui	Non

Suite à leur découverte, les archées ont d'abord été classées en deux phyla : *Euryarcheota* et *Crenarcheota*. Puis, à partir des années 2000, grâce aux progrès des technologies de séquençage, plusieurs nouveaux phyla ont été définis, tels que *Konarchaeota* (Barns *et al.*, 1996) et *Nanoarchaeota* (Huber *et al.*, 2003). Actuellement, les archées sont classées en 27 phylums (Baker *et al.*, 2020), regroupés en 4 « super-phylums » (Figure 3) : *Euryarchaeota*,

TACK (*Thaumarchaeota*, *Aigarchaeota*, *Crenarchaeota*, *Korarchaeota*, *Verstraetearchaeota*, *Nezhaarchaeota*, *Geothermarchaeota*, *Geoarchaeota*, *Marsarchaeota* et *Bathyarchaeota*), Asgard (*Odinarchaeota*, *Helarchaeota*, *Lokiarchaeota*, *Thorarchaeota* et *Heimdallarchaeota*) et DPANN (*Diapherotrites*, *Parvarchaeota*, *Aenigmarchaeota*, *Nanohaloarchaeota*, *Nanoarchaeota*, *Micrarchaeota*, *Pacearchaeota*, *Diapherotites*, *Woesearchaeota*, *Altiarchaeota* et *Huberarchaeota*).

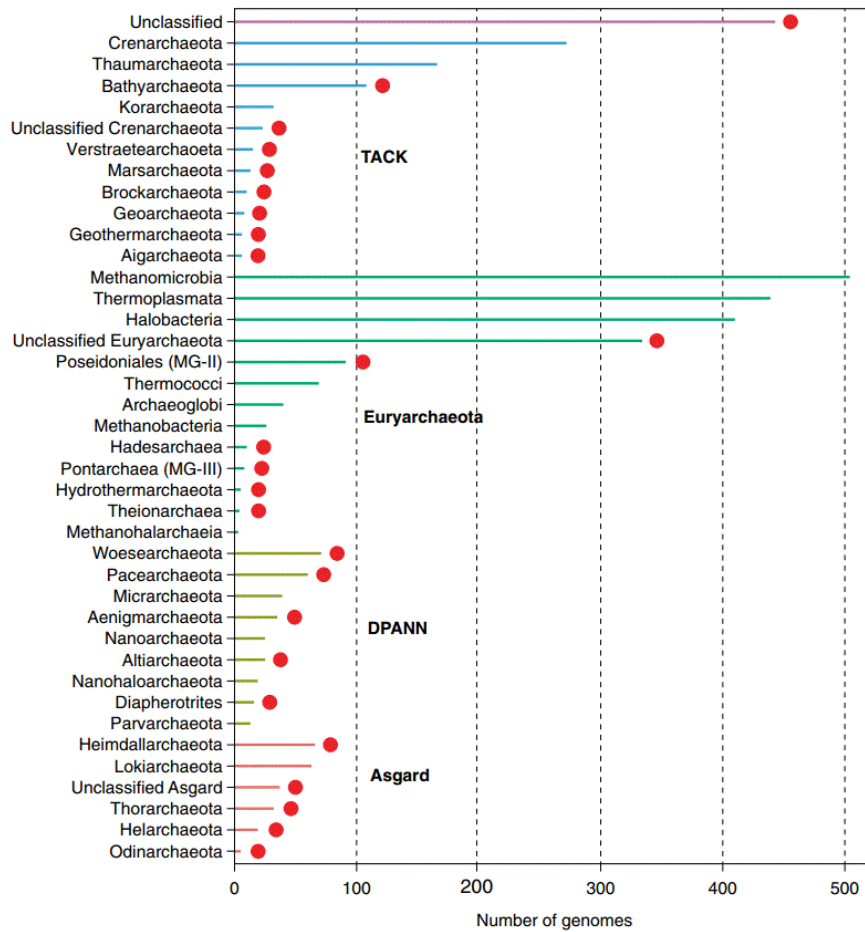


Figure 3. Nombres de génomes répertoriés dans les différents super-phylums et phylums d'archées (Baker et al., 2020). Les groupes avec point rouge ne possèdent aucune espèce cultivable.

## b. Archées méthanogènes : acteurs clés de la production de méthane

Dans le monde des archées, les premières identifiées ont été des méthanogènes, en 1933 (Stephenson and Stickland, 1933). Depuis, plusieurs études ont démontré l'importance du métabolisme du méthane dans le cycle globale du carbone (Evans *et al.*, 2019), notamment la production de méthane par les archées (méthanogenèse) qui représente plus de la moitié du méthane produit sur Terre chaque année (Reeburgh, 2007; Kirschke *et al.*, 2013). De nombreuses archées méthanogènes ont été décrites dans la littérature. Elles sont ubiquitaires dans les environnements anoxiques (Figure 4), tels que les sédiments marins profonds, les marais, le sol ou encore les tractus digestifs des animaux (insectes, vertébrés). Les méthanogènes colonisent donc une large gamme de biotopes, caractérisés par des conditions physico-chimiques (pH, température, salinité, nature et disponibilité des substrats) très variées. Bien que les archées méthanogènes soient très diverses et que nombre d'entre elles soient extrémophiles, la plupart de celles connues sont mésophiles, non-halophiles et ont un pH optimal de croissance proche de la neutralité (Zinder, 1993). De plus, certaines sont largement distribuées, tels que les archées des ordres Methanosarcinales, Methanomassiliicoccales, Methanomicrobiales et Methanobacteriales, tandis que d'autres semblent plus spécialistes, nécessitant un habitat particulier, par exemple les archées halophiles de l'ordre Methanonatronarchaeales (Figure 4).

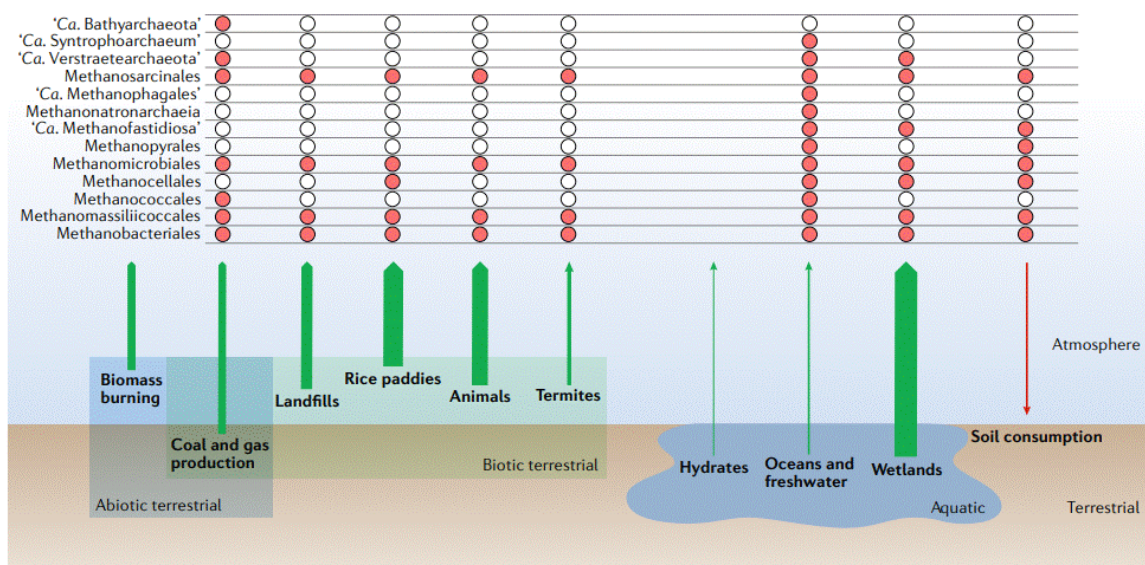


Figure 4. Flux globaux du méthane et distribution des méthanogènes (Evans *et al.*, 2019).

Les archées méthanogènes étaient historiquement limitées au phylum Euryarchaeota et regroupées en 8 ordres (Evans *et al.*, 2019; L. Bräuer *et al.*, 2020) : Methanopyrales, Methanococcales, Methanomicrobiales, Methanocellales, Methanonatronarchaeales, Methanomassiliicoccales, Methanosarcinales, et Methanobacteriales. L'analyse des données métagénomiques ont abouti à l'identification de nombreux groupes de méthanogènes putatifs, appartenant à d'autres phylums qu'Euryarchaeota, tels que « Candidatus Verstraetearchaeota » (Berghuis *et al.*, 2019) ou « Candidatus Bathyarchaeota » (Evans *et al.*, 2015). Récemment, la nouvelle phylogénie des archées basée sur la comparaison du complexe de méthyl-coenzyme M réductase (MCR) ou son homologue, une protéine marqueur du métabolisme de méthane (méthanogenèse et méthanotrophie) (Borrel *et al.*, 2019; Evans *et al.*, 2019) a montré une distribution plus complexe des taxons liés au métabolisme du méthane (Figure 5-6), ce qui a relancé le débat sur l'origine mono- ou polyphylétique de la méthanogenèse et l'évolution des archées méthanogènes. (Berghuis *et al.*, 2019).

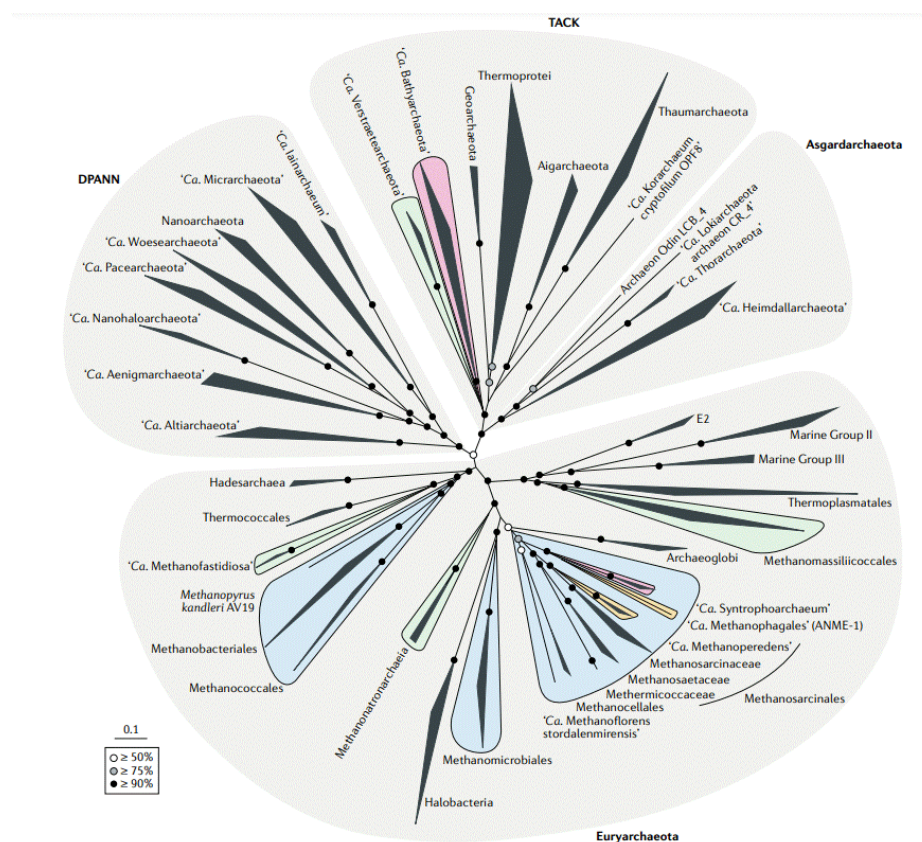


Figure 5. Phylogénie des archées basée sur les gènes *mcr* et *mcr*-like (Evans *et al.*, 2019)

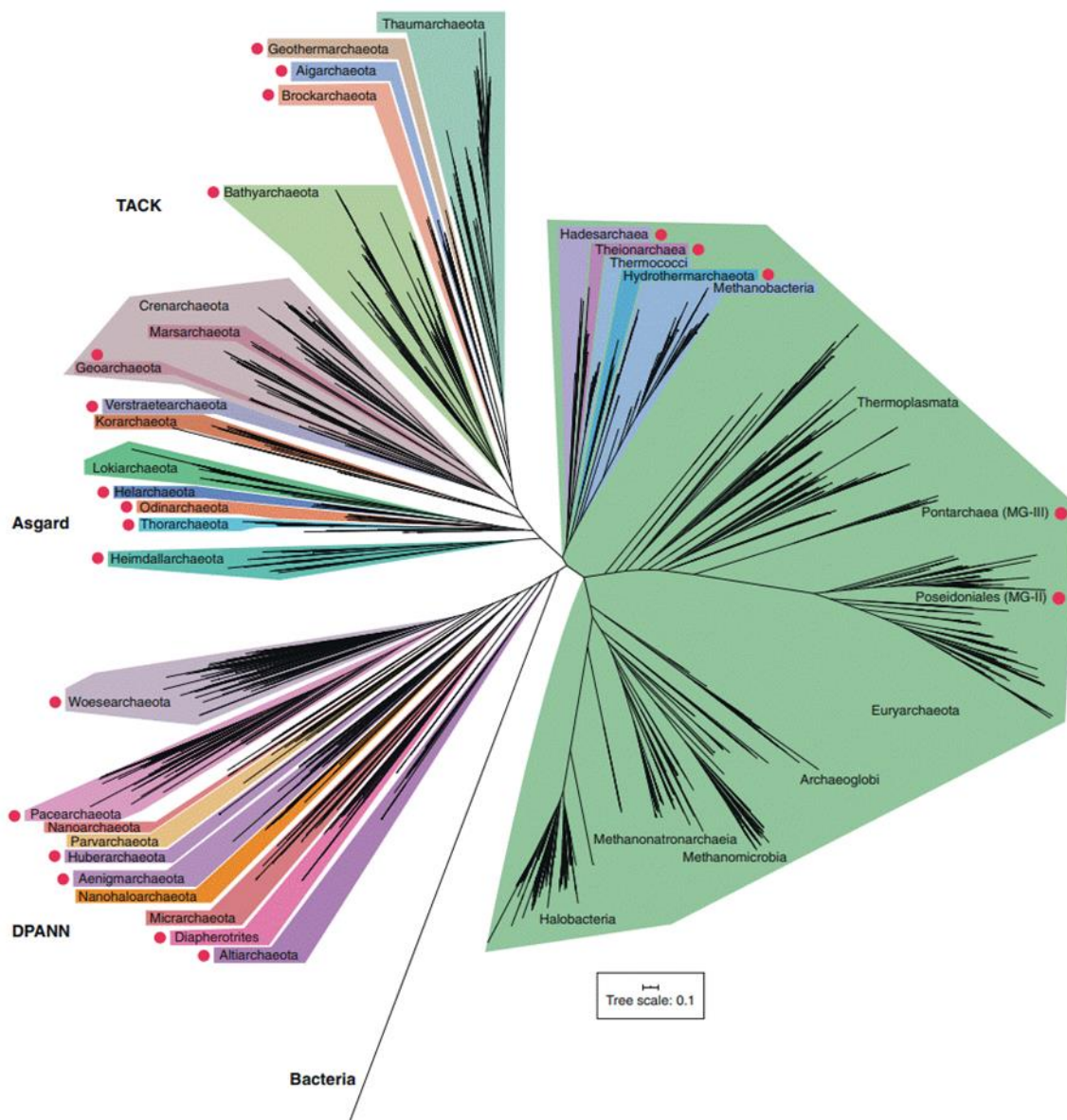


Figure 6. Phylogénie des archées basée sur 36 gènes marqueurs (Baker et al., 2020). Les pylums avec point rouge ne possèdent aucune espèce cultivable.

Les archées méthanogènes sont également présentes dans des habitats non-naturels, tels que les digesteurs anaérobies. La composition de la communauté des méthanogènes dans ces milieux peut varier en fonction de la nature du substrat (Kim *et al.*, 2015), de la température, de la composition chimique du milieu ou d'autres conditions opératoires (Amani *et al.*, 2010; Yan *et al.*, 2015; He *et al.*, 2018; Eryildiz *et al.*, 2020; J. Zhang *et al.*, 2020; S. Zhang *et al.*, 2021).

### c. Différentes voies de méthanogénèse

Il existe trois voies principales de méthanogénèse bien connues à l'heure actuelle, différant notamment par la nature des substrats : hydrogénotrophe ( $H_2$  et  $CO_2$ ), acétoclaste (acétate) et méthylootrophe (composés méthylés :  $X-CH_3$ ) (Figure 7).

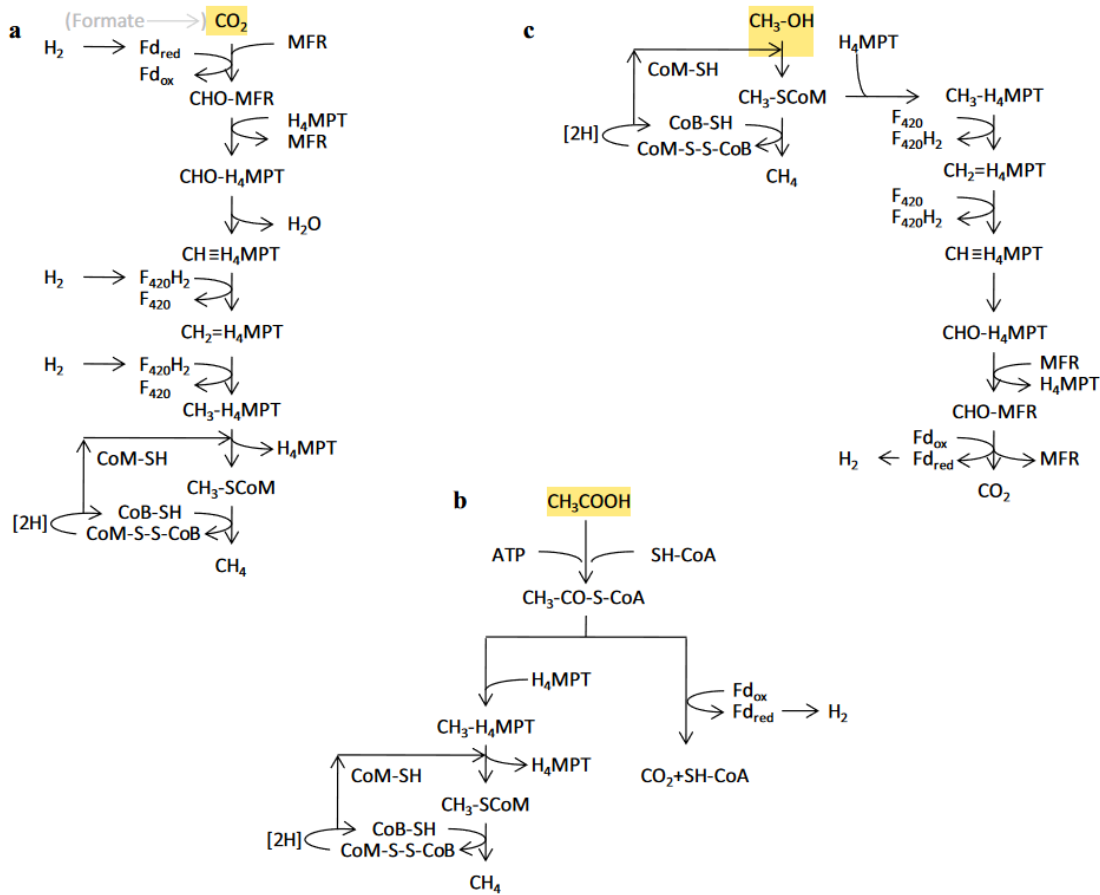


Figure 7. Trois voies distinctes de la méthanogénèse (Enzmann et al., 2018) (a : hydrogénotrophe, b : acétoclaste, c : méthylootrophe)

### La méthanogénèse hydrogénotrophe

La plupart des archées méthanogènes sont capables de réduire le  $CO_2$  en  $CH_4$  en utilisant  $H_2$  comme donneur d'électrons (Figure 7a). Il faut également noter que plusieurs d'entre elles peuvent utiliser le formate au lieu de  $H_2$  et  $CO_2$  pour former du méthane (Enzmann et al., 2018). La méthanogénèse à partir de  $CO_2$  et  $H_2$  nécessite 8 électrons pour réduire une molécule  $CO_2$  en  $CH_4$  donc 4 molécules  $H_2$  (Figure 1) et implique différentes molécules intermédiaires. Le  $CO_2$  est d'abord réduit en formyle-méthanofuran (CHO-MFR) à l'aide de



la ferrédoxine réduite ( $Fd_{red}$ ) comme donneur d'électron. Ensuite, le groupe formyle est transféré à la tétrahydrométhanoptérine ( $H_4MTP$ ), déshydraté, puis réduit en cascade en méthyle- $H_4MTP$  avec le  $F_{420}$  ( $F_{420}H_2$ ) réduit comme donneur d'électrons. Le groupe méthyle est par la suite transféré à la coenzyme M ( $HS-CoM$ ) et est finalement réduit en méthane avec la coenzyme B ( $HS-CoB$ ) comme donneur d'électrons. Le complexe enzymatique formé, l'hétérodisulfure ( $CoM-S-S-CoB$ ) est réduit par du  $H_2$  pour régénérer les coenzymes.

**La méthanogenèse acétoclaste** : Les membres des genres *Methanosarcina* peuvent produire le méthane par réduction de l'acétate ( $CH_3COO^-$ ), et le genre *Methanosaeta* est exclusivement acétoclaste. Dans ce processus, l'acétate est converti en acétyl-CoA avec de l'ATP et la coenzyme A. Le groupe méthyle de l'acétate ( $-CH_3$ ) est ensuite réduit en méthane grâce à un mécanisme similaire à la réduction du  $CO_2$  décrite ci-dessus. Le groupe carbonyle ( $-CO^-$ ) est oxydé en  $CO_2$  et fournit ainsi les électrons pour la réduction du groupe méthyle (Figure 7b).

**La méthanogenèse méthylootrophe** : Certaines archées méthanogènes (principalement les membres de l'ordre *Methanosarcinales*) utilisent des composés méthylés pour produire le méthane, tels que le méthanol, les méthylamines, le méthaneithiol, ou le diméthylsulfide. Le groupe méthyle de ces composés est converti, avec la coenzyme M ( $CoM$ ), en méthyle- $CoM$ , sous l'action de différentes méthyltransférases. Une molécule de méthyle- $CoM$  est oxydée en  $CO_2$  via la voie hydrogénotrophe inversée pour générer six électrons, permettant de réduire trois autres molécules de méthyle- $CoM$  en  $CH_4$  (Figure 7c). Récemment, d'autres voies possibles de méthanogenèse (de type méthylootrophe) ont été décrites dans la littérature, notamment dans les environnements hypersalins ou riches en composés sulfurés (Lyu *et al.*, 2018; Conrad, 2020). En outre, les voies de méthanogenèse des nouveaux candidats venus des phyla *Verstraetearchaeota* et *Bathyarchaeota* restent à étudier de manière plus approfondie. Il est possible que des alcanes ou des composés à base de méthoxydes soient également des substrats de méthanogenèse (Mayumi *et al.*, 2016).

## 2.4. Métabolisme du formate : intérêt pour la bioraffinerie

Durant le processus de digestion anaérobie, différentes molécules intermédiaires sont produites et métabolisées pour mener à la formation du produit final : le méthane. Parmi ces molécules, le formate présente un intérêt particulier dans le cadre du développement de la bioéconomie (Yishai *et al.*, 2016). En effet, cette molécule peut être produite non seulement à partir de biomasse mais aussi à partir de CO<sub>2</sub> et d'énergie renouvelable, par de nombreuses voies physiques, chimiques et biologiques (Figures 8-9). Le formate constitue alors une molécule plate-forme qui peut être consommée par des microorganismes comme source de carbone et/ou d'énergie pour produire d'autres molécules plus complexes à haute valeur ajoutée (Yishai *et al.*, 2016) (Figure 8). Le formate est impliqué dans différents voies métaboliques, telles que le cycle de Calvin, le cycle de la sérine, l'acétogénèse, ou la méthanogénèse.

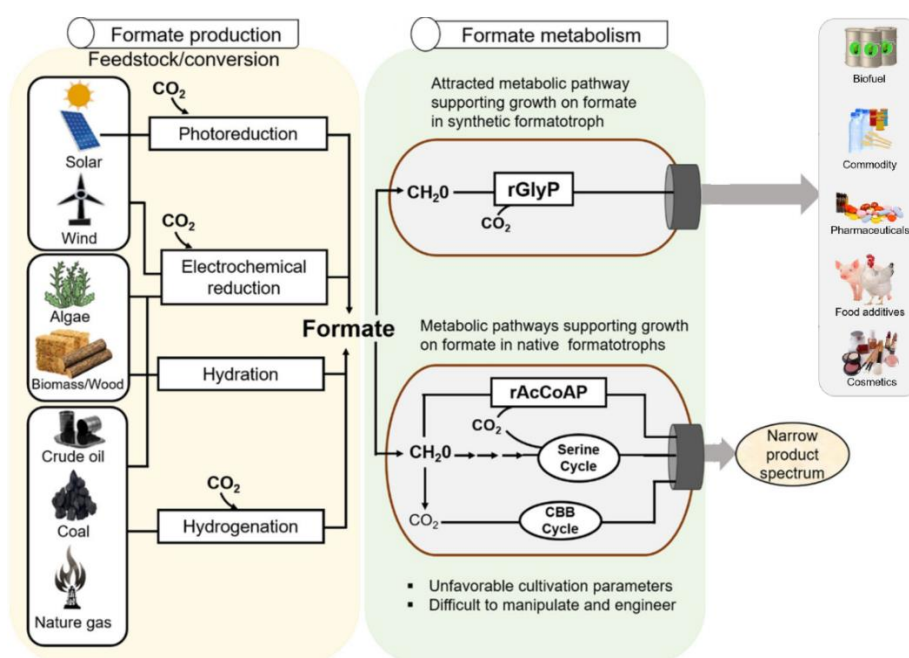


Figure 8. La bioéconomie du formate, adaptée de (Tuyishime and Sinumvayo, 2020)

Les formatotrophes désignent les microorganismes qui sont capables d'assimiler le formate comme source de carbone et/ou qui l'utilisent comme source d'électrons. Outre des archées méthanogènes, différentes bactéries sont également capables de consommer le formate au sein du processus de digestion anaérobie, notamment les bactéries acétogènes et sulfato-réductrices (Brysch *et al.*, 1987; Cotton *et al.*, 2020; Lemaire *et al.*, 2020) (Figure 9). L'étude du métabolisme anaérobie du formate en culture mixte permet donc de mieux comprendre son importance dans le processus global de la digestion anaérobie, en identifiant les voies métaboliques en jeu, les microorganismes impliqués et leurs interactions. Cela pourrait aussi permettre d'identifier des pistes biotechnologiques pour le développement du formate en tant que molécule plate-forme pour la bioéconomie (Yishai *et al.*, 2016).

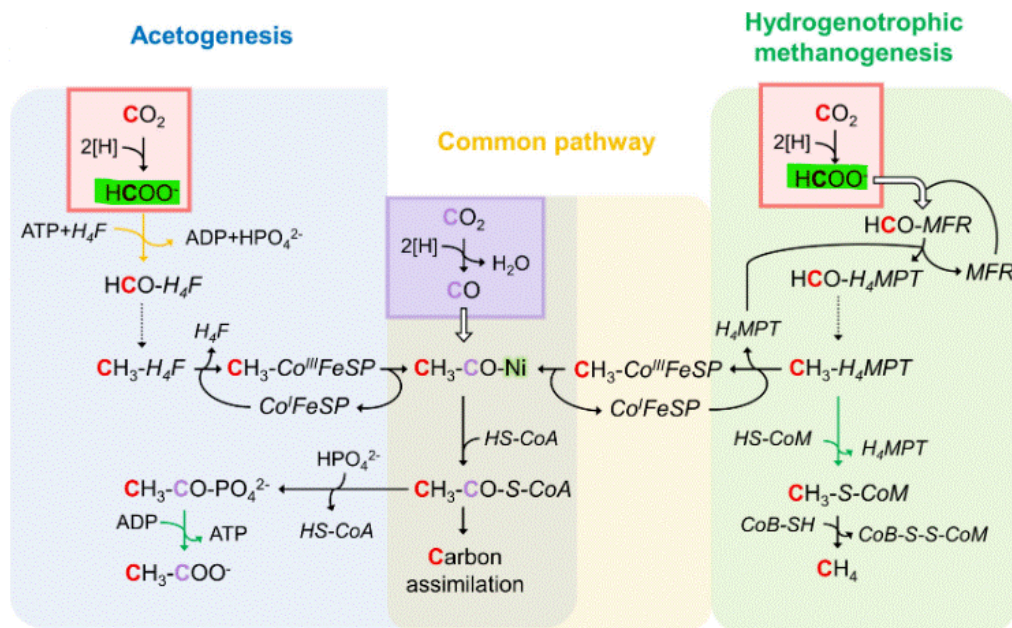


Figure 9. Différentes voies métaboliques d'utilisation du formate (en vert) au cours de la DA (Lemaire *et al.*, 2020)

### 3. Virus au sein du processus de digestion anaérobie

#### 3.1. Virus : découverte et définition

Le terme « virus » vient du latin « *vīrus* » (venin, poison), et désigne historiquement tous les agents infectieux sans indication de leur nature. Les virus sont d'abord connus par leur manifestation pathologique chez l'Homme. Les maladies infectieuses d'origine virale ont été observées et rapportées dès l'Antiquité dans les écrits gréco-romains, égyptiens ou chinois, tels que les notes d'Hippocrate (460-377 av. J-C), la mise en évidence de la poliomyélite dans l'Égypte antique ou la description de la variole par les chinois il y a près de 4 500 ans. Cependant, la découverte, l'isolation et l'observation des virus ne débute qu'à la fin du XIX<sup>ème</sup> siècle, grâce aux progrès technologiques.

Suite à l'invention du vaccin contre la rage par L. Pasteur et E. Roux en 1885, et à la mise en évidence du premier virus d'eucaryote - le virus de la mosaïque du tabac (TMV : Tobacco mosaic virus) par D. Ivanovsky et M. Beijerinck en 1899, puis la découverte des virus de bactéries (bactériophages) par F. Twort et F. d'Hérelle en 1915, la recherche en virologie a continué à se développer et a mené à la découverte d'un monde divers et complexe. Le concept, la définition de « virus » ont également évolués. Bien que la question « Virus : vivant ou non-vivant ? » soit toujours en discussion avec différents concepts proposés (Raoult and Forterre, 2008; Forterre, 2011), le monde viral est totalement différent du monde cellulaire. Un virus est une entité biologique qui peut être définie par les caractères suivantes :

- un virus possède souvent une petite taille (< 0,20 µm), à l'exception des virus géants (Ex : *Mimivirus*, *Pandoravirus*)
- un virus est constitué d'un seul type d'acide nucléique comme support de l'information génétique, soit de l'ADN soit de l'ARN.
- les virus sont des parasites obligatoires. Ils infectent les cellules dans l'immense majorité des cas, mais ils peuvent aussi infecter un autre virus (d'où la notion « virophage ») (Smallridge, 2008).

### 3.2. Diversité virale

Au cours de ce siècle, l'ère de la métagénomique a mis en lumière un nombre énorme de virus découverts dans différents environnements (Jiang *et al.*, 2015; Suttle, 2016; Graham *et al.*, 2019). Malgré le débat scientifique sur le nombre estimé de  $10^{31}$  particules virales sur Terre (Mushegian, 2020), il est devenu clair que les virus sont ubiquitaires et jouent un rôle important dans la régulation de la biosphère. En effet, les virus sont les entités biologiques les plus abondantes dans les océans (Suttle, 2005, 2007) et aussi certainement sur la planète (Suttle, 2016). Considérant que pratiquement chaque forme de vie est parasitée par au moins un virus, ces derniers présentent potentiellement une très grande diversité comparable à l'estimation de plusieurs millions d'espèces d'organismes cellulaires sur la Terre (Mora *et al.*, 2011). Bien que cette immense diversité ne soit encore que partiellement explorée, et constitue encore une matière noire (*dark matter*) à découvrir, les virus connus à l'heure actuelle présentent déjà une grande variabilité morphologique, génomique et génétique.

#### a. Diversité morphologique

Il existe une grande diversité de taille et de structure au sein des virus. En effet, la taille des particules virales (virions) varie entre  $\approx 27$  nm et  $1,5 \mu\text{m}$  de diamètre. Un virion est composé du génome viral, empaqueté dans la capsid - une structure protéique qui le protège. Chez certains virus, une enveloppe est également présente. Sa nature peut varier, elle peut notamment être lipidique et être agrémentée de glycoprotéines (Murin *et al.*, 2019) (Figure 10). La présence ou l'absence de cette enveloppe joue un rôle dans la stabilité des particules dans l'environnement, dans l'infectiosité des virus, ainsi que dans leurs stratégies d'infection (voir Section « interaction hôte-virus »).

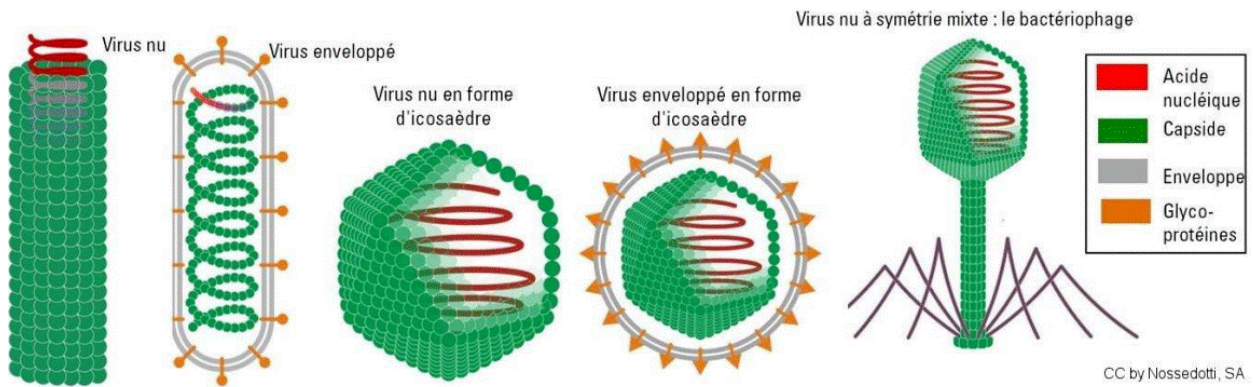


Figure 10. Structures de virus présentant différents types de symétrie (modifiée à partir de <https://commons.wikimedia.org/wiki/User:Nossedotti> )

La capside est constituée de protéines virales polymérisées en formes géométriques souvent régulières. Les capsides présentent quasiment toujours des symétries, qu'elles soient de type hélicoïdal, icosaoédrique ou mixte (virus tête-queue par exemple) (Figure 10). En outre, la morphologie des capsides virales présente également une grande diversité, notamment chez les virus d'archées avec des morphotypes uniques en forme de gouttes d'eau (*Guttavirus*), de bobine (*Alphalipothrixvirus*) de bouteilles (*Ampullavirus*) ou de citrons (*Fusellovirus*, *Bicaudovirus*) (Figure 11).

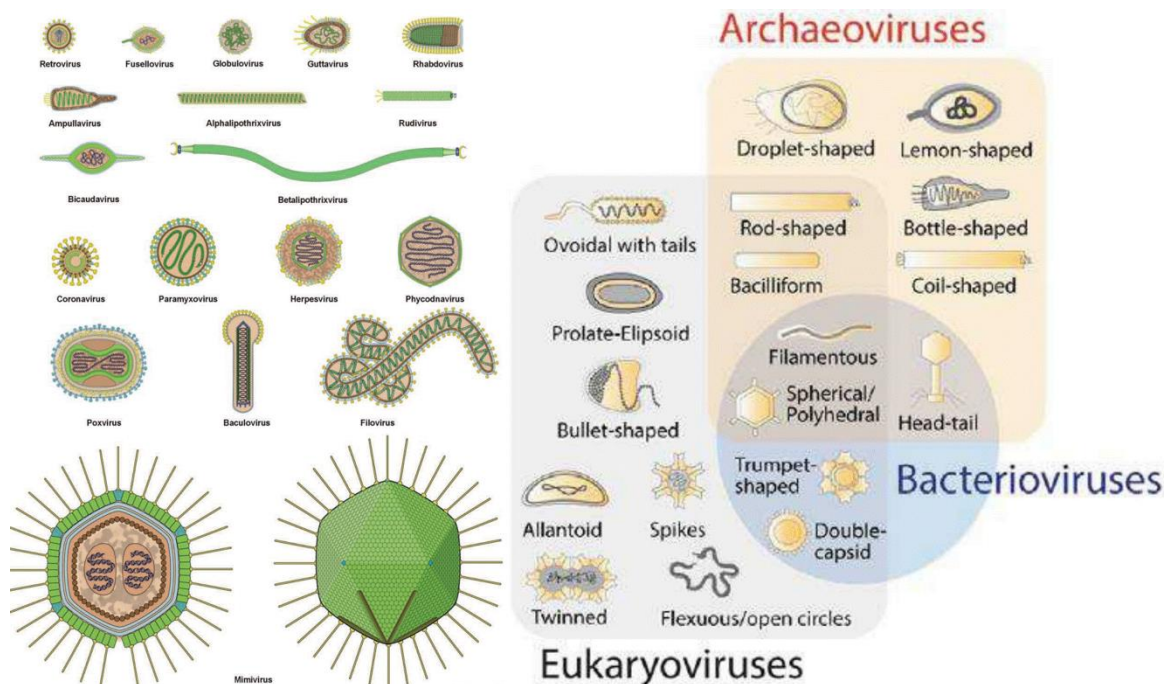


Figure 11. Illustration de la diversité morphologique des virus

## b. Diversité génomique et génétique

Au-delà de la diversité morphologique, les génomes viraux présentent également une grande diversité de structure génétique. En effet, la taille des génomes viraux varie de moins de 2000 bases jusqu'à plus de 2,5 Mb chez les virus géants (Koonin and Yutin, 2019) et ces génomes peuvent être circulaires ou linéaires, monocaténaires ou bicaténaires. Il existe différents types de support de l'information génétique chez les virus : ADN double brin (ADNdb), ADN simple brin (ADNsb), ARN simple brin positif (ARNsb+), ARN simple brin négatif (ARNsb-) et ARN double brin (ARNdb). De plus, les virus possèdent différentes stratégies de réplication et d'expression, en lien avec la nature de leur matériel génétique. En plus de la stratégie qui est utilisée chez toutes les formes cellulaires et est basée sur la réplication de l'ADNdb, la transcription d'ARN messagers suivie par leur traduction en protéines, des stratégies alternatives telles que la transcription inverse ou la réplication de l'ARN sont aussi couramment employées chez les virus (Figure 12).

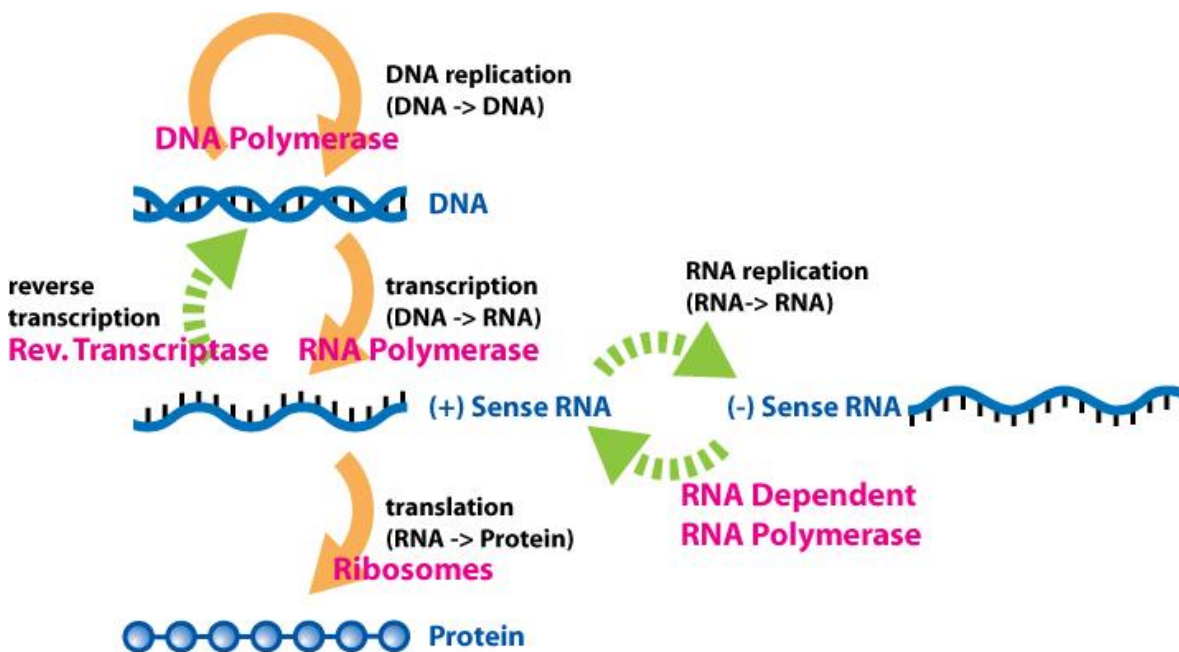


Figure 12. Dogme central de la biologie moléculaire. Les flèches vertes reflètent les flux inhabituels, dont certains sont liés aux virus ([https://commons.wikimedia.org/wiki/File:Extended\\_Central\\_Dogma\\_with\\_Enzymes.jpg](https://commons.wikimedia.org/wiki/File:Extended_Central_Dogma_with_Enzymes.jpg))

Les génomes viraux contiennent des gènes codant des éléments viraux fondamentaux,

tels que des protéines majeures de capsid ou des protéines impliquées dans la réplication (primases, hélicases, polymérases, etc.). Cependant, des protéines de même fonction peuvent être codées par un ou plusieurs types ou lignées de gènes, comme dans le cas des capsides icosaédriques pour lesquelles existent au moins quatre lignées différentes, et sans lien évolutif (Abrescia *et al.*, 2012). Il n'existe pas de gène universel chez les virus, qui serait par exemple l'équivalent des gènes codant pour l'ARN ribosomique dans le monde cellulaire, ce qui rend contraignante la classification des virus.

### 3.3. Taxonomie et classification des virus

Historiquement, les virus étaient nommés et classés en se basant sur la maladie qu'ils causaient, tels que : le virus de la rage, celui de la variole, etc. Au XX<sup>ème</sup> siècle, avec les progrès en biologie moléculaire, les premières classifications des virus basées sur la nature de l'acide nucléique génomique et sur les caractéristiques de la capsid (présence/absence de l'enveloppe, type de symétrie, diamètre) sont apparues. En 1971, Baltimore a proposé une classification prenant en compte le mode de réplication et la nature du génome viral (Baltimore, 1971). Les six « classes de Baltimore » de virus, avec une 7<sup>ème</sup> classe ajoutée par la suite (Figure 13), sont devenues une référence pour développement de la virologie jusqu'à nos jours (Koonin *et al.*, 2021).

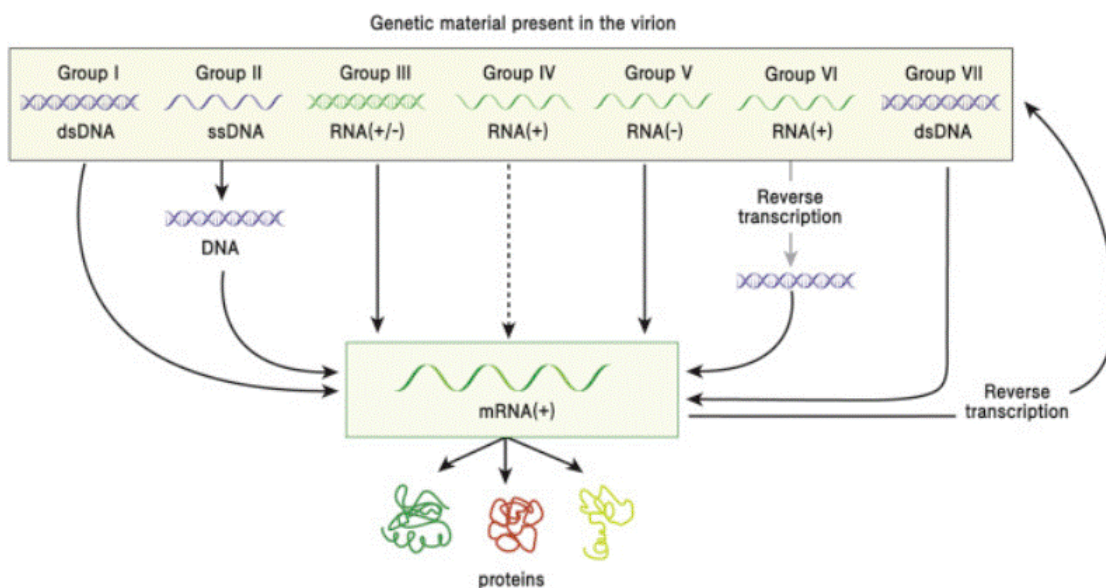


Figure 13. Classification Baltimore récente des virus (Ryu, 2017)



De nos jours, la comparaison de divers caractères décrivant un virus est utilisée pour définir et distinguer les taxons viraux. Ces caractères incluent la composition moléculaire du génome, la présence d'une capsid virale et ses propriétés structurales, l'expression des gènes utilisés pour produire des protéines virales, la nature des hôtes, la pathogénicité, et les similarités de séquence. L'ICTV (International Committee on Taxonomy of Viruses), fondé en 1966, définit la taxonomie et la classification de l'ensemble des virus (infectant les bactéries, les archées, ou les eucaryotes). Actuellement, 10 434 virus distincts sont recensés (selon la version disponible en ligne depuis juillet 2021), et classés en 6 règnes (realms), 1 classe, 19 familles et 2 genres disjoints. La très grande majorité des virus décrits sont à ADN double brin, présentent une morphologie binaire de type tête-queue et appartiennent à la classe *Caudoviricetes*. Malgré le faible nombre de virus d'archées décrits (environ 136) et le fait que ceux-ci sont tous à ADN, ces virus sont classés en plus de 20 familles différentes, en raison de leur grande diversité de morphologie et de contenu génétique : le plupart de leurs gènes n'ont pas de correspondant dans les bases de données publiques (Krupovic *et al.*, 2018).

### 3.4. Interactions hôte-virus

#### a. Cycles viraux

Les virus virulents sont caractérisés par le fait qu'ils ne peuvent effectuer qu'un seul cycle suite à leur infection : le **cycle lytique**, qui provoque la lyse de la cellule hôte (Figure 14). Durant ce cycle, le virus injecte son génome dans le cytoplasme de l'hôte. En détournant la machinerie cellulaire, le génome viral est répliqué et les différents composants structuraux du virus sont produits. Les virions sont ensuite assemblés et la lyse de la cellule hôte est déclenchée, aboutissant à la libération des virions dans l'environnement. Ces derniers peuvent ensuite rencontrer un nouvel hôte, et le cycle lytique peut être répété une nouvelle fois.

Les virus tempérés sont capables non seulement d'utiliser le cycle lytique mais également le **cycle lysogénique**. Au cours de ce dernier, le génome viral est intégré, le plus souvent de manière stable, au sein du chromosome de l'hôte, à des sites spécifiques, formant alors un « provirus ». Plus rarement, chez certains virus, le génome reste sous forme d'épisome

dans la cellule. Dans les deux cas, l'ADN viral est répliqué lorsque la cellule hôte réplique son propre ADN, au cours de la division cellulaire. Le provirus est donc transféré de la cellule mère aux cellules filles et il peut rester dans le chromosome de l'hôte pendant plusieurs générations. La sortie du cycle lysogénique peut être déclenchée par induction, souvent lorsque la cellule subit un stress (densité cellulaire élevée, carence nutritive, exposition aux antibiotiques, etc.) et conduit à l'activation du cycle lytique. Les virions relargués peuvent ensuite rencontrer un nouvel hôte. Lorsque l'on considère une population de cellules lysogènes, il peut exister, même en l'absence de stress connu, une activation à bas bruit du cycle lytique, il s'agit de l'induction spontanée des prophages (Owen *et al.*, 2017).

Il existe d'autres types d'interaction hôte-virus. **La pseudo-lysogénie** est un phénomène observé chez certains virus, virulents et tempérés (Łoś and Węgrzyn, 2012). Celle-ci peut être définie comme l'état figé d'un virus dans la cellule hôte. Le génome viral n'est pas intégré dans le génome de leur hôte mais reste stable dans la cellule, permettant ainsi le redémarrage du développement du virus. Ce phénomène a lieu lorsque les conditions sont défavorables à la croissance de la cellule hôte, par exemple lors d'un manque de substrat, et s'achève avec une initiation du cycle lytique ou lysogénique lorsque les conditions s'améliorent.

**L'infection chronique** correspond à un état où la production et l'export des virions n'induisent pas la lyse de l'hôte. Ce type d'infection chronique est souvent observé chez les virus filamenteux des bactéries (tels que M13) et les virus enveloppés (Klumpp and Crépin, 2014). La cellule hôte produit les virions de façon continue et les secrète soit *via* des pores protéiques, dans le cas des virus filamenteux, soit par bourgeonnement cellulaire. Cet état chronique est maintenu tant que la cellule hôte est en croissance. En générale, lorsque l'hôte atteint sa phase stationnaire, la lyse cellulaire se déclenche.

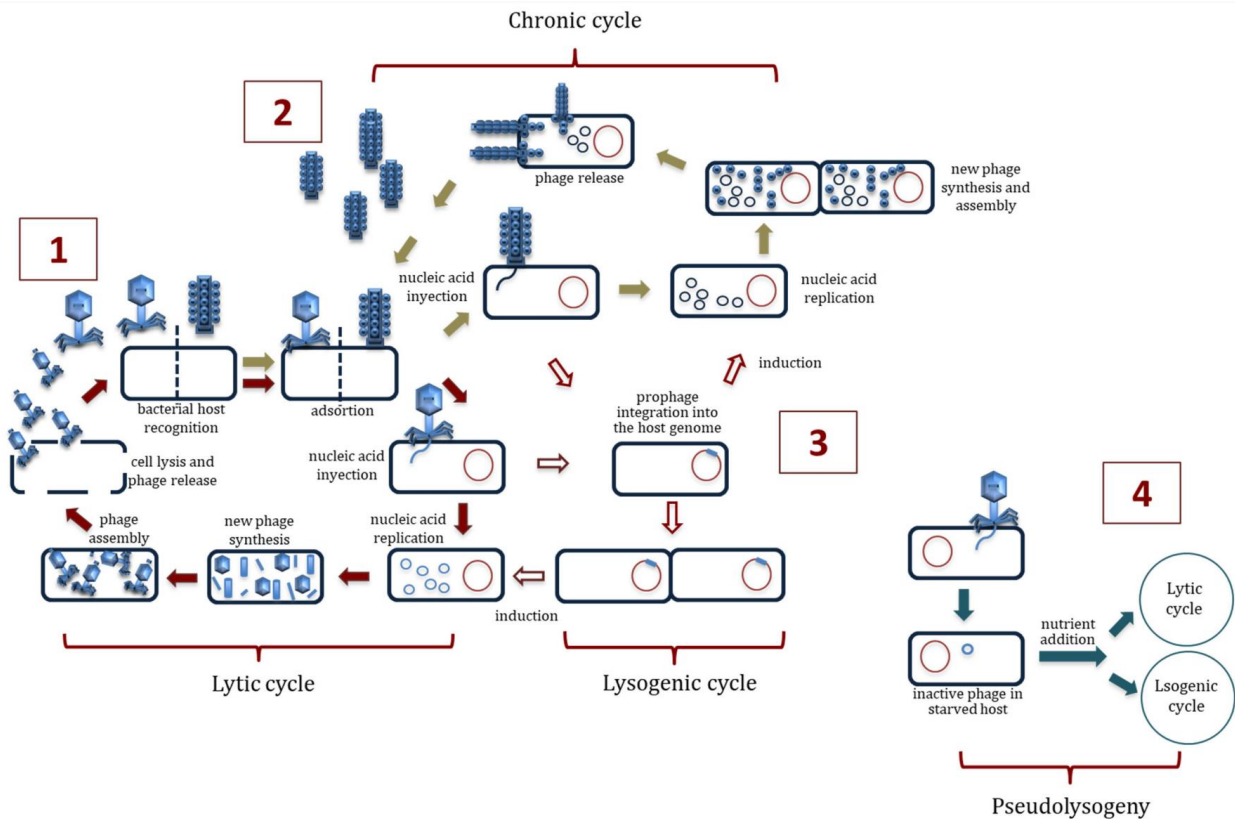


Figure 14. Différents types de cycles viraux (Sieiro et al., 2020)

## b. Mécanismes de défense de l'hôte

La mortalité cellulaire causée par les virus crée une forte pression de sélection chez l'hôte pour acquérir une résistance à ces virus. Chez les eucaryotes pluricellulaires, il existe différents mécanismes/ cellules spécifiques participant au système immunitaire. Les procaryotes ont développé également une diversité de stratégies pour résister aux infections virales (Figure 15):

- Modification des récepteurs viraux soit en les cachant avec une protéine de masque soit en les modifiant.
- Reconnaissance et clivage des ADN exogènes
- Activation de la mort cellulaire programmée ou de l'entrée en dormance des cellules infectées

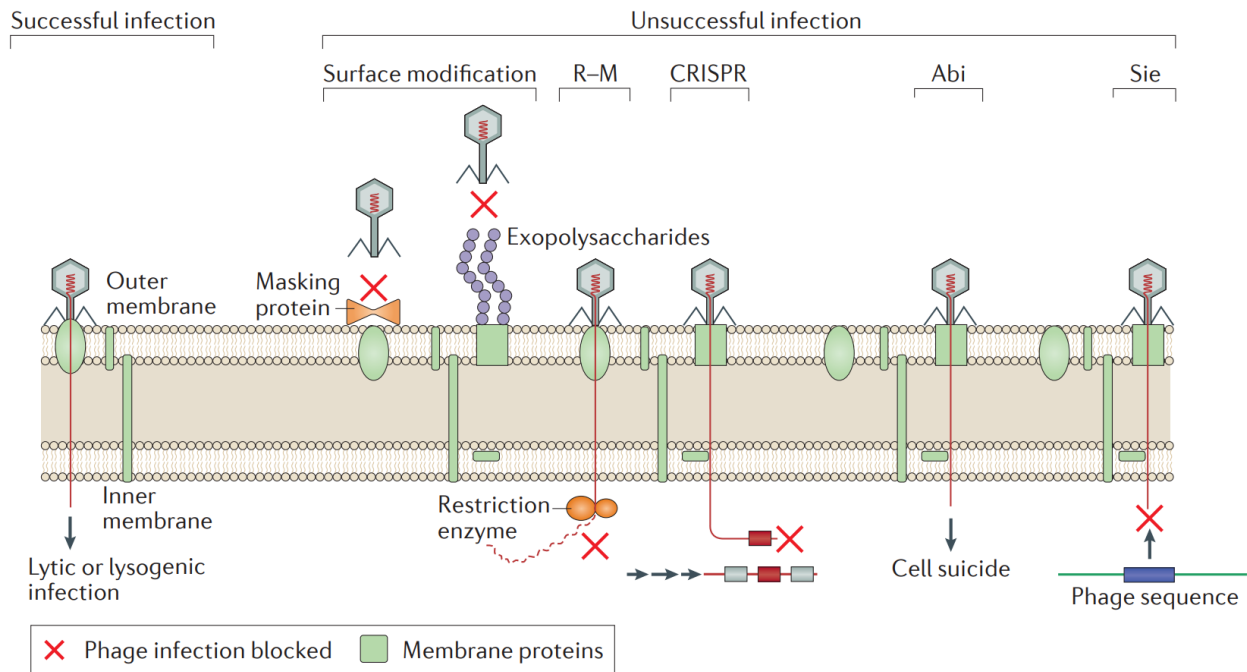


Figure 15. Stratégies de défense des hôtes procaryotes contre les virus (Mirzaei and Maurice, 2017)

**Le système Sie (SuperInfection Exclusion)** est un système codé par les virus et permettant, suite à leur infection, de protéger l'hôte contre la surinfection par d'autres virus (Folimonova, 2012). Ce système a un mode d'action permettant soit d'empêcher l'injection du génome du deuxième virus dans la cellule hôte, soit d'inhiber la réplication de ce deuxième virus.

➤ Immunité innée

**Les systèmes de restriction-modification (R-M)** sont les premiers mécanismes de défense de l'hôte découverts et décrits (Chater and Wilde, 1980). Ces systèmes assurent la méthylation de l'ADN endogène à l'aide de méthyltransférases (MT) et permettent le clivage des ADN non méthylés, considérés comme du « non-soi », sous l'action des endonucléases de restriction.

Le système de défense centré sur les **protéines argonautes (Ago)** repose également sur la reconnaissance de soi et non-soi à l'aide des molécules guide (ADN ou ARN). Après avoir reconnu la molécule guide grâce à un groupe phosphate placé à l'extrémité 5' du guide, les protéines Ago se lient à l'acide nucléique cible pour effectuer leur activité endonucléase (Höck and Meister, 2008). Dans le cas de *Methanocaldococcus jannaschi*, les

protéines MjAgo sont capables de cliver les longs ADNdb sans avoir besoin de guide. Le produit de ce clivage sert alors de guide pour une seconde étape de clivage qui devient plus efficace (Zander *et al.*, 2017). En outre, les protéines Ago sont également impliquées dans la régulation de la traduction, en inhibant la transcription des ARNm sans clivage (Figure 16) (Höck and Meister, 2008).

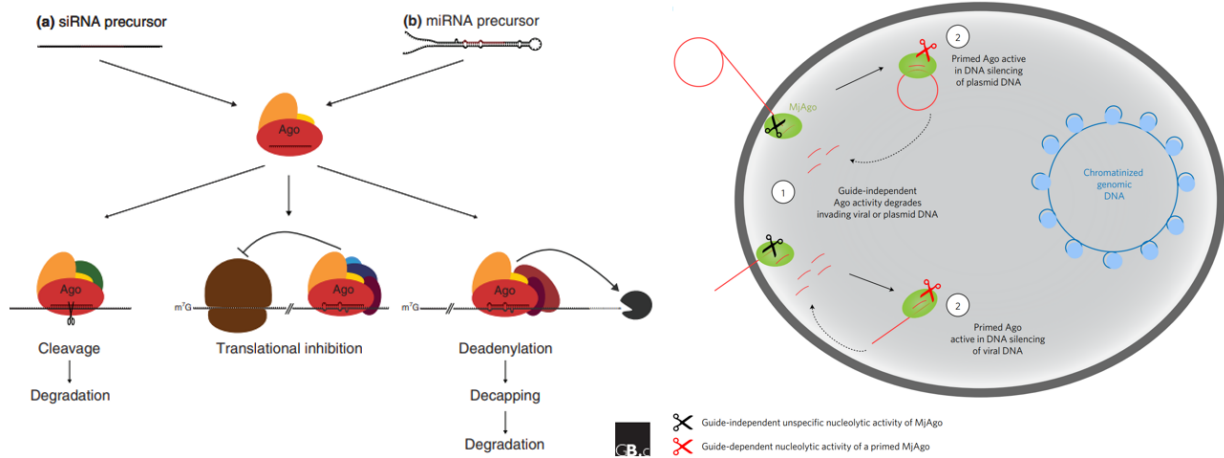


Figure 16. Rôles des protéines argonautes (Ago) (Höck and Meister, 2008; Zander *et al.*, 2017)

Un autre mécanisme connu pour empêcher la propagation des virus est présenté par les **systèmes Abi (Abortive Infection)**. Il s'agit de systèmes d'apoptose qui entraînent la mort prématurée de la cellule infectée, permettant la survie de la communauté microbienne (Lopatina *et al.*, 2020).

➤ Immunité adaptative (Crispr-Cas)

Le **système CRISPR-Cas** (« *Clustered Regularly Interspaced Short Palindromic Repeats-associated with cas genes* ») est l'un des systèmes de résistance les mieux décrits, dont le mécanisme est basé sur la reconnaissance d'acide nucléique exogène. Ce système microbien utilise son locus CRISPR pour conserver les informations génétiques exogènes hypervariables (origine virale ou plasmidique), sous la forme de séquences « *spacers* » (espaceurs), séparées par des séquences courtes conservées et répétées, partiellement palindromiques appelées « *repeats* » (répétitions) (Figure 17). Le locus CRISPR joue donc un rôle de « dictionnaire » permettant la reconnaissance des intrusions. Les gènes *cas* codent pour différentes familles de protéines fonctionnelles dont le rôle est d'éliminer

l'ADN étranger (Hille *et al.*, 2018) en dirigeant les trois phases d'immunité : acquisition, biogenèse de l'ARN CRISPR (ARNcr) et interférence.

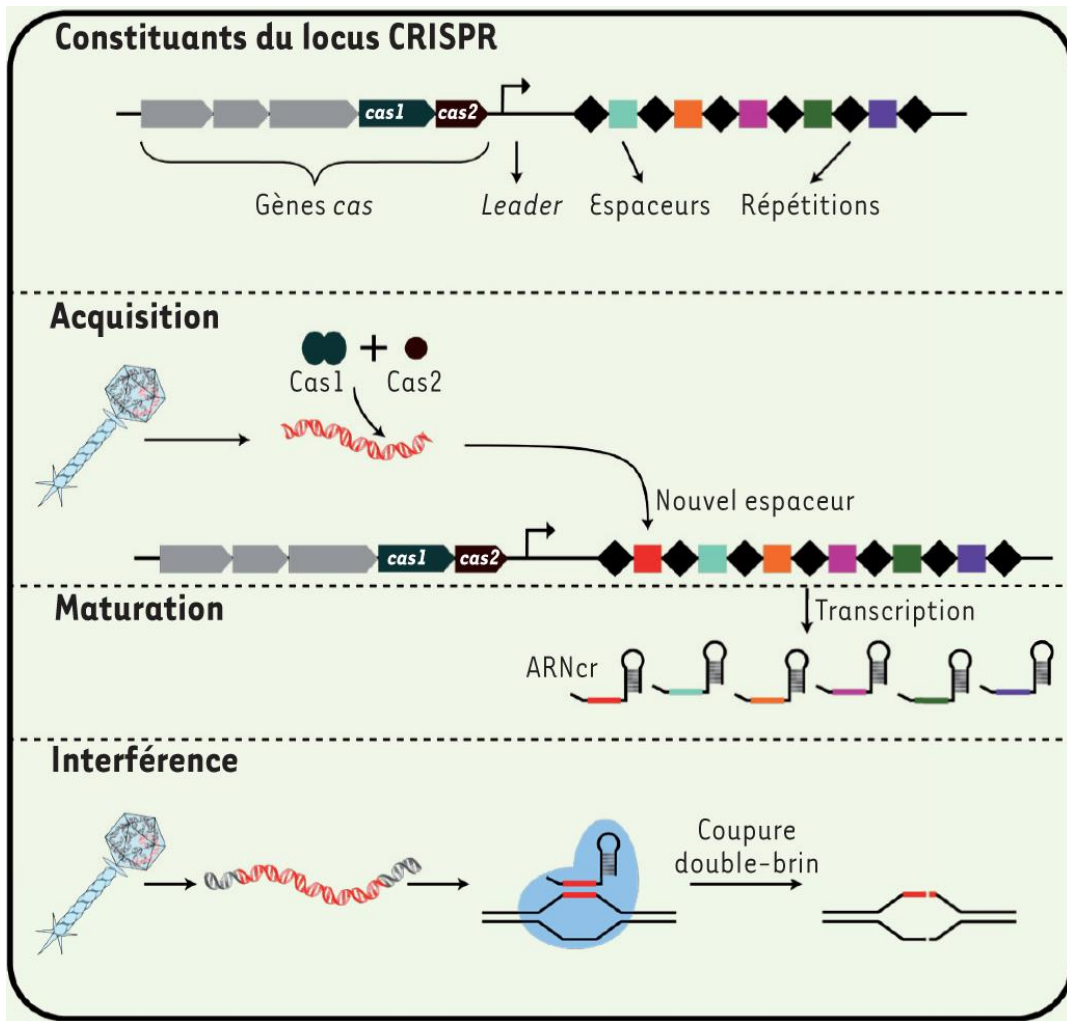


Figure 17. Mode d'action du système CRISPR-Cas (Croteau *et al.*, 2018)

Durant la phase d'acquisition, le complexe *Cas1-Cas2* capture un fragment de l'ADN étranger, appelé « *protospacer* », et l'intègre dans le locus CRISPR. A l'étape suivante, ce locus est transcrit en long pré-ARNcr, puis mûri sous l'action des protéines *Cas* ou, dans certains cas, par les ribonucléases cellulaires. Ces ARNcr servent de guide pour recruter les protéines *Cas* nucléases lors de la présence de l'ADN exogène connu afin de le cliver.

Ce système est fréquemment présent chez les procaryotes, notamment chez les archées. En effet, il est identifié chez 85% des archées et 40% des bactéries (Makarova *et al.*, 2020). Les *spacers* sont considérés comme une preuve de l'infection virale dans le passé, et sont de plus en plus utilisés pour la prédiction d'hôte par des approches bioinformatiques

(Alkhnabashi *et al.*, 2019) qui seront également détaillées dans ce manuscrit.

### c. Coévolution hôte-virus : course à l'armement

Face à différents mécanismes de défense de l'hôte, les virus ont développé différentes stratégies pour leur échapper, tels que le système toxine-antitoxine, permettant de maintenir la viabilité de la cellule hôte face à l'induction de la mort cellulaire du système Abi (Stern and Sorek, 2011), ou encore le mécanisme le plus courant pour éviter la reconnaissance : la mutation.

Cette course à l'armement entre un virus et son hôte se poursuit dans le temps, et conduit donc à une adaptation et à une modification de leurs génomes. En effet, plusieurs expériences à l'échelle du laboratoire ont montré que la coévolution entre les virus et leur hôte entraînait l'accélération de la vitesse d'évolution du génome de l'hôte en y favorisant la création de nouveaux gènes, ainsi que potentiellement de nouvelles fonctions (Pal *et al.*, 2007; Meyer *et al.*, 2012; Thompson, 2012). Ainsi, les virus ont une grande influence sur l'histoire et l'évolution des organismes cellulaires.

### 3.5. Virus : régulateurs majeurs des écosystèmes

L'impact des virus dans différents écosystèmes est de nos jours une certitude. Ainsi, différentes études ont montré leur importance dans les cycles biogéochimiques, notamment dans les cycles du carbone et de l'azote (Suttle, 2007, 2016; Roux *et al.*, 2016; Gazitúa *et al.*, 2021; Rodríguez-Ramos *et al.*, 2022; Lara *et al.*). Cet impact est lié principalement aux effets de la lyse cellulaire par les virus, induisant un relargage d'éléments fondamentaux (C, N, P, S) soit sous la forme de matières organiques dissoutes (« *viral shunt* ») soit sous la forme d'agrégats de particules/composants cellulaires (« *viral shuttle* ») (Zimmerman *et al.*, 2020). Ces matières sont par la suite consommées par d'autres (micro)organismes, ce qui forme donc une boucle de recyclage de nutriments.

Les virus peuvent aussi affecter les flux de nutriments lorsque leurs génomes contiennent des gènes associés aux processus métaboliques. Différentes recherches, notamment sur les cycles de nutriments dans les océans (Roux *et al.*, 2016; Gazitúa *et al.*, 2021), ont relevé

la présence de plusieurs gènes métaboliques auxiliaires (AMGs : *auxiliary metabolic genes*) putatifs codant pour des protéines métaboliques essentielles impliquées dans la réduction et l'oxydation du soufre ou dans la transformation de l'azote.

Ces recherches ont démontré que la virosphère est fortement intégrée dans la diversité biologique et dans les processus métaboliques responsables des grands cycles biogéochimiques de la planète (Suttle, 2016).

### 3.6. Virus au sein du processus de digestion anaérobie

#### a. Etat de l'art

Comme mentionné précédemment, les virus sont ubiquitaires sur la planète. Plusieurs articles ont décrit la diversité virale dans les environnements anaérobies, qu'il s'agisse d'environnements naturels (océans profonds, sédiments marins, marais, microbiotes d'animaux, etc.) ou créés par l'activité humaine (méthaniseurs) (Danovaro *et al.*, 2001; Ortmann and Suttle, 2005; Suttle, 2007; Jackson and Jackson, 2008; Mora *et al.*, 2011; Robinson and Pfeiffer, 2014; Calusinska *et al.*, 2016; Manrique *et al.*, 2016; Hernández and Vives, 2020). Ainsi, les observations au microscope électronique ont mis en évidence une grande diversité morphologique des particules virales dans ces environnements. De plus, avec les progrès des technologies du séquençage, l'analyse de l'ensemble du contenu génomique viral d'un échantillon (métavirome), a élargi considérablement nos connaissances sur les virus.

Cependant, malgré cette grande diversité virale, très peu de virus anaérobies ont été bien identifiés et décrits dans la littérature. En effet, les recherches en virologie sont historiquement concentrées sur les virus pathogènes causant les maladies chez l'Homme, les animaux et les plantes. Le nombre de virus anaérobies isolés est donc limité par comparaison avec les virus aérobies et anaérobies facultatifs (Hernández and Vives, 2020) et la plupart de ces virus anaérobies décrits sont principalement liés aux bactéries pathogènes tels que celles appartenant aux genres *Clostridium* (par exemple : *C. perfringens*, *C. difficile*, *C. botulinum*) (K. Inoue and Iida, 1968; Roseman and Richardson,



1969; Morales *et al.*, 2012; Seal, 2013; Hargreaves and Clokie, 2015; Nale *et al.*, 2016) ou *Bacteroidetes* (par exemple : *B. fragilis*, *B. intestinalis*) (Vijayavel *et al.*, 2010; Koonin and Yutin, 2020; Park *et al.*, 2020; Tominaga *et al.*, 2020; Y. Li *et al.*, 2021). Quelques virus infectant les espèces non pathogènes de *Desulfovibrio* spp. ont été également rapportés (Handley *et al.*, 1973; Rapp and Wall, 1987; Kamimura and Araki, 1989; Walker *et al.*, 2006; Eydal *et al.*, 2009).

Concernant les communautés virales au sein des méthaniseurs, une grande diversité de particules virales a été observée dans plusieurs type de méthaniseurs (Figure 18) (Wu and Liu, 2009; Chien *et al.*, 2013; Calusinska *et al.*, 2016; Willenbücher *et al.*, 2022). Les analyses des métaviromes ont montré également une grande diversité génomique virale, permettant l'identification de plusieurs nouveaux virus. Toutefois, ces derniers ne sont pas bien caractérisés ou isolés en raison de la complexité des communautés microbiennes et de leurs interactions dans le procédé de digestion anaérobie. La mise en évidence des liens hôte-virus demeure un défi majeur dans ce type d'environnement complexe.

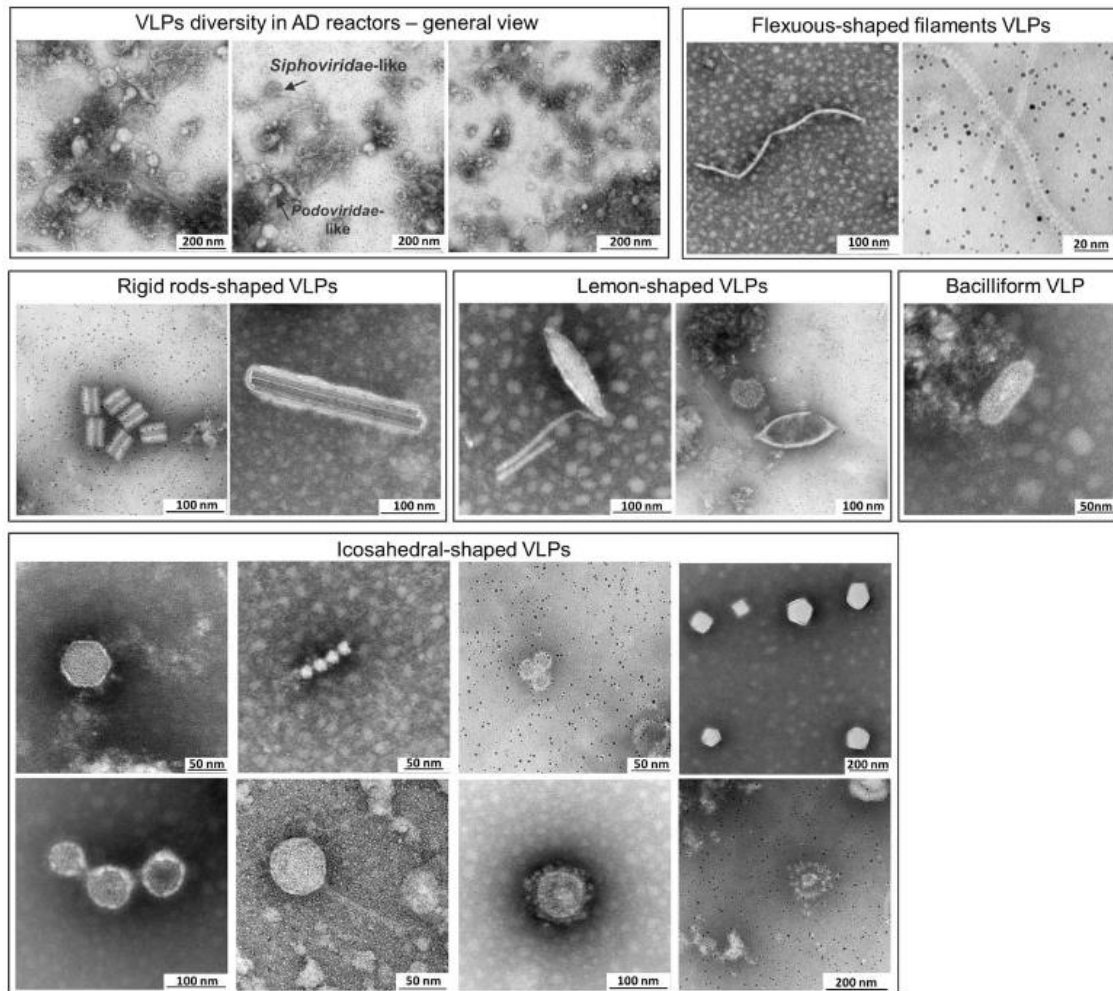


Figure 18 Aperçu de la diversité des particules virales au sein des méthaniseurs (Calusinska et al., 2016)

### b. Virus d'archées méthanogènes

Parmi les virus présents dans les écosystèmes de digestion anaérobie, les virus infectant les archées méthanogènes, actrices de l'étape ultime de la méthanisation, sont les plus décrits. Cependant, la connaissance sur ces virus reste limitée. A ce jour, seuls 9 virus sont décrits dans la littérature (Wood *et al.*, 1989; Meile *et al.*, 1989; Nölling *et al.*, 1993; Wolf *et al.*, 2019; Thiroux *et al.*, 2021; Weidenbach *et al.*, 2021; Molnár *et al.*, 2020; Pfister *et al.*, 1998; Weidenbach *et al.*, 2017; Krupovič, Forterre, *et al.*, 2010a), ainsi que quelques provirus (Luo *et al.*, 2001; Krupovič and Bamford, 2008; Krupovič, Gribaldo, *et al.*, 2010a; Chibani *et al.*, 2022) (Tableau 5). De plus, MetMV, un virus putatif détecté via des analyses métagénomiques, pourrait être le second virus décrit infectant un hôte du genre *Methanosarcina* (Molnár *et al.*, 2020), mais cela reste à confirmer.

Tableau 5. Virus et provirus d'archées méthanogènes

Nom	Morphologie	Genre de l'hôte	Origine	Génome à ADN	Référence
ΨM1	Tête-queue	<i>Methanothermobacter</i>	DA thermophile	30,4 kb - linéaire	(Meile <i>et al.</i> , 1989; Liu <i>et al.</i> , 2021)
ΨM2	Tête-queue	<i>Methanothermobacter</i>	DA thermophile	26,1 kb - linéaire	(Pfister <i>et al.</i> , 1998; Liu <i>et al.</i> , 2021)
Provirus ΨM100	-	<i>Methanothermobacter</i>	DA thermophile	28,8 kb - linéaire	(Luo <i>et al.</i> , 2001)
ΦF1	Tête-queue	<i>Methanobacterium</i>	DA thermophile	85 kb - linéaire	(Nölling <i>et al.</i> , 1993)
ΦF3	Tête-queue	<i>Methanobacterium</i>	DA thermophile	36 kb – linéaire ou circulaire	(Nölling <i>et al.</i> , 1993)
Drs3	Tête-queue	<i>Methanobacterium</i>	DA mésophile	37 kb - linéaire	(Wolf <i>et al.</i> , 2019; Liu <i>et al.</i> , 2021)
Bif4	Tête-queue	<i>Methanoculleus</i>	DA mésophile	37 kb - potentiellement circulaire	(Weidenbach <i>et al.</i> , 2021)
MetSV	Sphérique	<i>Methanosarcina</i>	DA mésophile	11 kb - linéaire	(Weidenbach <i>et al.</i> , 2017)
A3-VLP	Oblat/ fusiforme	<i>Methanococcus</i>	DA mésophile	23 kb - circulaire	(Wood <i>et al.</i> , 1989)
MFTV1	Tête-queue	<i>Methanocaldococcus</i>	Cheminée hydrothermale	31 kb -linéaire	(Thiroux <i>et al.</i> , 2021)
Provirus <i>M. fervens</i>	-	<i>Methanocaldococcus</i>	Sédiment marin	30,1 kb - linéaire	(Krupovič, Gribaldo, <i>et al.</i> , 2010a)
Provirus <i>M. janaschii</i>	-	<i>Methanocaldococcus</i>	Cheminée hydrothermale	30,4 kb - linéaire	(Krupovič, Gribaldo, <i>et al.</i> , 2010a)
Provirus <i>M. vulcanius</i>	-	<i>Methanocaldococcus</i>	Cheminée hydrothermale	30,1 kb - linéaire	(Krupovič, Gribaldo, <i>et al.</i> , 2010a)

Le virus virulent **ΨM1**, infectant *Methanothermobacter marburgensis* (ordre Methanobacteriales), a été isolé d'un digesteur thermophile (température opérationnelle de 55-60°C) (Meile *et al.*, 1989). Il s'agit d'un virus tête-queue (caudovirus) avec une tête icosaédrique d'environ 55 nm de diamètre et une queue d'environ 210 nm de longueur (Figure 19). Il contient de l'ADNdb linéaire, d'une longueur d'environ 30,4 kb. **ΨM2**, un mutant du virus ΨM1, a montré également la capacité à infecter *M. marburgensis* (Pfister *et al.*, 1998). Cette mutation est la délétion d'un fragment du génome viral de ΨM1. Le génome du virus ΨM2 a une longueur de 26,1 kb. Ces virus appartiennent à la famille *Leisingerviridae*.

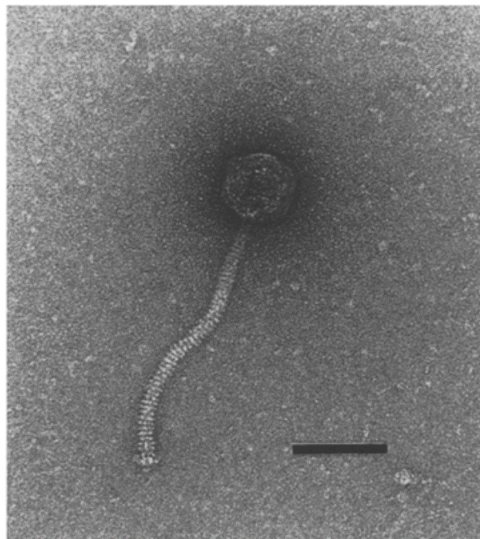


Figure 19. ΨM1 observé au MET (Meile *et al.*, 1989). L'échelle représente 70 nm

Le prophage défectif **ΨM100**, identifié au sein du génome d'un autre membre du genre *Methanothermobacter*, *Methanothermobacter wolfeii*, montre une grande similarité (70,8%) avec le génome de ΨM2 (Luo *et al.*, 2001). Son génome a une longueur de 28,8 kb et code pour une pseudomuréine endopeptidase, *PeiW*, qui est responsable de l'autolyse de l'hôte.

**ΦF1**, **ΦF3** et **Drs3** sont des virus virulents infectant des membres du genre *Methanobacterium* (ordre Methanobacteriales) (Figure 20). ΦF1 et ΦF3 ont été isolés de digesteurs thermophiles (température opérationnelle de 55°C) (Nölling *et al.*, 1993). Les deux virus possèdent une morphologie similaire mais leur taille diffère. ΦF1 est composé

d'une tête icosaédrique d'environ 70 nm de diamètre et une queue non-contractile d'environ 160 nm de long et 20 nm de large.  $\Phi$ F3 est aussi un virus tête-queue avec une tête icosaédrique d'environ 55 nm de diamètre et une queue flexible de 230 nm de long et 9 nm de large. Le génome de  $\Phi$ F1 est estimé à 83 kb, et constitué d'ADNdb linéaire, tandis que celui de  $\Phi$ F3 est estimé à 36 kb et constitué d'ADNdb linéaire ou circulaire. De plus, les deux génomes ne présentent pas de similarité. Ces différences en taille et topologie de génomes peuvent affecter l'infectiosité de ces deux virus. En effet,  $\Phi$ F1 présente une large gamme d'hôtes incluant *M. thermoformicum* (souches Z-245, FTF, FF1, FF3 et CSM3) et *M. thermoautotrophicum* alors que  $\Phi$ F3 n'infecte que la souche FF3 de *M. thermoformicum*. Concernant Drs3 (famille *Anaeroviridae*), ce virus virulent a été isolé d'un digesteur mésophile expérimental (Wolf *et al.*, 2019) avec son hôte, la souche Kh10 de *Methanobacterium formicum*. Comme  $\Phi$ F1 et  $\Phi$ F3, il s'agit d'un caudovirus avec une tête icosaédrique d'environ 60 nm de diamètre et une queue non-contractile d'environ 230 nm de longueur. Son génome complet est constitué d'ADNdb linéaire de longueur 37 kb, et contient 39 ORFs (Open reading frame : cadre de lecture ouvert) dont seulement 12 ont montré une similarité avec des gènes de fonction prédite connue.

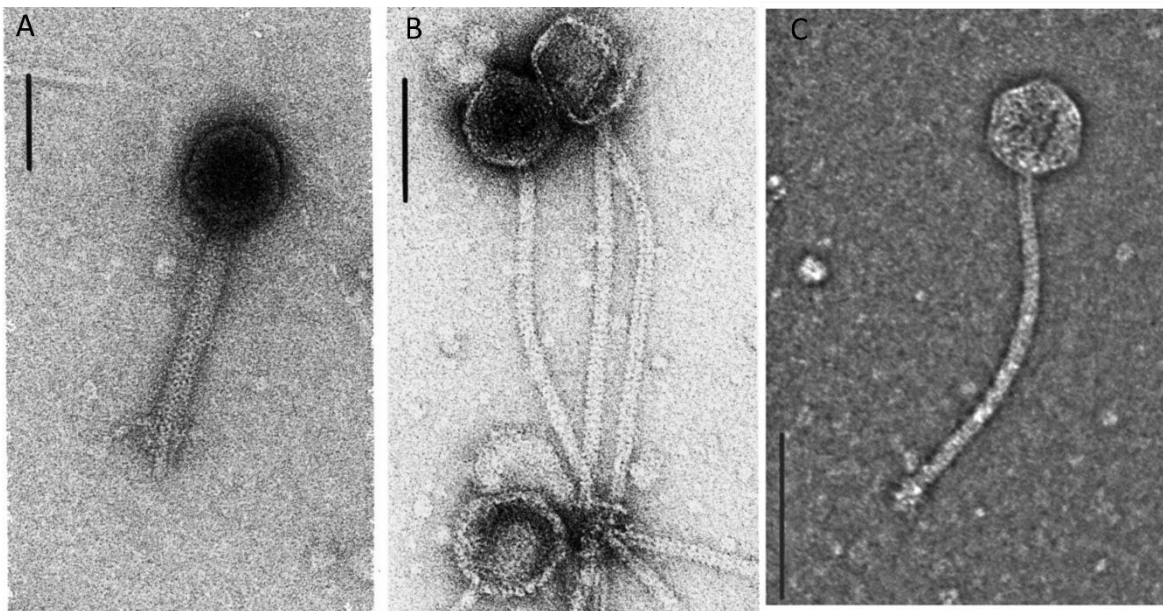


Figure 20. Observations au MET de virions infectant des membres du genre *Methanobacterium* :  $\Phi$ F1 (A),  $\Phi$ F3 (B) et Drs3 (C). Les barres représentent 70 nm pour A et B, et 100 nm pour C. (Nölling *et al.*, 1993; Wolf *et al.*, 2019)

*Methanosarcina* Spherical Virus (**MetSV**) est un virus virulent infectant plusieurs espèces mésophiles du genre *Methanosarcina* (*Methanosarcina mazei*, *Methanosarcina barkeri* et *Methanosarcina soligedili*). MetSV a été enrichi et isolé d'un méthaniseur qui digère les boues de station d'épuration à Göttingen, Allemagne. Il s'agit d'un virus dont la capsid est sphérique de  $56 \pm 1,9$  nm de diamètre, et entourée par une enveloppe (Figure 21). Son génome est constitué d'ADNdb linéaire (10 567 pb) et contient 22 ORFs dont les fonctions ont pu être prédites pour certains d'entre eux : ADN polymérase, ATPase, protéine se liant à l'ADN (*DNA-binding protein*) et protéine structurale de l'enveloppe. Il s'agit du premier virus décrit infectant des membres de l'ordre Methanosarcinales.

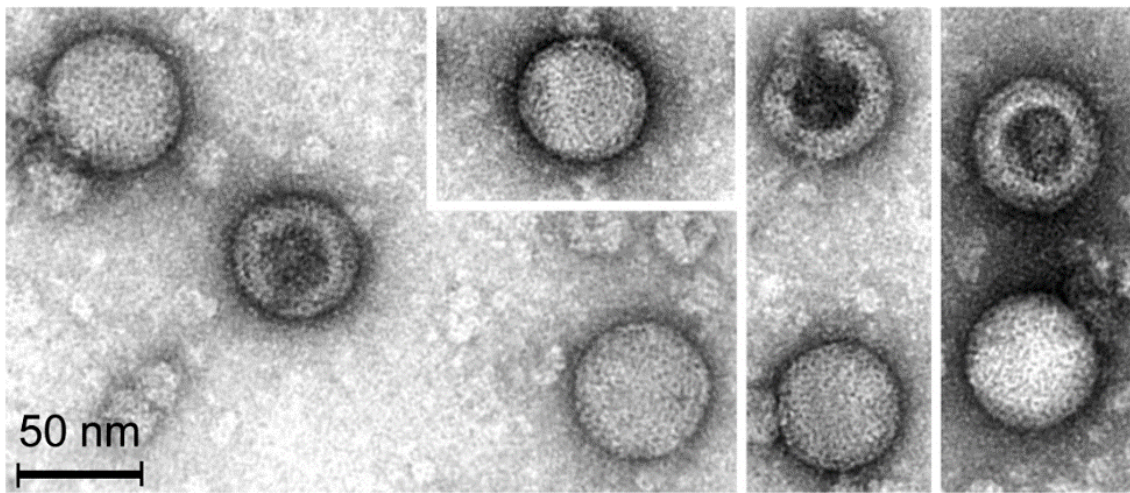


Figure 21. Virions de MetSV observés au MET (Weidenbach et al., 2017)

Récemment, Weidenbach *et al* ont décrit un nouveau virus virulent, **Blf4**, qui infecte la souche *Methanoculleus bourgensis* E02.3. Il s'agit du premier virus infectant un membre de l'ordre Methanomicrobiales décrit dans la littérature. Le virus et son hôte ont été isolés d'un digesteur commercial en Allemagne, qui traite des déchets de ferme (ensilage de maïs, fumier bovin, excréments de poulet). Les particules de Blf4 ont une tête icosaédrique de 60 nm de diamètre et une queue non contractile de 125 nm de longueur (Figure 22). Son génome complet a une longueur de 37 kb, d'ADNdb. Le génome pourrait être circulaire et il contient 63 ORFs, dont la plupart ont de fonctions potentielles prédites.



Figure 22 Virions de Blf4 observés au MET (Weidenbach et al., 2021)

Les particules pseudovirales A3 (*virus-like particles A3*, **A3-VLP**) ont été mises en évidence à partir de la culture de *Methanococcus voltae* A3, une archée méthanogène mésophile, appartenant à l'ordre Methanococcales (Wood *et al.*, 1989). Les particules présentent une morphologie oblongue, une longueur d'environ 50-70 nm, et possèdent une capsidie enveloppée (Figure 23). Elles contiennent de l'ADNdb circulaire de longueur 23 kb. L'ADN viral d'A3-VLP est capable de s'intégrer dans le génome de l'hôte. Ces particules ne semblent pas montrer de caractère infectieux.

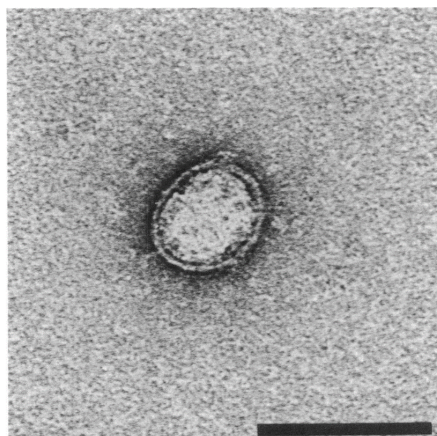


Figure 23 Particule A3-VLP observée au MET. La barre représente 100 nm (Wood et al., 1989)

Certains provirus autres que  $\Psi$ M100 ont été identifiés dans le milieu marin profond (Krupovič, Gribaldo, *et al.*, 2010a) (Tableau 5) et différents génomes de virus infectant les archées méthanogènes ont été détectés par l'analyse métagénomique dans le microbiote intestinal humain (Chibani *et al.*, 2022).

Ces virus et provirus décrits ci-dessus infectent donc quatre ordres de méthanogènes : Methanobacterales, Methanomicrobiales, Methanosarcinales et Methanococcales. Concernant les autres ordres, aucun virus n'est rapporté dans la littérature à ce jour, suggérant une grande diversité de virus d'archées méthanogènes à découvrir.

#### 4. Les méthodes moléculaires pour l'écologie microbienne

Le monde des « invisibles », des microbes, représente la diversité génétique la plus importante. En effet, ces microorganismes colonisent tous les environnements de la biosphère grâce à leur capacité d'adaptation aux variations de leurs milieux de vie. La compréhension des interactions de ces différentes communautés entre-elles et avec leurs biotopes fait partie des grandes questions en biologie. Avec l'apparition des méthodes de séquençage dans les années soixante-dix (Maxam and Gilbert, 1977; Sanger *et al.*, 1977), l'utilisation des techniques de biologie moléculaire a révolutionné la recherche sur les communautés microbiennes en permettant de s'affranchir des limites des méthodes classiques basées sur la mise en culture ou sur l'observation microscopique. Avec les progrès actuels de la technologie, les approches méta-omiques constituent des ressources importantes pour la caractérisation de la diversité génétique et fonctionnelle des communautés microbiennes au sein de leur environnement en analysant à haut débit leur ADNs (métagénomique), leur ARNs (métatranscriptomique), leur protéines (méta-protéomique) ou leurs métabolites (métabolomique). Parallèlement, de nouvelles méthodes de mise en culture et de biologie moléculaire ont également été développées, permettant la mise en évidence des interactions microbes-environnement (Figure 24).



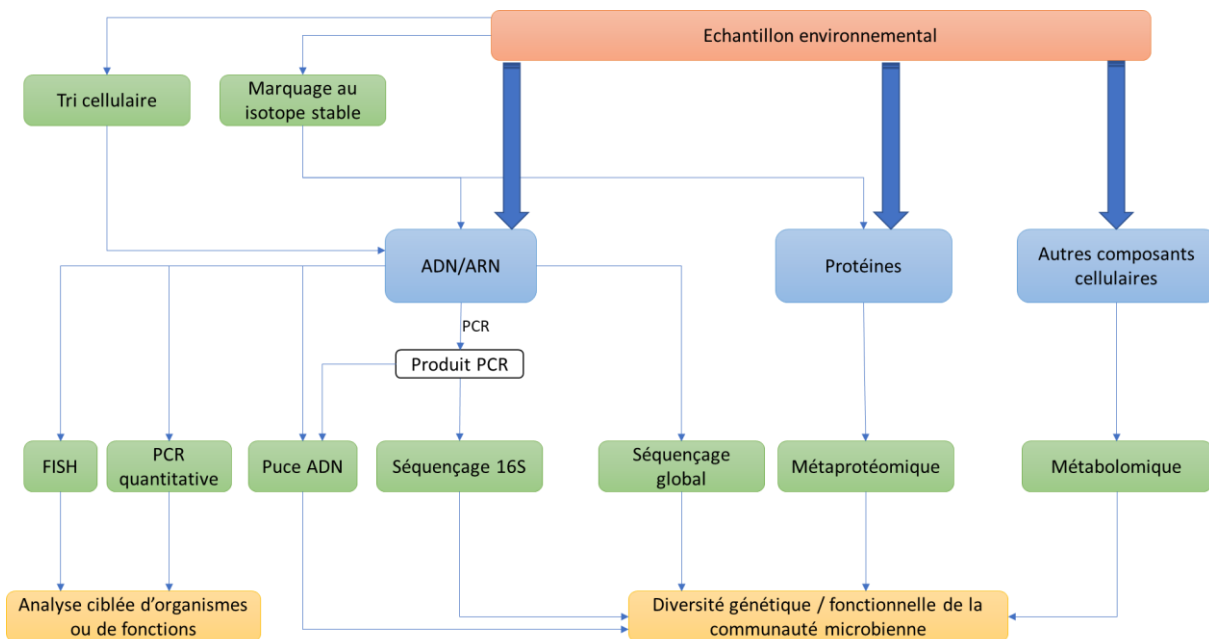


Figure 24 Méthodes culture-indépendantes pour l'étude des communautés microbiennes complexes.

#### 4.1. Marquage isotopique

L'une des plus grandes questions en microbiologie environnementale est de déterminer quels microorganismes sont impliqués dans quelles voies métaboliques, dans leur environnement naturel, autrement dit, de répondre à la question « Qui fait quoi ? ». En effet, au sein de communautés microbiennes complexes, la présence d'un microorganisme ne signifie pas nécessairement qu'il participe au métabolisme étudié. Pour établir le lien entre les microorganismes et le métabolisme d'un substrat donné, le traçage isotopique est reconnu comme une méthode efficace. En 2000, Radajewski et ses collaborateurs ont développé la méthode de marquage par un isotope stable (*Stable isotope probing* : SIP), appliquée en écologie microbienne (Radajewski *et al.*, 2000). L'avantage de cette méthode par rapport au marquage radioactif est de ne pas affecter les propriétés physico-chimiques du substrat utilisé, ainsi que l'activité métabolique des microorganismes dans l'environnement étudié, et de permettre la stabilité du signal au cours du temps. Cette méthode, comme son nom l'indique, consiste à utiliser des isotopes stables rares, tels que le  $^{13}\text{C}$  ou le  $^{15}\text{N}$ , qui représentent respectivement 1,1% et 0,364% de l'abondance de ces éléments dans le milieu naturel. L'échantillon environnemental à étudier est incubé avec un substrat

marqué avec l'isotope stable rare. Les microorganismes qui consomment ce substrat incorporent cet isotope rare dans leur matériel cellulaire et deviennent ainsi « marqués » (Figure 25). Ainsi, tous les microorganismes qui assimilent ce substrat ou l'un de ses produits de dégradation sont également « marqués ». Le SIP permet donc de suivre le devenir d'un composé étudié au sein d'un écosystème complexe et d'étudier simultanément les microorganismes associés.

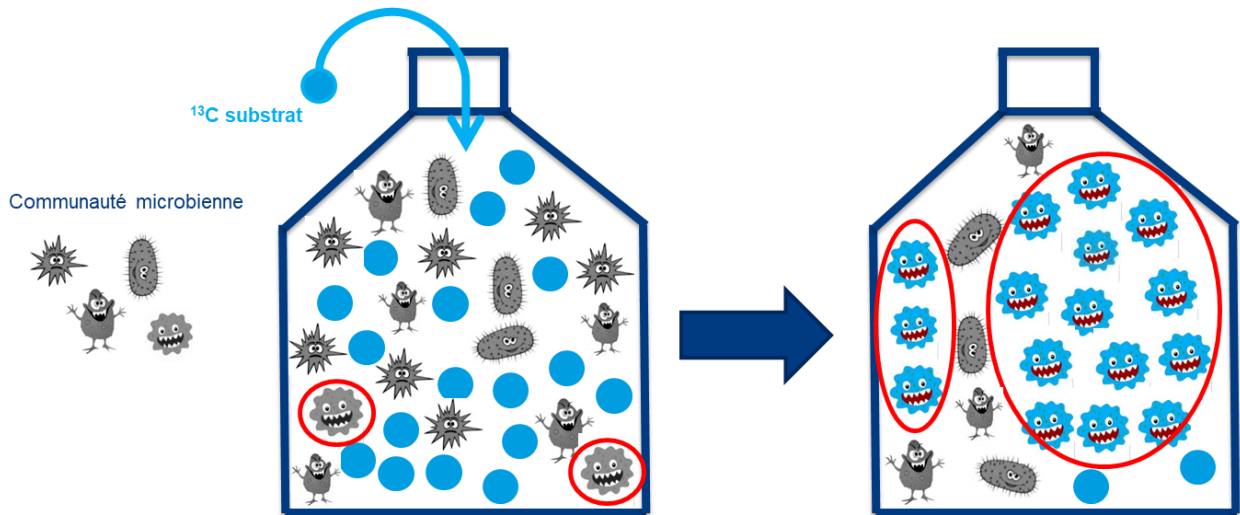


Figure 25. Illustration de l'approche de marquage isotopique

Il existe trois principales approches d'analyse de SIP, basées sur la nature du matériel cellulaire à étudier : SIP ADN, SIP ARN et SIP protéines. Le SIP ADN est le plus utilisé en raison de la faisabilité de l'approche et de la stabilité de l'ADN. Cette approche a été appliquée dans plusieurs études en écologie microbienne sur différents substrats, au sein de différents environnements (Neufeld, Schäfer, *et al.*, 2007; T. Li *et al.*, 2009; Herrmann *et al.*, 2010; Martineau *et al.*, 2010; Redmond *et al.*, 2010; Webster *et al.*, 2010; Bell *et al.*, 2011; Jones *et al.*, 2011; Eichorst and Kuske, 2012; Haichar *et al.*, 2012; Wang *et al.*, 2015; Chapleur *et al.*, 2016; Mosbæk *et al.*, 2016; Yin *et al.*, 2019). L'ADN total extrait de l'échantillon environnemental est ultracentrifugé dans une solution de Chlorure de Césium (CsCl). Cette forte centrifugation (plusieurs dizaines de milliers de g) crée un gradient de densité dans la solution et permet de séparer les ADNs « lourds » (marqués au  $^{13}\text{C}$  ou  $^{15}\text{N}$ ) qui vont migrer plus profondément que les ADNs « légers » (Figure 26).

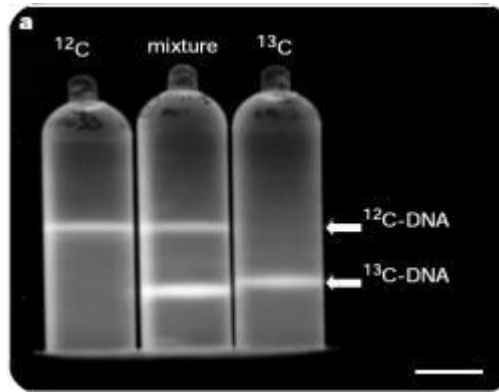


Figure 26. Tubes contenant de l'ADN  $^{12}\text{C}$  et  $^{13}\text{C}$  ultracentrifugés dans une solution de CsCl, mettant en évidence leur différence de densité (Radajewski et al., 2000)

Cette séparation dépend non seulement de l'enrichissement en isotope marqué, mais aussi du pourcentage de Guanine et Cytosine des acides nucléiques. Dans le cadre du marquage avec du  $^{13}\text{C}$ , la densité d'ADN dans un gradient de CsCl est calculée avec l'équation suivante (Birnie and Rickwood, 2014) :

$$\rho = 0.098 * (\%GC) + 0.035 * (\%^{13}\text{C}) + 1.660$$

#### 4.2. Approches métagénétiques et méta-omiques

##### a. Le gène d'ARN ribosomique 16S : marqueur universel des procaryotes

En 1977, Carl Woese et ses collaborateurs ont bouleversé la taxonomie des procaryotes en proposant l'idée d'utiliser le gène codant la petite sous-unité 16S de l'ARN ribosomique (ARNr) comme outil de classification taxonomique (Woese *et al.*, 1975; Fox *et al.*, 1977; Woese and Fox, 1977). Contrairement à la taxonomie classique basée sur les caractéristiques morphologiques ou phénotypiques à l'époque, l'idée de Woese et al. est basée sur l'accumulation de mutations au sein des génomes durant l'évolution, selon deux principes : (i) les mutations se produisent de façon aléatoire et (ii) les mutations se produisant chez un individu se conservent dans sa descendance. Le choix du gène d'ARNr 16S comme marqueur phylogénétique repose sur ses principaux avantages :

- L'ARNr est présent dans toutes les cellules procaryotes, que ce soit en une seule copie ou en de multiples copies.
- Les ARNr ont évolué plus lentement que la plupart des autres gènes et constituent

donc une horloge moléculaire lente (Alauzet, 2009).

- Le gène codant l'ARNr 16S est constitué de différentes zones hautement conservées et d'autres au contraire hypervariables. Les zones conservées servent de cibles pour des amorces utilisées lors de l'amplification d'ADN par PCR puis lors du séquençage (Woese, 1987), tandis que les zones hypervariables constituent une signature moléculaire spécifique pour chaque groupe microbien.
- Ce gène présente une taille raisonnable pour le séquençage (environ 1540 nucléotides) et contient suffisamment d'information.

Grace à cet outil, en 1990, Woese et ses collaborateurs ont redéfini le système de classification des êtres vivants (Woese *et al.*, 1990) en démontrant que ces derniers étaient divisés en trois domaines (Eucaryotes, Bactéries et Archées) (Figure 27), d'après les caractères moléculaires sur la base des séquences d'ADN codant ARNr 16S (Woese, 1987).

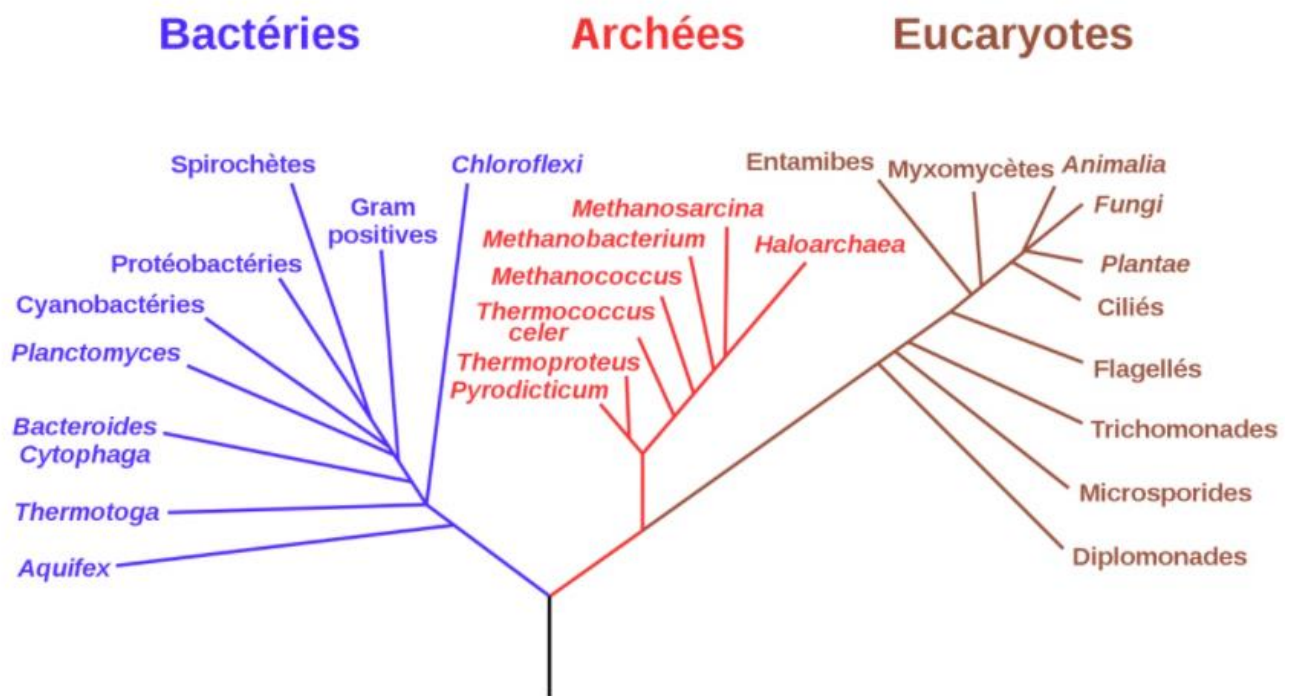


Figure 27. L'arbre des trois domaines du vivant (Source <https://papiermachesciences.org/>)

Appliquée à la microbiologie environnementale, cette approche permet d'identifier la présence de différents procaryotes au sein d'un échantillon, reflétant la diversité microbienne.

## b. Les approches méta-omiques

Pour décrire une communauté microbienne, on peut utiliser les données de séquençage du gène d'ARNr 16S pour y identifier les microorganismes présents. Cependant, pour le mieux comprendre, il est nécessaire d'analyser le lien entre la diversité génétique/fonctionnelle et leur régulation répondant aux différents facteurs environnementaux. Il existe quatre niveaux d'analyse, correspondant à quatre approches méta-omiques distinctes (Figure 28) :

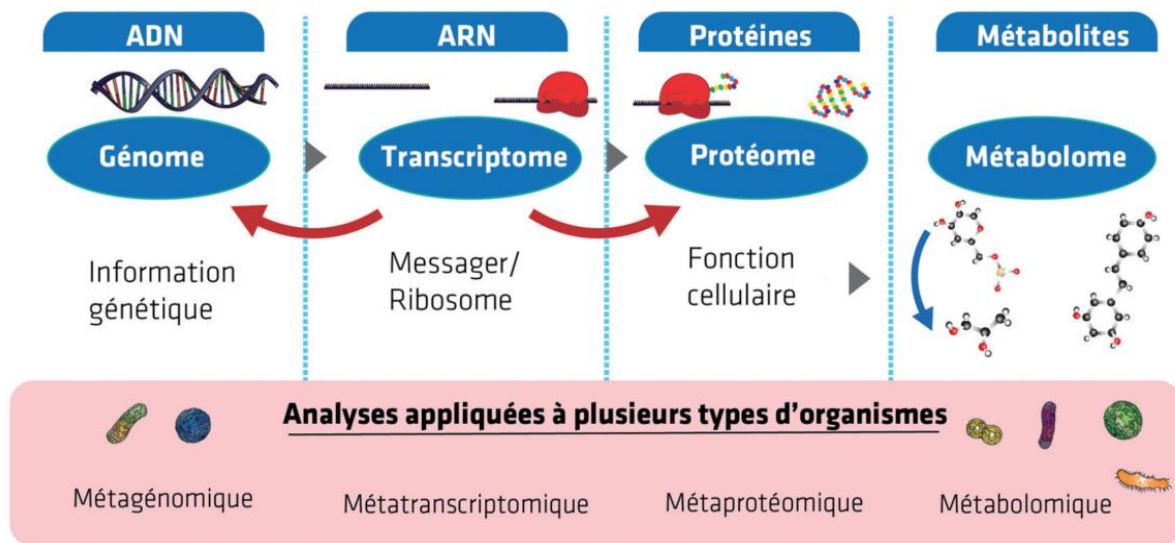


Figure 28. Différents types d'approches méta-omiques pour l'étude des communautés microbiennes (messenger = ARN messenger, Ribosome = ARN ribosomique), modifiée à partir de (Bouchez et al., 2016)

- L'analyse d'ADN (**Métagénomique**) : le séquençage du contenu génomique entier d'un échantillon permet non seulement d'identifier les microorganismes présents dans cet échantillon mais aussi de fournir un répertoire de gènes pour identifier les différentes fonctions potentielles et les voies de régulation de certains gènes. Avec les progrès de la technologie de séquençage permettant d'obtenir un haut débit et de diminuer les coûts expérimentaux, cette approche est actuellement très utilisée dans l'étude des communautés microbiennes. Cependant, la présence de l'ADN d'un microorganisme dans un échantillon ne signifie pas que ce microorganisme est vivant ou que tous les gènes associés sont actifs.
- L'analyse d'ARN (**Métatranscriptomique**) : cette approche basée sur le séquençage

de l'ensemble des transcrits (ARN) des gènes permet d'identifier la composition taxonomique des microorganismes actifs d'un échantillon en analysant le contenu en ARN ribosomique. Par ailleurs, la caractérisation de l'ARN messenger met en évidence les fonctions actives des microorganismes viables dans cet échantillon, permettant de s'affranchir d'une grande limite de l'approche métagénomique. Cette approche métatranscriptomique est de plus en plus utilisée pour étudier l'aspect fonctionnel ou l'effet d'un stimulus sur un écosystème microbien. Toutefois, le lien entre l'expression des gènes et l'activité protéique est affecté par différents mécanismes, tels que, la dégradation des ARN ou la régulation de la synthèse des protéines.

- L'analyse des protéines (**Métaprotéomique**) : cette approche consiste à analyser l'ensemble des protéines synthétisées par les communautés microbiennes dans un échantillon à un instant donné. Elle apporte des informations complémentaires à celles issues de la métagénomique ou de la métatranscriptomique. L'intérêt de cette approche est non seulement d'identifier et de quantifier les protéines mais aussi d'évaluer la régulation de l'expression des protéines dans des conditions déterminées en ciblant les protéines (enzymes) liées au stress (Ameen and Raza, 2017). Deux techniques sont principalement utilisées. La première est basée sur le couplage de l'électrophorèse sur gel 1D ou 2D avec la spectrométrie de masse (MS) (Maron *et al.*, 2007; Bouchez *et al.*, 2016). La deuxième, appelé protéomique *shotgun* est basée sur le couplage de chromatographie en phase liquide (HPLC) et de spectrométrie de masse en tandem (MS/MS), suite à une étape de digestion des protéines (McDonald and Yates III, 2002; Zhu, 2018) (Figure 29).

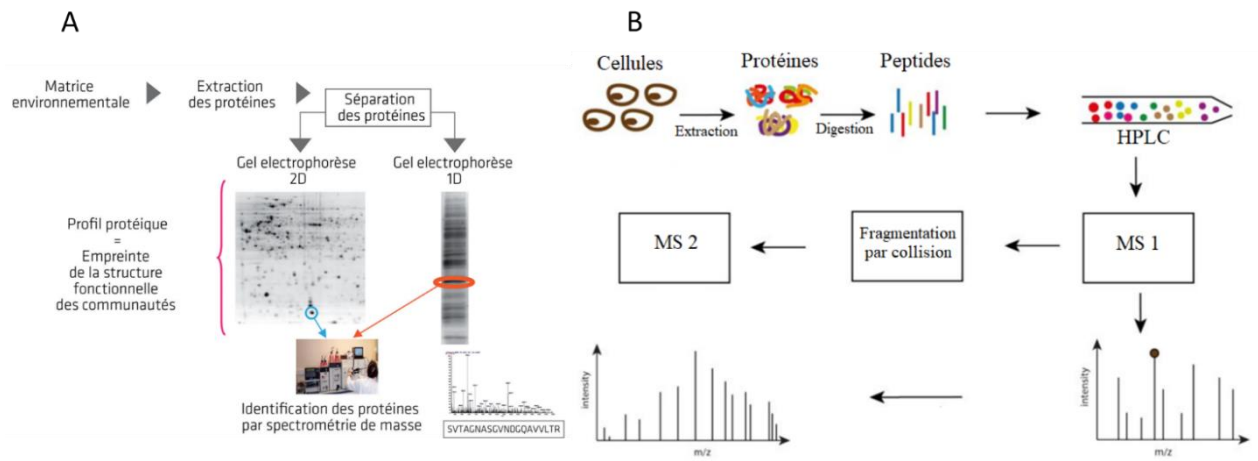


Figure 29 Techniques d'analyse de métaprotéomique (A : couplage entre de l'électrophorèse sur gel 1D ou 2D et de la MS, B : couplage entre de la HPLC et de la MS/MS) (Bouchez et al., 2016; Zhu, 2018)

- L'analyse des métabolites (**Métabolomique**) : la métabolomique environnementale implique la recherche de tous les métabolites primaires et secondaires produits par les microorganismes. Cette approche permet (i) d'identifier les voies métaboliques actives, (ii) d'évaluer la dynamique de ces voies dans des conditions déterminées (mécanismes d'adaptation au stress par exemple). Les techniques principalement utilisées sont la MS et la résonance magnétique nucléaire (RMN).

L'analyse des résultats de ces approches méta-omiques nécessite la mise en place d'outils bioinformatiques performants pour traiter ces flux de données de « haut débit ».

#### 4.3. Focus sur la prédiction d'hôte pour les virus

Dans le cadre des études en écologie microbienne, il est primordial de déterminer les hôtes des virus, puisque ces derniers sont dépendant de leur hôte pour leur réplication. Ces connaissances sont indispensables pour pouvoir appréhender l'impact des virus au niveau des communautés ou de l'écosystème. Différentes nouvelles approches expérimentales et bioinformatiques ont été développées depuis la dernière décennie afin de mettre en évidence le lien entre un virus et son hôte (Figure 30) (Edwards *et al.*, 2016; Coclet and Roux, 2021).

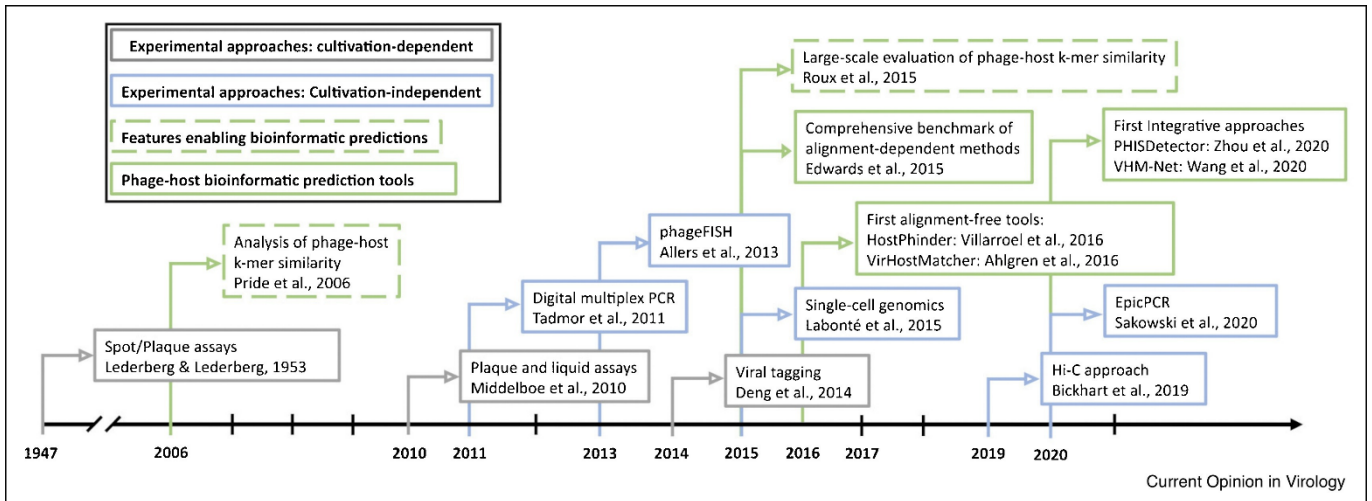


Figure 30 Chronologie globale des approches bioinformatiques et expérimentales utilisées pour établir les liens hôtes-virus (Coclet and Roux, 2021).

### a. Approches expérimentales

Les méthodes dites « traditionnelles » sont basées sur la formation des plaques de lyse, ce qui nécessite le travail en cultures de souches pures (virus et microorganisme). Ces méthodes dépendent de la cultivabilité de l'hôte et de l'infectiosité du virus. En effet, les virus peu virulents ou lysogènes ne tuent pas (ou pas suffisamment) de cellules pour être détectés. Afin de contourner ces biais, de nouvelles méthodes ont été développées, permettant d'analyser plusieurs interactions dans un échantillon environnemental.

Le marquage viral (*Viral tagging*) consiste à trier des cellules auxquelles des virus, marqués par un fluorophore, sont adsorbés, en utilisant la technique de cytométrie en flux. Par la suite, l'ADN total issu des cellules triées est extrait. Les cellules ayant été préalablement marquées avec un isotope stable, les ADN viraux et cellulaires peuvent ensuite être séparés au sein d'un gradient de densité, par ultracentrifugation. Les ADN viraux peuvent alors être séquencés (Mosier-Boss *et al.*, 2003; Deng *et al.*, 2014). Cette approche permet de détecter un lien entre le virus et la cellule hôte. Cependant, l'adsorption n'équivaut pas nécessairement à l'infection virale. Il existe des cas où les virus sont adsorbés aux cellules mais ne les infectent pas. En particulier, certaines bactéries ont la capacité de consommer les virus pour leur croissance, selon des mécanismes qui restent à déterminer (Godon *et al.*, 2021).



Alternativement, des approches indépendantes de la culture ont été développées. La PCR digitale microfluidique (Tadmor *et al.*, 2011) ou PCR digitale en gouttelettes (Morella *et al.*, 2018) permet de co-localiser les gènes marqueurs des virus et des cellules. Pour cela, les cellules sont triées, isolées dans les chambres de réaction de la puce microfluidique, ou encapsulées dans des microgouttelettes, de sorte que la majorité des chambres/gouttelettes soient vides ou contiennent une seule cellule microbienne. Certaines microgouttelettes unicellulaires peuvent contenir une cellule comprenant des virions ou des génomes viraux dans le milieu intracellulaire, ou une cellule à laquelle le virus qui l'infecte a adhéré. Une PCR avec des rapporteurs fluorescents viraux est utilisée pour détecter ces chambres/microgouttelettes contenant un virus associé à une cellule. Une autre méthode, basée sur la fluorescence, le *PhageFISH* (Allers *et al.*, 2013), combine la co-détection d'un gène viral spécifique et de l'ARN ribosomique, à l'échelle de la cellule, par des observations en microscopie. Cette approche permet de détecter l'ADN phagique en répllication ou encapsidé, tout en identifiant et quantifiant les cellules hôtes à tous les stades de l'infection. Sans utiliser de marqueurs fluorescents, l'epicPCR (Emulsion, Paired Isolation and Concatenation PCR) (Spencer *et al.*, 2016) vise à créer un lien physique entre le génome du virus et celui de son hôte. Cette approche comprend une étape d'encapsulation de cellules individuelles dans des microgouttelettes, suivie par une étape de PCR de fusion d'un gène marqueur viral avec celui cellulaire. De nos jours, ces approches sont de plus en plus utilisées (Jahn *et al.*, 2021; Sakowski *et al.*, 2021; Tran and Anantharaman, 2021; Jang *et al.*, 2022; Ostenfeld *et al.*, 2022). Toutefois, une grande limite est la conception des amorces de PCR. Comme il n'existe pas de gène universel chez les virus, ces approches sont spécifiques et nécessitent donc de cibler un groupe de virus et d'avoir des connaissances préalables sur leurs génomes.

Le séquençage de cellule unique est basé sur l'amplification puis le séquençage de l'ADN total présent dans une cellule unique. Cet ADN total peut contenir de l'ADN viral (provirus ou virus en cours de cycle infectieux) et peut donc être un autre moyen pour identifier les associations hôte-virus sans avoir besoin de connaissance préalable sur le génome viral.

En analysant l'ADN des cellules uniques venues d'échantillons marins, l'hôte de bactériophages précédemment inconnus a été identifié (Roux *et al.*, 2014). L'ADN viral et celui de son hôte peuvent aussi être liés *via* une approche basée sur le contact physique (structure spatiale) de leur chromosome : Hi-C (high-throughput chromosomal confirmation capture) ou sa variante Meta3C (Marbouty *et al.*, 2014; Marbouty and Koszul, 2015), ou encore plus récemment XRM-Seq (Ribosome cross-linking and sequencing) (Brito, 2021), une version ciblant les ARN.

#### a. Approches bioinformatiques

La prédiction des hôtes des virus fait partie des défis principaux dans les études de métagénomique virale, pour mieux comprendre le rôle des virus au sein de l'échantillon environnemental étudié. De nos jours, le développement d'approches et d'outils bioinformatiques pour la prédiction des hôtes est en croissance. Ces approches impliquent l'analyse d'un ou plusieurs signaux génétiques, tels que l'homologie entre les séquences virale et cellulaires, ou la similarité des profils nucléotidiques (signatures génomiques) du virus et de l'hôte en raison de leur co-évolution. Les principaux signaux, le principe et les limites de chaque signal sont présentés dans le tableau 6.

Tableau 6. Différentes stratégies pour la prédiction d'hôte des virus, modifié à partir de (Edwards *et al.*, 2016)

Signal	Principe	Commentaires (limites)
<b>Profil d'abondance</b>	Les virus se reproduisent dans un environnement seulement si leur hôte est également présent. Les modèles d'abondance de virus et de microorganismes dans les métagénomiques peuvent être utilisés pour identifier leur association par corrélation	Le protocole de métagénomique affecte la sensibilité de détection des virus et des procaryotes dans un échantillon. Des processus écologiques tels que Kill-the-Winner (Rodriguez-Brito <i>et al.</i> , 2010) peuvent conduire à une dynamique non linéaire qui confond les corrélations directes. La stratification des échantillons par environnement peut améliorer les performances
<b>Homologie génétique</b>	L'homologie génétique entre le virus et les séquences nucléotidiques et protéiques des procaryotes peut représenter des séquences qui ont été acquises par un virus au cours d'un événement d'infection passé.	Cette approche dépend de la richesse de la base de données de référence. Certaines familles de gènes sont plus susceptibles au transfert horizontal, ce qui entraîne des difficultés dans la prédiction d'hôte.
<b>CRISPR</b>	Suite à une infection virale, il peut arriver qu'un fragment de 25 à 75 pb d'une séquence virale soit intégrée dans des CRISPR arrays du génome de l'hôte, sous la forme d'un spacers. Les régions CRISPR peuvent être identifiées dans les génomes procaryotes, et les spacers alignés contre les génomes viraux pour détecter les infections récentes. La présence de plusieurs spacers entre un microorganisme et un virus améliorent le signal.	Seulement ~40 % des bactéries et ~85 % des archées possèdent un système CRISPR (Makarova <i>et al.</i> , 2020), et les spacers d'un réseau CRISPR sont rapidement retournés dans l'environnement. La plupart des spacers CRISPR ne correspondent à aucune séquence connue, donc, bien que cette approche soit spécifique (peu de faux positifs), elle est peu sensible (beaucoup de faux négatifs).
<b>Alignement exact</b>	Les alignements exacts entre les génomes viraux et des microorganismes peuvent se trouver au niveau des sites d'intégration, des spacers CRISPR, des régions d'homologie génétique ou des prophages intégrés.	Les alignements de longueur courte ne contiennent pas de signal significatif car ils peuvent être produits de façon aléatoire.
<b>Profil d'oligonucléotide (Signatures de k-mers)</b>	Au cours du temps, la composition en k-mers (mots de taille k dans le génome) des virus devient généralement similaire à celle de l'hôte, en raison de leur co-évolution. Cette composition est influencée par les pools de nucléotides intracellulaires, l'usage des codons et la disponibilité des ARN de transfert (ARNt), ainsi que la présence de systèmes de restriction-modification.	Contrairement à ce signal, il a souvent été observé que les prophages ont un profil de composition en oligonucléotides différent de celui du génome de l'hôte. En effet, certains génomes viraux codent pour des gènes d'ARNt, si bien que le profil optimal d'utilisation des codons est différent pour le virus et pour l'hôte. Le contenu en GC est une mesure d'identification qui n'a pas beaucoup de pouvoir discriminant.

En se basant sur ces signaux, différents outils de prédiction ont été développés ces dernières années (Tableau 7). On peut les classer en trois catégories :

- Les approches alignement-dépendantes : certaines évaluent la similarité entre les génomes viraux et ceux des procaryotes, telles que Blastn (outil générique pour la comparaison de génomes) et SpacePharer (R. Zhang *et al.*, 2021), en supposant que l'échange d'information génétique entre ces derniers est indicative d'une association hôte-virus (Coutinho *et al.*, 2021). D'autres se basent sur la similarité entre les génomes viraux et des gènes marqueurs connus, par exemple : RaFAH (Coutinho *et al.*, 2021), vHULK (Amgarten *et al.*, 2020) et VPF-Class (Pons *et al.*, 2021). Dans tous les cas, ces approches dépendent fortement de la richesse des bases de données.
- Les approches alignement-indépendantes : ces approches établissent les couples hôte-virus sur la base soit du profil nucléotidique de génome viral entier, tel que HostPhinder (Villarroel *et al.*, 2016), ILMF-VH (Liu *et al.*, 2019), VirHostMatcher (Ahlgren *et al.*, 2017), WiSH (Galiez *et al.*, 2017) et PHP (Lu *et al.*, 2021) ; soit de l'ensemble des protéines prédites pour ce virus tel que: PredPhi (M. Li *et al.*, 2021), et la méthode de Leite (Leite *et al.*, 2018) ou de Boeckeaerts (Boeckeaerts *et al.*, 2021). Ces méthodes permettent de prédire les hôtes des virus sans avoir besoin de la présence d'un virus de référence similaire. Néanmoins, leur précision est plus faible par rapport à celle des méthodes d'alignement.
- Les approches intégratives : celles-ci reposent sur le couplage d'analyses issues des deux approches précédentes, afin d'améliorer les prédictions. Les deux approches connues à l'heure actuelle, PHISDetector (Phage–Host Interaction Signals Detector) (F. Zhang *et al.*, 2020) et VirHostMatcher-Net (Wang *et al.*, 2020), analysent les signaux basés sur (i) l'alignement entre séquences virales et celles de l'hôte, (ii) la présence de *spacers* partagés et (iii) la similarité de la composition en k-mer entre le virus et l'hôte (Coclet and Roux, 2021; Pappas *et al.*, 2021). Ces approches récentes pourraient améliorer la précision de la prédiction, mais elles nécessitent encore d'être évaluées dans un contexte de communautés microbiennes complexes.

Tableau7. Différents outils pour la prédiction d'hôte des virus, d'après (Coclet and Roux, 2021)

Catégorie	Outil	Date	Base de données de référence	Signaux	Avantages	Limites
Approches alignement-dépendantes	Blastn	NA	Gènes, CRISPR spacers et génomes d'hôtes	CRISPR spacers, AMGs, ARNt	Haute spécificité, jusqu'au niveau de la souche	Forte dépendance à la base de données des séquences de l'hôte, entraînant souvent un faible rappel. Cela peut être partiellement atténué en utilisant des génomes microbiens non cultivés (SAG* et MAG) provenant d'échantillons / écosystèmes pertinents.
	SpacePHARER	2020	CRISPR spacers d'hôtes	CRISPR spacers	Potential d'identification d'hôte(s) de virus sans rapport avec un génome de référence de virus.	Ne peut être appliqué qu'aux hôtes qui encodent un système CRISPR. Aspect aléatoire en fonction des spacers présents dans les génomes des hôtes et du contenu des bases de données.
	RaFAH	2020	Gènes marqueurs/ profil HMM des virus	Profil HMM	Haute précision pour les virus liés à des références connues, ne dépend pas d'une base de données de référence d'hôte.	Applicable uniquement aux virus partageant au moins 1 gène marqueur avec un virus de référence déjà connu.
	vHULK	2020	Gènes marqueurs/ profil HMM des virus	Familles de protéine (pVOGs)		
	VPF-Class	2021	Gènes marqueurs/ profil HMM des virus	Familles de protéine (VPF)		

<b>Approches alignement-indépendantes</b>	HostPhinder	2016	Génomomes viraux	Fréquence nucléotidique 16-mer	Indépendant de la prédiction des gènes viraux et de la base de données de référence de l'hôte.	Applicable uniquement aux virus similaires à au moins 1 génome viral de référence.
	ILMF-VH	2019	Génomomes viraux	Fréquence nucléotidique 6-mer		
	VirHostMatcher	2017	Génomomes d'hôtes	Fréquence nucléotidique 6-mer	Indépendant de la prédiction des gènes et de la base de données de référence sur les virus. Capable d'identifier des hôtes pour des virus entièrement nouveaux.	Le taux élevé de faux positifs par rapport aux autres méthodes nécessite souvent des tests statistiques supplémentaires et/ou des approches consensuelles pour traiter les résultats.
	WiSH	2017	Génomomes d'hôtes	Fréquence nucléotidique 8-mer		
	PHP	2020	Génomomes d'hôtes	Fréquence nucléotidique 4-mer		
	PredPhi	2020	Génomomes d'hôtes	Propriétés des séquences protéiques	Généralement une bonne précision, en particulier lorsque la base de données d'hôte comprend des microbes non cultivés (SAG/MAG) provenant d'échantillons/ écosystèmes pertinents.	Manque de benchmark pour la méthode de Boeckaerts (testée uniquement sur <i>E.coli</i> et <i>C.difficile</i> ).
	Méthode de Leite	2018	Génomomes d'hôtes	Propriétés des séquences protéiques		
	Méthode de Boeckaerts	2021	Protéomes partiels d'hôtes	Receptor-binding proteins (RBP)		
<b>Approches intégratives</b>	PHISDetector	2020	Génomomes d'hôtes et de virus	Critères multiples	Précision et sensibilité améliorées en considérant et en intégrant plusieurs signaux.	Approches récentes nécessitant une évaluation plus avancée. Nécessite plus de temps d'analyse et de ressources informatique
	VirHostMatcher Net	2020	Génomomes d'hôtes	Critères multiples		

\*SAG : single-cell assembled genome

## Présentation générale des travaux de thèse

Le chapitre « Synthèse bibliographique » a apporté les connaissances de base sur la digestion anaérobie et les communautés microbiennes (procaryotes et virus) associées. Ce chapitre a également présenté les différentes techniques couramment utilisées en microbiologie environnementale, ainsi que leurs avantages et limites.

Etablir le lien entre les virus et leurs hôtes dans les procédés de digestion anaérobie est toujours difficile en raison de la diversité et de la complexité de ces communautés microbiennes. Dans ce manuscrit, j'ai développé une méthode couplant le marquage isotopique (SIP) et différentes approches bioinformatiques pour prédire les relations hôte/virus. Ainsi, le SIP permet de simplifier les communautés microbiennes à étudier en sélectionnant par marquage les populations impliquées dans le métabolisme d'une molécule pré-sélectionnée. Le contenu génomique de ces microbes est utilisé pour la construction de bases de données manuelles (de *spacers*, de profils de signature des hôtes) permettant d'améliorer la précision des méthodes de prédiction d'hôte par des approches bioinformatiques.

Mes travaux de thèse sont présentés sous forme de trois chapitres. **Le chapitre 1** décrit des tests effectués pour optimiser les expérimentations ultérieures, ainsi qu'une première série d'incubations en microcosmes, pour identifier les virus infectant des archées méthanogènes, acteurs clés de la méthanisation. Durant cette première incubation, les populations bactériennes et virales impliquées dans le métabolisme du formate ont également été analysées. **Le chapitre 2** présente une deuxième série d'expériences qui a pour but de mieux comprendre les interactions hôte-virus au cours de la méthanisation de formate, notamment grâce à l'obtention de davantage de jeux de données métagénomiques permettant de comparer avec ceux obtenus précédemment, et d'étendre l'analyse aux hôtes bactériens comme archées. Parallèlement, un développement méthodologique s'appuyant sur l'approche SIP appliquée directement à l'ADN viral a été effectué. Comme les virus utilisent les matières de leur hôte pour se répliquer, les virus qui infectent les cellules marquées par SIP devraient théoriquement être marqués également, y compris leur ADN. **Le chapitre 3** présente la preuve de concept du SIP-ADN viral avec un modèle de souche pure, *Escherichia coli* et son bactériophage T4.

# Matériels et Méthodes

## 1. Système expérimental

Pour étudier la diversité microbienne au sein du processus de méthanisation, des incubations en microcosmes de digestion anaérobie ont été effectuées (Figure 31). Il s'agissait de flacons plasma en verre borosilicaté dans lesquels un inoculum anaérobie était introduit (culot de digestat d'ordures ménagères stabilisé), ainsi qu'une solution minérale pour la méthanisation (Norme NF EN ISO 11734) et un substrat spécifique servant de source de carbone unique et favorisant la méthanogenèse. Le formate, l'acétate ou le méthanol (Concentration de C à 200 mM) ont été employés selon les cas. Ces flacons avaient un volume total de 500 mL ou 2 000 mL selon les expériences. Le ratio entra du volume de travail et du ciel gazeux est 6 : 10 dans toutes les séries d'expériences. Ils étaient fermés hermétiquement par un septum en caoutchouc d'une épaisseur de 5,5 mm et par une bague à vis en aluminium de 40 mm de diamètre, permettant de maintenir des conditions anaérobies. Un système de vanne a aussi été installé afin de faciliter les prélèvements de gaz et de liquide. Ces microcosmes ont été incubés à 35°C pour toutes les expériences décrites dans ce manuscrit.

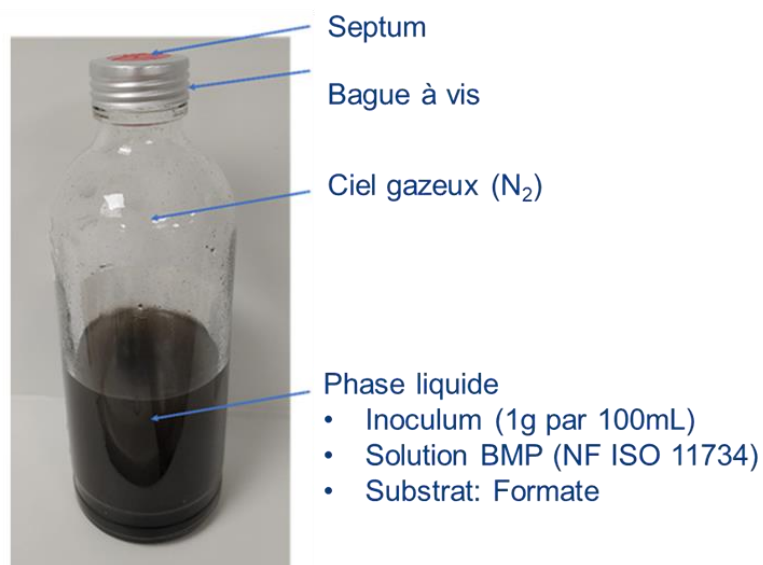


Figure 31. Système expérimental pour les incubations en microcosmes



## 1.1. Phase liquide

### a. Inoculum

Pour pouvoir effectuer les incubations en microcosmes, la préparation d'un inoculum issu de digestats de méthaniseur a été réalisée. Un volume de 10 L de digestat a été prélevé dans un méthaniseur de laboratoire de 100 L qui digère des biodéchets en mode semi-continu et fonctionne en conditions mésophiles ( $35 \pm 1^\circ\text{C}$ ). Le digestat prélevé a été par la suite incubé en conditions anaérobies à  $35^\circ\text{C}$  pendant 20 jours, jusqu'à l'arrêt complet de la production de biogaz. Cette étape de stabilisation permettait la dégradation de l'essentiel des sources carbonées fermentescibles présentes dans le milieu. Elle avait pour but d'éviter un apport trop important de  $^{12}\text{C}$ , dans la mesure où un substrat marqué au  $^{13}\text{C}$  a été utilisé dans la plupart des incubations, et qu'une dilution du signal n'était pas souhaitable.

Après cette étape de stabilisation, le liquide a été centrifugé à 10 000g pendant 15 minutes à  $15^\circ\text{C}$ . Le culot a été collecté, aliquoté puis conservé à  $-80^\circ\text{C}$ . Ce culot été utilisé comme inoculum pour l'ensemble des expériences en microcosmes. La caractérisation de cet inoculum est présentée dans le tableau 8.

Tableau 8. Caractéristiques de l'inoculum préparé

<b>Demande chimique en oxygène (mg/mL)</b>	<b>800</b>
<b>Matière volatile/inoculum (g/g)</b>	<b>0.064</b>
<b>Archées méthanogènes dominantes</b>	<b>Ordre Methanosarcinales</b>

### b. Substrat

Pour étudier l'effet du substrat sur la production de biogaz, ainsi que sur la composition des communautés microbiennes au sein du processus de digestion anaérobie, et en particulier des archées méthanogènes, trois substrats ont été employés au cours d'une première série d'expérience : le formate, l'acétate et le méthanol. Les résultats sont présentés

dans le chapitre 1. Par la suite, le formate a été choisi pour des analyses plus approfondies grâce à deux séries d'expériences supplémentaires, notamment en utilisant du  $^{13}\text{C}$ -formate.

### c. Solution minérale

La solution minérale employée est nommée BMP (Biochemical Methane Potential). Elle permet la viabilité des microorganismes impliqués dans la digestion anaérobie et répond à leurs besoins en macro- et micro-nutriments. Elle ne comporte aucune source de carbone organique ou inorganique et est souvent employée dans des incubations ayant pour but de déterminer la quantité maximale de méthane pouvant être produit à partir de la matière organique introduite. La composition de ce tampon est décrite dans la norme NF EN ISO 11734 (Tableau 9). Dans le cadre des expériences présentées dans ce manuscrit, le milieu BMP a été préparée à une concentration intermédiaire de 5X et la solution d'éléments traces à une concentration intermédiaire de 1000X.

Tableau 9. Milieu BMP 1X et solution mère d'éléments traces 100X (Norme NF ISO 11734)

Milieu BMP		Solution mère d'élément traces	
$\text{KH}_2\text{PO}_4$	0,27 g	$\text{MnCl}_4 \cdot 4 \text{H}_2\text{O}$	0,05 g
$\text{Na}_2\text{HPO}_4 \cdot 12 \text{H}_2\text{O}$	1,12 g	$\text{H}_3\text{BO}_3$	0,005 g
$\text{NH}_4\text{Cl}$	0,53 g	$\text{ZnCl}_2$	0,005 g
$\text{CaCl}_2 \cdot 2 \text{H}_2\text{O}$	0,075 g	$\text{CuCl}_2$	0,003 g
$\text{MgCl}_2 \cdot 6 \text{H}_2\text{O}$	0,10 g	$\text{Na}_2\text{MnO}_4 \cdot 2 \text{H}_2\text{O}$	0,001 g
$\text{FeCl}_2 \cdot 4 \text{H}_2\text{O}$	0,02 g	$\text{CoCl}_2 \cdot 6 \text{H}_2\text{O}$	0,1 g
$\text{Na}_2\text{S} \cdot 9 \text{H}_2\text{O}$	0,1 g	$\text{NiCl}_2 \cdot 6 \text{H}_2\text{O}$	0,01 g
Solution mère d'éléments traces	10 mL	$\text{NaSeO}_3$	0,005 g
Eau désoxygénée, q.s.p	1 L	Eau, q.s.p	1 L

## 1.2. Phase gazeuse

L'air initialement présent dans les microcosmes a été purgé et remplacé par de l'azote 99% (Linde), en effectuant plusieurs cycles consécutifs « vide-gaz » à l'aide d'une rampe à vide. Le temps de chaque cycle dépendait du volume du ciel gazeux des flacons utilisés (15 secondes pour ceux de 500 mL et 1 minute pour ceux de 2 L).

## 1.3. Mise en incubation des microcosmes

Lors des différentes séries d'expériences, les diverses conditions opératoires ont systématiquement été réalisées en triplicata. Le volume de travail (volume de la phase liquide) était de 300 mL pour les réacteurs de 500 mL, et de 1,2 L pour ceux de 2 L. Pour chaque 100 mL de phase liquide, ont été employés : 1 g d'inoculum décongelé, 20 mL de solution minérale BMP 5X, 0,1 mL de solution d'éléments traces 1 000X, le substrat choisi en quantité correspondant à une concentration finale de 200 mM de carbone, et de l'eau distillée pour ajuster le volume.

Après préparation de la phase liquide, le flacon a été fermé hermétiquement avec un septum en caoutchouc et une bague à vis métallique. Le ciel gazeux des microcosmes a été purgé à l'azote à l'aide d'une rampe à vide. Trois cycles vide-gaz ont été effectués afin d'assurer des conditions anaérobies. Un prélèvement liquide de 5 mL a été réalisé, afin de pouvoir caractériser l'état initial du système (appelé par la suite T0). Les incubations des microcosmes ont été effectuées à l'obscurité en conditions mésophiles, dans une pièce thermostatée à  $35 \pm 1$  °C, et sans agitation.

## 1.4. Suivi, prélèvement et conservation des échantillons pendant l'incubation

La production de biogaz a été suivie par des mesures régulières de pression à l'aide d'un manomètre différentiel (Digitron 2020P) équipé d'un tube relié à une aiguille ( $\emptyset$  de 0,45 mm, Terumo). Les analyses de composition du gaz ont été réalisées à l'aide d'un appareil de micro-chromatographie portable en phase gazeuse ( $\mu$ GC) (Varian CP4900), lorsque la pression différentielle interne des microcosmes dépassait 250 mbar. A la fin de l'analyse, la pression différentielle dans les microcosmes était à nouveau mesurée. Parfois,

une étape de dégazage était nécessaire lorsque cette pression était trop élevée (> 500 mbar) : le dégazage était alors réalisé à l'aide d'une seringue à gaz en verre (Terumo). En relevant systématiquement la température ambiante, la température d'incubation, la pression atmosphérique, et la pression différentielle avant analyse et à l'issue de l'ensemble des analyses ( $\mu$ GC, dégazage), la quantité de biogaz produite a été calculée et exprimée en conditions normales de pression et de température ( $T = 273 \text{ K}$ ,  $P = 1013,25 \text{ mbar}$ ).

Pendant l'incubation, différents prélèvements liquides et gazeux ont été effectués à différents points de temps pour pouvoir déterminer la composition isotopique du biogaz, ainsi que l'évolution de la communauté microbienne lors du processus de digestion anaérobie. Pour le prélèvement de gaz, 5 mL de biogaz ont été prélevés dans le ciel gazeux à l'aide d'une seringue en verre (Terumo) et conservés dans un vacutainer (BD Vacutainer Hermogard) et conservés à température ambiante jusqu'à l'analyse isotopique des gaz par GC-C-IRMS (Delta plus Isotope Ratio Mass Spectrometer, Thermo Electron Corporation, USA). Concernant le prélèvement de liquide, après avoir mélangé par agitation le contenu du microcosme, 5 mL de liquide ont été prélevés à l'aide d'une seringue, puis répartis dans trois microtubes de 2 mL (avec des volumes respectifs de 2 mL, 1,5 mL, et 1,5 mL). L'aliquot de 2 mL a été utilisé pour la mesure immédiate du pH. Les deux autres tubes ont été centrifugés à 8000 g pendant 20 minutes à 15°C. Les surnageants ont été ensuite transférés dans deux autres microtubes. Les culots et surnageants ainsi obtenus ont été conservés séparément à -80°C, et ont été utilisés respectivement pour les différentes analyses de biologie moléculaire et chimiques (Figure 32).

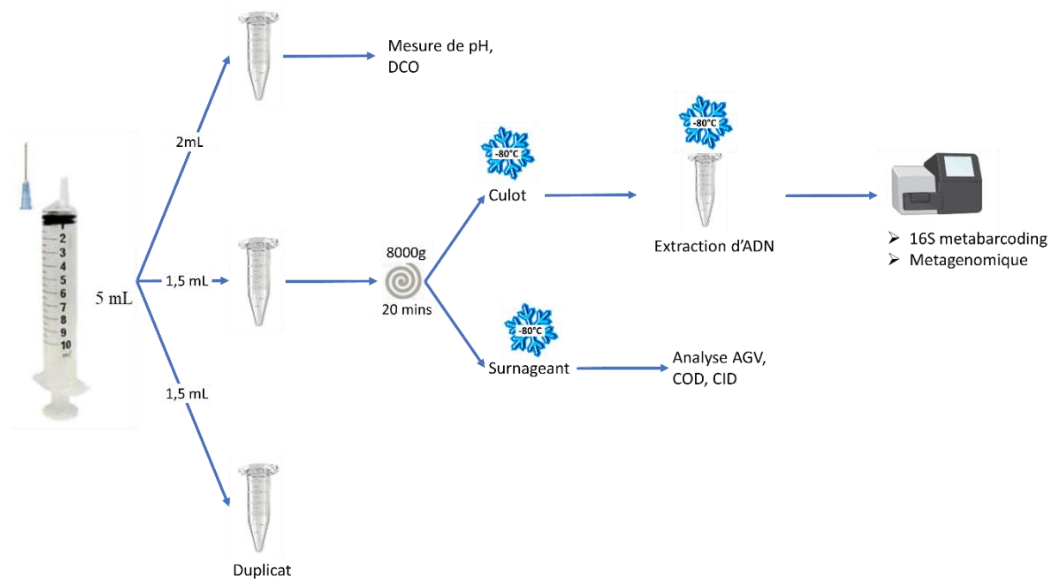


Figure 32. Illustration schématique des prélèvements d'échantillons liquides et de leurs analyses

A la fin de l'incubation, la totalité de la phase liquide des flacons a été centrifugée à 8000 g pendant 20 minutes à 15°C. Le surnageant (contenant les particules virales) a été collecté, puis conservé à 4°C pour la préparation des particules virales.

## 2. Préparation des particules virales issues des microcosmes de digestion anaérobie

Afin d'éliminer les cellules microbiennes encore présentes dans les surnageants contenant les particules virales, ceux-ci ont été soumis à une préfiltration sous vide en trois étapes successives, avec des filtres de diamètre de 45 mm en Polyethersulfone (PES), avec des tailles de pore de 5  $\mu\text{m}$ , 1,2  $\mu\text{m}$  et 0,8  $\mu\text{m}$  (Sartorius), placés dans des unités de filtration de 500 mL (Thermo Scientific) connectées à une pompe à vide (Fisher Scientific). La dernière étape de filtration a été appliquée avec des filtres en PES de taille de pore 0,2  $\mu\text{m}$ . Le filtrat obtenu a été centrifugé à 40 000 g pendant 3 h à 4°C, pour culoter les particules virales. Les culots ont été collectés, puis suspendus dans 2 mL de tampon SM (NaCl 0,1 M, MgSO<sub>4</sub> 0,1 M, Tris-HCl 0,05 M, pH 7,5), adapté à la conservation des particules virales. Les solutions virales ainsi obtenues ont été conservées à 4°C et utilisées ultérieurement pour des observations en microscopie électronique en transmission (MET) et pour l'extraction d'ADN viral.

### 3. Observation des particules virales

Toutes les observations des particules virales au microscope électronique en transmission (MET) réalisées pendant ces travaux ont été effectuées à la plateforme INRAE-MIMA2 (Microscopie et Imagerie des Micro-organismes, Animaux et Aliments), localisée sur le site de Jouy-en-Josas. Chaque solution contenant des particules virales préparées comme décrit dans la section précédente a été diluée 10 fois dans du tampon SM. Les particules qu'elle contenait ont été adsorbées sur une membrane fine de carbone sur une grille de cuivre de 300 mesh, puis colorées avec une solution d'1% d'acétate d'uranyle dilué dans de l'eau distillée. La grille a été séchée à température ambiante. Les grilles ont été examinées à l'aide d'un microscope électronique Hitachi HT7700 fonctionnant à 80kV (Elexience - France), et les images ont été acquises avec une caméra à dispositif à couplage de charge (AMT).

### 4. Méthodes d'analyses physico-chimiques

#### 4.1. pH

La valeur de pH a été mesurée directement dans 2 mL d'échantillon brut le jour du prélèvement de liquide, à l'aide d'un pH-mètre IQ160 (IQ scientific) de résolution de 0,01.

#### 4.2. Demande chimique en oxygène (DCO)

La mesure de DCO est une analyse couramment utilisée dans les études environnementales. Ce paramètre indique la quantité en oxygène nécessaire pour oxyder les composés organiques et minéraux présents dans l'échantillon. L'analyse a été effectuée en utilisant un kit commercial (LCK514, Hach), selon les instructions du fabricant, avec le spectrophotomètre (DR3900, Lange).

#### 4.3. Matière sèche/Matière volatile (MS/MV)

Ces analyses ont permis de caractériser l'inoculum utilisé pour les séries d'incubation en microcosmes. Les protocoles utilisés sont décrits dans les normes NF EN 14346 (pour les MS) et NF EN 15169 (pour les MV). Ces analyses ont été réalisées en triplicats pour pallier

à l'effet d'échantillonnage.

Environ 20 g d'inoculum ont été transférés dans un creuset préalablement pesé à 0,001 g de précision ( $m_a$ ). Ces creusets ont été à nouveau pesés ( $m_b$ ) pour déterminer la masse exacte d'inoculum ajouté. Ils ont été ensuite incubés dans un four (Memmert) à 105°C pendant au minimum 24 h, pour éliminer toutes les traces d'eau dans l'inoculum, puis à nouveau pesés ( $m_c$ ). Le résidu sec, ou matière sèche, exprimé en fraction massique en pourcentage, a été calculé avec l'équation :

$$MS(\%) = \frac{m_b - m_c}{m_b - m_a} * 100$$

Puis, ces creusets ont été ensuite placés dans un autre four (Nabertherm) et chauffés à 550°C pendant 20 minutes. Ces premiers ont été ensuite transférés dans un dessiccateur pour se refroidir avant d'être à nouveau pesés ( $m_d$ ). La matière volatile, exprimée en fraction massique en pourcentage, est calculée avec l'équation :

$$MV(\%) = \frac{m_c - m_d}{m_b - m_a} * 100$$

#### 4.4. Carbone organique/inorganique dissous (COD/CID)

Les concentrations en carbone organique et inorganique dissous ont été déterminées selon la norme NF EN 1484, en utilisant un analyseur TOC-L (Shimadzu). Le domaine de validité se situe entre 0,5 à 200 mg C/L pour le COD et entre 1 à 100 mg C/L pour le CID.

##### a. Analyse du COD

La technique utilisée était le NPOC (Non-Purgeable Organic Carbon). Après avoir acidifié (HCl 2N) l'échantillon à pH 2-3, on y faisait barboter un gaz vecteur pour éliminer tout le carbone inorganique. L'échantillon était ensuite introduit dans un tube de combustion contenant un catalyseur et chauffé à 680 °C. Le carbone contenu dans l'échantillon était converti en dioxyde de carbone et transporté par le gaz vecteur jusqu'au détecteur NDIR (Non-Dispersive Infra-Red), où il était analysé.

## b. Analyse du CID

L'analyse était réalisée directement dans la seringue. Après avoir acidifié l'échantillon à pH 2-3, on y faisait barboter un gaz vecteur afin que tout le carbone inorganique soit converti en dioxyde de carbone. Le dioxyde de carbone était ensuite transporté par le gaz vecteur jusqu'au détecteur NDIR (Non-Dispersive Infra-Red), où il était analysé.

### 4.5. Acides gras volatils (AGV) et autres acides organiques

Les concentrations en 7 AGV ou autres acides organiques distincts (acide lactique, acide formique, acide acétique, acide propionique, acide butyrique, acide valérique, acide caproïque) ont été mesurées par chromatographie ionique (ICS 5000+, Thermo Fisher Scientific).

La séparation s'effectuait grâce à une colonne spécifique d'exclusion stérique, une colonne IonPac ICE-AS1 (9 X 250 mm). Les acides faiblement ionisés étaient séparés en fonction de leur pKa, leur taille, et leur hydrophobicité. Après la séparation, les molécules étaient détectées par un détecteur conductimétrique.

La gamme de validité s'étendait de 5 à 500 mg/L.

### 4.6. Analyse de composition des gaz

La composition du biogaz a été analysée à l'aide d'un  $\mu$ GC Varian CP 4000, un système miniaturisé d'analyse de gaz, par chromatographie en phase gazeuse. Il contenait 4 colonnes couplées à un détecteur de thermoconductivité, permettant de séparer les gaz produits ( $O_2$ ,  $N_2$ ,  $CH_4$ ,  $CO_2$ ,  $N_2O$ ,  $H_2S$ ,  $H_2$ ,  $NH_3$ ) avec de l'Helium et de l'Argon, les gaz vecteurs, et aussi de quantifier ces différents gaz. Chaque analyse a été réalisée en deux répliques techniques, pour s'assurer de la bonne répétabilité de la mesure.

### 4.7. Analyse de la composition isotopique des gaz

Les analyses de composition isotopique du gaz ont été réalisées à l'aide d'un chromatographe en phase gazeuse (Trace GC Ultra, Thermo Fisher Scientific) couplé à un spectromètre de masse isotopique (Delta V PLUS, Thermo Fisher Scientific), via une



interface de combustion GC combustion III (Thermo Fisher Scientific) (Figure 33).

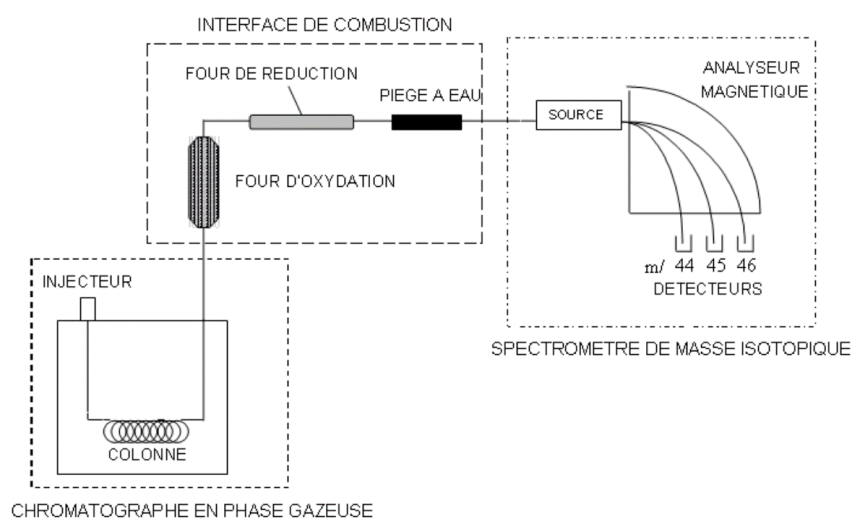


Figure 33. Système de couplage de chromatographie en phase gazeuse-combustion-spectrométrie de masse isotopique (Chapleur, 2012).

A l'aide d'un passeur automatique (Triplus AS, Thermo Electron Corporation, USA), du biogaz était prélevé et injecté en mode split avec un ratio de 20 et à une température de 150°C dans le chromatographe en phase gazeuse (GC). Le CH<sub>4</sub> et le CO<sub>2</sub> étaient séparés par une colonne (Rt-Q-BOND PLOT, L 30m, DI 0.32mm, EF 10µm).

La programmation pour le four était la suivante : 3 min à 35°C puis une montée en température de 60°C par minute jusqu'à 180°C, suivie d'un plateau de 24 secondes à 180°C, avec un flux d'hélium constant de 1,8 mL.min<sup>-1</sup> dans la colonne. Ensuite, les gaz passaient dans le four de combustion porté à 940°C, où ils étaient transformés en CO<sub>2</sub>. Le CO<sub>2</sub> était transporté jusqu'à la source du spectromètre grâce à l'hélium. Les ions générés dans la source par impact électronique étaient accélérés et focalisés sur le spectromètre de masse isotopique. Celui-ci comportait un secteur magnétique qui séparait les ions en fonction de leur ratio masse/charge, et détectait les masses 44, 45 et 46 pour le CO<sub>2</sub>, dans trois collecteurs (Figure 33). Chaque analyse était réalisée en triplicats techniques, pour s'assurer de la bonne répétabilité de la mesure.

Les analyses de COD, CID, AGV, et composition isotopique du gaz ont été réalisées par le pôle analytique de l'unité PROSE.

## 5. Méthodes de biologie moléculaire

### 5.1. Extraction d'ADN cellulaire

L'ADN cellulaire des échantillons a été extrait avec le kit PowerSoil en suivant le protocole du fabricant (Qiagen). Ce dernier comprend la lyse mécanique et chimique des cellules, une étape d'élimination des inhibiteurs de PCR (de type acides humiques), et une purification de l'ADN sur colonne. L'ADN extrait a été élué dans du tampon Tris HCl 10 mM à pH 7,5 fourni par le fabricant.

### 5.2. Extraction d'ADN viral

La procédure d'extraction d'ADN était basée sur la méthode Phenol-Chloroforme (Pickard, 2009). 360  $\mu$ L de solution contenant les virions ont été traités avec 2 U de Turbo DNase (2 U. $\mu$ l<sup>-1</sup>, Ambion) et 22  $\mu$ L de RNaseA (1 mg/mL, Ambion) pendant 1 h à 37°C, pour éliminer l'ADN extracellulaire contaminant. Le tampon d'inactivation du kit d'Ambion a été ajouté pour arrêter l'activité de la DNase, selon les recommandations du fabricant. La digestion de la capsid a été réalisée en ajoutant 10  $\mu$ L de SDS 20% et 4  $\mu$ L de protéinase K à 10 mg/mL (Ambion), puis en incubant à 37°C pendant 30 min. Après digestion, la solution obtenue a été transférée dans un tube Phase Lock Gel light, de 1,5 mL (VWR). Pour extraire l'ADN viral, 0,5 mL de phenol:chloroforme:alcool isoamylique (25:24:1) (Fisher) ont été ajoutés et une centrifugation à 1500 g pendant 5 min a été effectuée. Cette étape a été répétée une fois. Ensuite, 0,5 mL de chloroforme:IAA (24:1) (Fisher) ont été ajoutés, avant de centrifuger à 6 000 g à 15°C pendant 5 min.

L'ADN a été précipité à température ambiante pendant 20 minutes en ajoutant 45  $\mu$ L d'acétate de sodium 3 M (Thermo) et 500  $\mu$ L d'isopropanol à 100% (Thermo). Le culot d'ADN a été lavé deux fois avec de l'éthanol à 70% puis centrifugé à 14 000 g pendant 20 min à 15°C. Le culot a finalement été séché sous hotte et mis en suspension dans 100  $\mu$ L de tampon Tris-HCl 10 mM (Thermo).

### 5.3. Dosage d'ADN

La concentration des ADNs (cellulaires et viraux) a été déterminée avec le kit Qubit dsDNA HS assay (Invitrogen), en suivant les instructions du fabricant. Il s'agit d'une technologie fluorimétrique.

### 5.4. SIP-ADN

Les échantillons d'ADN ont été analysés par paires. Ils étaient issus d'un microcosme alimenté avec du formate  $^{13}\text{C}$ , d'une part, et d'un microcosme contenant du formate non marqué, d'autre part, systématiquement analysé de la même manière, comme contrôle.

Environ 20  $\mu\text{L}$  d'échantillon d'ADN avec une concentration de l'ordre de 20-30  $\text{ng}/\mu\text{L}$  ont été introduits dans le tube Beckman Quick-seal en Polyallomer de 2,2 mL, ainsi qu'une solution de CsCl dans du tampon TE, pour atteindre un volume total de 2 mL. La densité de la solution de CsCl était comprise entre 1,718 et 1,728, ce qui correspond à des indices de réfractométrie de 1,4010 et 1,4020, respectivement, mesurés à l'aide d'un réfractomètre d'Abbe, semi-automatique (ARIAS 500). La densité était calculée selon la formule suivante à partir de l'indice de réfraction (Chapleur, 2012):

$$\text{Densité} = 10,8601 * (\text{indice de réfraction}) - 13,4974.$$

Les tubes ont été ensuite scellés, leur masse équilibrée par paires à 0,001 g de précision. Ces tubes ont été introduits dans un rotor TLA-120.2 (Beckman), puis ultracentrifugés à 150 000g pendant 20 h à 20°C, à l'aide de l'ultracentrifugeuse Optima™ MAX (Beckman).

Après ultracentrifugation, des fractions de 100  $\mu\text{L}$  ont été collectées successivement par le fond des tubes, en pompant de l'eau en haut des tubes à l'aide du système Beckman Fraction Recovery system. Pour chaque fraction, la quantification de l'ADN a été effectuée avec le kit Qubit dsDNA HS assay (Invitrogen), et la densité des fractions a été calculée à partir de l'indice de réfractométrie, déterminé avec le réfractomètre Arias 500. En se basant sur la comparaison des profils de concentration en ADN en fonction de la densité, pour chaque paire d'échantillons, les fractions collectées issues de microcosmes avec formate marqué ( $^{13}\text{C}$ ) ont été ensuite combinées en trois groupes principaux, appelées fraction fr1

(contenant seulement de l'ADN non marqué), fraction fr2 (mélange d'ADN marqué au  $^{13}\text{C}$  et non marqué) et fraction fr3 (contenant uniquement de l'ADN marqué au  $^{13}\text{C}$ ).

Ces trois fractions principales ont été purifiées en utilisant la méthode de précipitation de l'ADN avec du glycogène et du PEG 6000 (Sigma-Aldrich). Pour 100  $\mu\text{L}$  de solution d'ADN, 1  $\mu\text{L}$  de glycogène (20  $\mu\text{g}/\mu\text{L}$ ) (Roche) et 200  $\mu\text{L}$  de la solution de 30% PEG 6000 et de NaCl 1,6M ont été ajoutés, puis incubés pendant la nuit à température ambiante. Ensuite, ce mélange a été centrifugé à 13 000 g pendant 30 minutes, à 15°C. Le culot contenant l'ADN a été lavé deux fois avec de l'éthanol 70°, puis suspendu dans 100  $\mu\text{L}$  de Tris HCl 10 mM pH 7,5 et enfin conservé à -20°C avant d'être amplifié ou séquencé.

#### 5.5. Amplification des ADNs préalable au séquençage « shotgun »

Les ADNs de certaines fractions ont été amplifiés pour obtenir des quantités suffisantes (minimum 100 ng) pour pouvoir réaliser le séquençage métagénomique shotgun. Pour cela, le kit RTG GenomiPhi™ V3 DNA Amplification (illustra) a été employé en suivant les instructions du fabricant. Les ADNs amplifiés ont été ensuite purifiés par précipitation avec du glycogène et du PEG 6000, selon la procédure décrite dans la section précédente.

#### 5.6. Métabarcoding basé sur le gène d'ARNr 16S

Des régions hyper variables archéennes et bactériennes du gène codant pour l'ARNr 16S ont été amplifiées puis séquencées selon le protocole décrit par Poirier et al. (2016) ou Madigou et al. (2019), avec quelques modifications. Les régions hypervariables V4-V5 ont ainsi été amplifiées par PCR avec les amorces de fusion 515F (5'- Ion A adapter-Barcode-GTGYCAGCMGCCGCGTA-3') et 928R (5'-Ion trP1 adapter-CCCCGYCAATTCMTTTRAGT-3'), qui comprenaient un code-barre et des adaptateurs de séquençage, selon les recommandations du manuel « Ion Amplicon Library Preparation (Fusion Method) ».

Le mélange PCR contenait du 1X SuperFi Buffer, 0,2 mM de chaque dNTP, 0,5  $\mu\text{M}$  de chaque amorce, 1 U d'enzyme Platinum SuperFi DNA Polymerase (Invitrogen) et 10 à 20 ng d'ADN matrice, dans un volume total de réaction de 50  $\mu\text{L}$ . L'amplification a été

réalisée avec les étapes suivantes : 30 s à 98°C, 30 cycles de 10 s à 98°C, 10 s à 50°C, 30 s à 72°C, et une extension finale de 5 min à 72°C.

Les produits de PCR ont été purifiés à l'aide de billes magnétiques d'immobilisation réversible en phase solide (SPRI) (billes magnétiques Mag-Bind TotalPure NGS, Omega Bio-Tek) conformément aux instructions du fabricant, avec un rapport billes/amplicon de 1,2. Ils ont été élués dans 45 µL de tampon TE (Tris-HCl 10 mM pH 8,0, EDTA 1 mM). Les amplicons purifiés ont été quantifiés par analyse des profils d'électrophorèse capillaire, obtenus à l'aide soit du kit D1000 ScreenTape avec l'appareil 2200 TapeStation, soit du kit DNA 1000 avec l'appareil 2100 Bioanalyzer (Agilent Technologies), en suivant les instructions du fabricant. Ensuite, tous les amplicons ont été pré-dilués à 500 pM dans de l'eau de qualité moléculaire et combinés en proportions équimolaire. Le mélange a ensuite été dilué à 100 pM pour le séquençage.

Afin de préparer des *Ion Sphere Particles* (ISP) positives à la matrice contenant de l'ADN obtenu par amplification clonale, par PCR en émulsion, la bibliothèque a été diluée à 26 pM et introduite dans l'instrument Ion OneTouch 2 (Life Technologies), en utilisant le kit Ion PGM Hi-Q View OT2 (Life Technologies) et en suivant les instructions du fabricant. Ces ISP ont ensuite été purifiées sur le Ion OneTouch ES (Life Technologies) conformément aux recommandations du fabricant.

Le séquençage a été effectué sur un séquenceur Ion Torrent Personal Genome Machine (PGM) en utilisant une puce Ion 316 Chip V2 (Life Technologies) et le kit de séquençage Ion PGM Hi-Q View (Life Technologies), conformément aux instructions du fabricant. Les données de séquençage ont été traitées par le logiciel Torrent Suite. Le logiciel a éliminé les lectures de séquences polyclonales et de mauvaise qualité, et les données filtrées ainsi obtenues ont été exportées sous la forme de fichiers FastQ.

### 5.7. Métagénomique *shotgun*

Les données de séquençage métagénomique *shotgun* présentées dans ce manuscrit ont été obtenues par deux prestataires différents, (I2BC et Eurofins) selon leurs propres protocoles. Le détail de chaque protocole sera précisé dans les chapitres suivants.

## 6. Analyses bioinformatiques

### 6.1. Traitement bioinformatique des séquences de metabarcoding d'ARNr 16S

Les analyses bioinformatiques des données de metabarcoding ont été effectuées avec l'outil FROGS (Escudié *et al.*, 2018), un outil convivial pour l'analyse des gros jeux de données produisant notamment des tables d'abondance des clusters de séquences, qui représentent des unités taxonomiques opérationnelles (*Operational Taxonomic Units*, OTUs), et des tables de leur affiliation taxonomique. Ces analyses ont été menées sur l'instance Galaxy du serveur d'INRAE MIGALE. Les différentes étapes sont illustrées schématiquement dans la Figure 34.

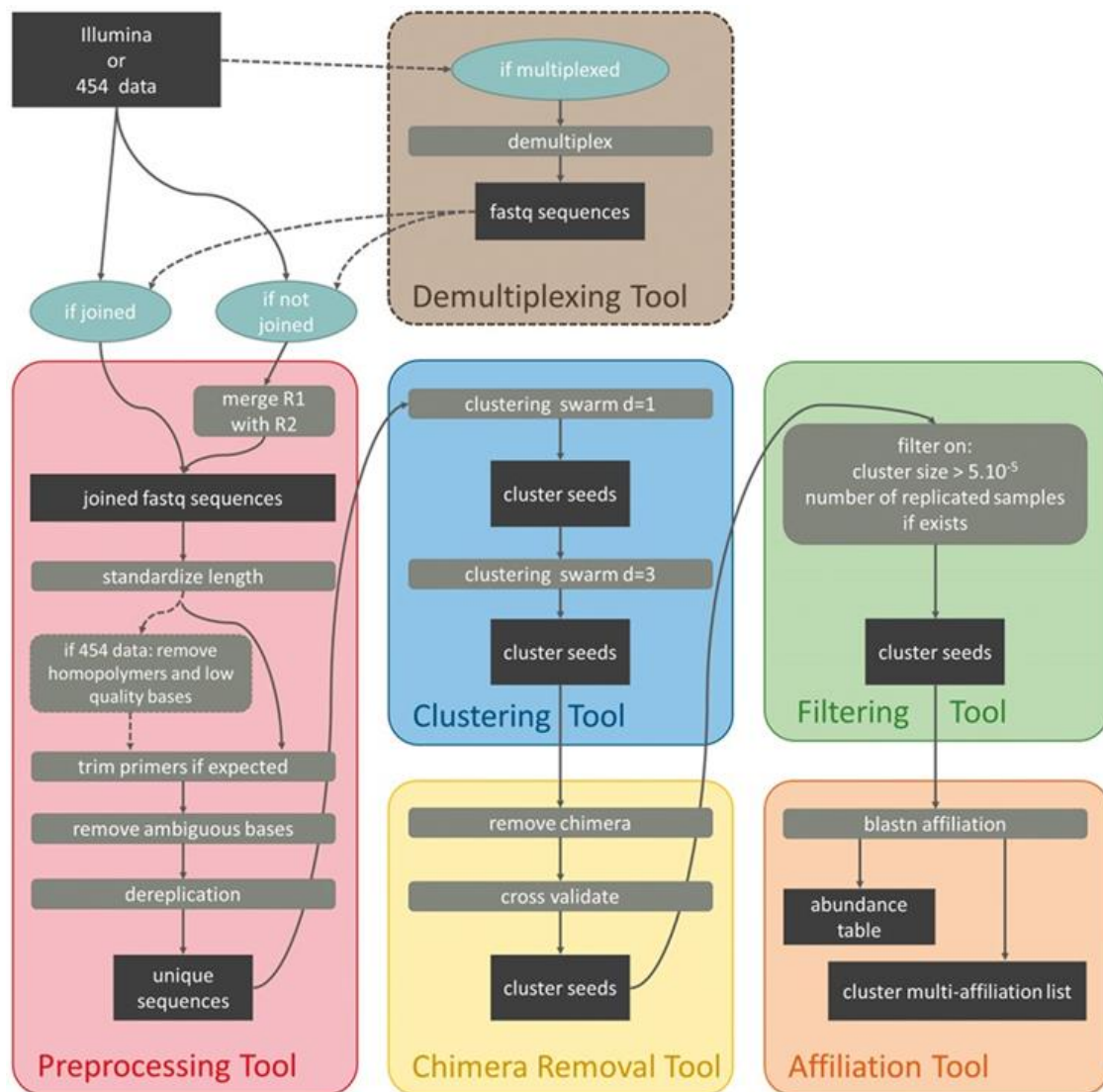


Figure 34. Pipeline bioinformatique de FROGS (Escudié *et al.*, 2018)

Dans le cadre du présent travail, les données FastQ avaient déjà été démultiplexées, c'est-à-dire que les différentes lectures avaient été attribuées à leur échantillon d'origine. Aussi, les fichiers FastQ ont été analysés directement à l'étape de pré-*processing* (rouge) qui permet de : i) éliminer les séquences de mauvaise qualité, ii) retirer les séquences des amorces, iii) ordonner et dénombrer les séquences strictement identiques. Ensuite, les séquences très similaires ont été regroupées en clusters en utilisant un algorithme itératif, *swarm* (bleu). Puis, l'étape de suppression des séquences chimériques, créées lors des cycles de PCR (jaune) a été appliquée pour éliminer les clusters chimériques. Elle a été suivie par une étape de filtration pour sélectionner les clusters pertinents sur la base de leur abondance (ratio  $\geq 10^{-5}$ ) et éliminer les contaminations (Phi X database). Enfin, la séquence représentative de chaque cluster a été comparée à une base de données de référence (silva138 16S pintail100), par alignement, afin de réaliser son assignation taxonomique.

FROGS permet également d'exporter des résultats sous différents formats (tableaux, arbres phylogénétiques, biom), ainsi que de fournir des statistiques sur les clusters et sur leur affiliation.

### 6.1. Pipeline bioinformatique d'analyses de métagénomique *shotgun*

Afin de favoriser la reproductivité des analyses de données, un pipeline bioinformatique a été développé sous la forme d'un *workflow snakemake* ([https://forgemia.inra.fr/cedric.midou/workflow\\_metagenomics](https://forgemia.inra.fr/cedric.midou/workflow_metagenomics)), développé sur le serveur INRAE-MIGALE (INRAE, Bioinformatics, MIGALE bioinformatics facility) par le bioinformaticien de l'UR PROSE, Cédric Midoux. Ce pipeline comprenait les étapes les plus génériques, appliquées aux métagénomomes cellulaire ainsi qu'aux métaviromes (Figure 35).

Les détails des versions et des paramètres de chaque outil utilisé pour chaque série d'expériences seront présentés dans le chapitre correspondant.

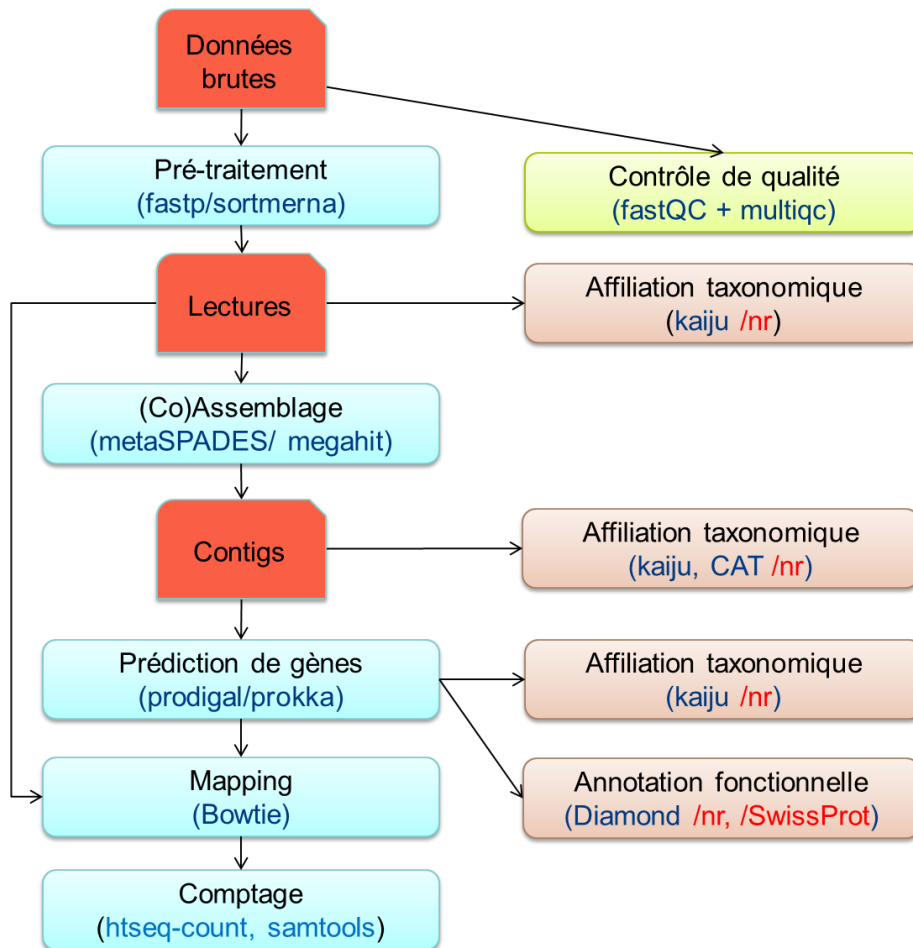


Figure 35. Etapes d'analyses bioinformatiques communes aux métagénomés cellulaires et métaviromes

Après des étapes de contrôle de qualité et de pré-traitement des données brutes, les lectures ont été assemblées individuellement ou co-assemblées selon les cas, avec metaSPADES (Nurk *et al.*, 2016) ou avec megahit (Li *et al.*, 2015) selon les séries d'expériences. Les régions codantes ont été prédites avec Prodigal (Hyatt *et al.*, 2010). L'assignation taxonomique des contigs a été obtenue avec Kaiju (Menzel *et al.*, 2016) ou CAT (von Meijenfeldt *et al.*, 2019a) et celle des gènes prédits avec Kaiju, contre la base de données NCBI nr dans tous les cas. Une première annotation des gènes prédits a été réalisée avec Diamond (Buchfink *et al.*, 2015) contre les bases de données NCBI nr ou Swiss-Prot (Bairoch and Apweiler, 2000). Pour chaque sous-jeu de données individuel, les lectures ont été alignées soit sur les gènes soit sur les contigs assemblés, à l'aide de Bowtie2 (Langmead and Salzberg, 2012) et comptées avec htseq-count (Anders *et al.*, 2015) et SAMtools (H. Li *et al.*, 2009a), respectivement.



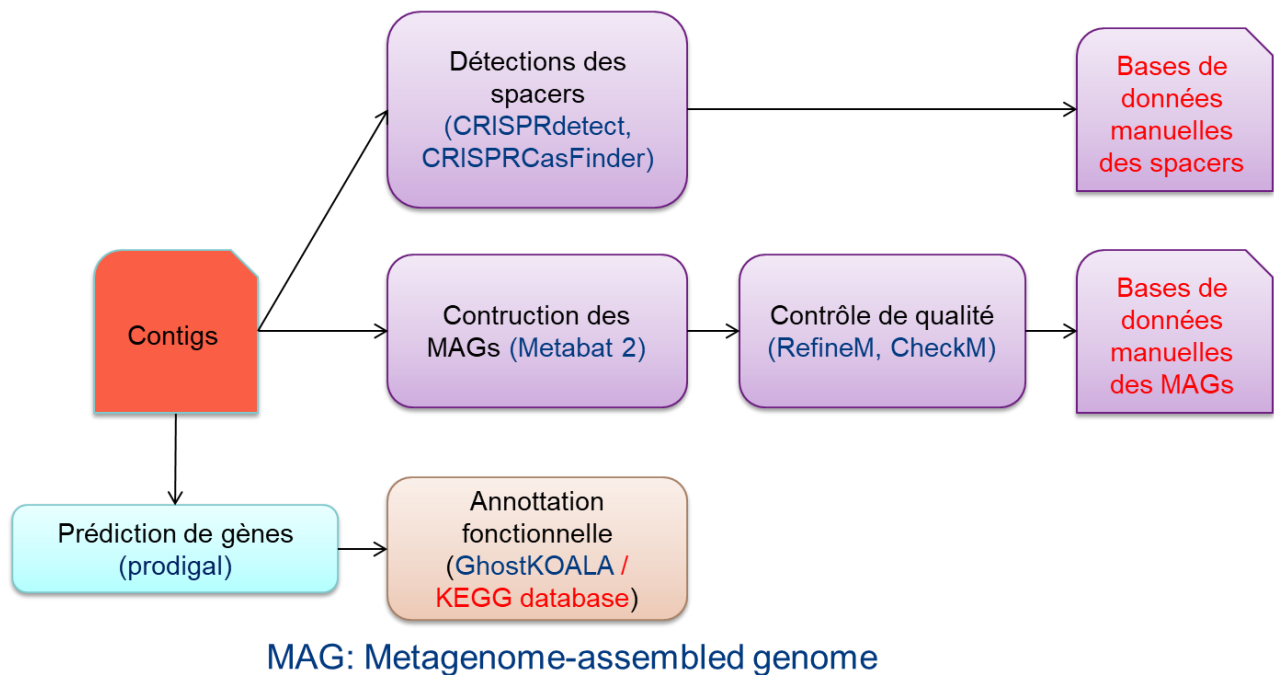


Figure 36. Etapes d'analyses bioinformatiques spécifiques aux métagénomes cellulaires

Différentes étapes spécifiques ont ensuite été effectuées en fonction du type de métagénome (cellulaire ou viral). Concernant les métagénomes cellulaires (Figure 36), une base de données non-redondante de *spacers* a été construite à partir de ceux détectés au sein des contigs avec les outils CRISPRCasFinder (Couvin *et al.*, 2018) et CRISPRdetect (Biswas *et al.*, 2016). Cette base de données a servi à la prédiction d'hôte pour les contigs viraux, grâce à des méthodes d'alignement avec les outils blastn et SpacePharer (R. Zhang *et al.*, 2021). De plus, une base de données de « *Metagenome assembled genomes* » (MAGs) a été également construite à partir des contigs cellulaires, à l'aide de Metabat2 (Kang *et al.*, 2019a). Cette base de données a été employée pour la prédiction d'hôte basée sur la composition en k-mer (signature), avec l'outil WISH (Galiez *et al.*, 2017). Par ailleurs, l'analyse des voies métaboliques impliquées dans le processus de digestion anaérobie a été effectuée avec l'outil GhostKOALA (Kanehisa *et al.*, 2016) contre la base de données KEGG, sur le serveur dédié (<https://www.kegg.jp/ghostkoala/>).

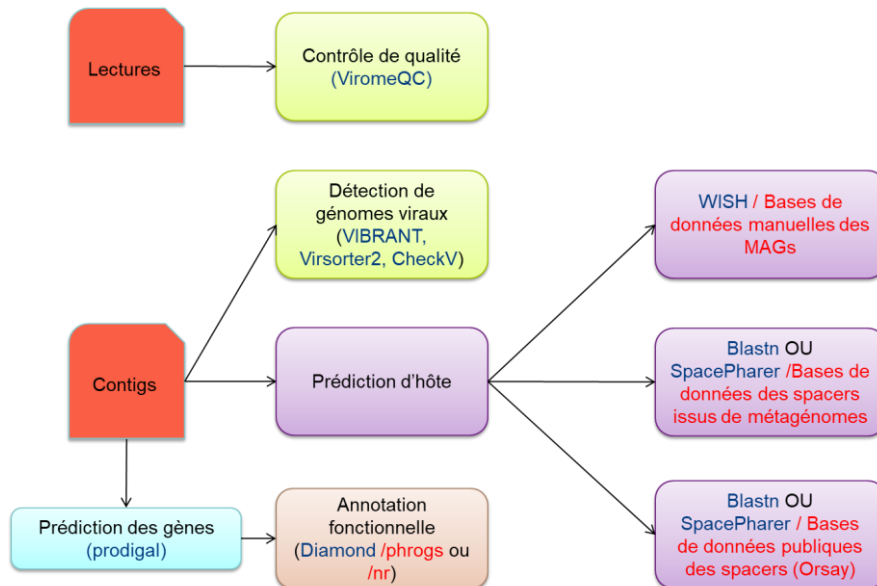


Figure 37. Etapes d'analyses bioinformatiques spécifiques aux métaviromes

Concernant les métaviromes, la qualité des lectures a été analysée avec ViromeQC (Zolfo *et al.*, 2019), qui permet d'estimer le taux de contamination des métaviromes par des séquences d'origine cellulaire (Figure 37). La détection des contigs d'origine virale a été effectuée avec VIBRANT (Kieft *et al.*, 2020a) et Virsorter2 (Guo *et al.*, 2021a), et la qualité des contigs viraux a été contrôlée avec CheckV (Nayfach *et al.*, 2021). Les gènes ont été annotés avec Diamond contre la base de données PHROGs (Terzian *et al.*, 2021), une base de données dédiée aux virus procaryotes et à leurs protéines.

Différents outils ont été utilisés pour prédire l'hôte d'origine des contigs viraux, tels que décrit dans la section précédente. Concernant la méthode basée sur l'alignement des spacers, en plus de la base de données construite à partir des métagénomes cellulaires, une base de données publique a été également employée (CRISPRCasdb spacer: [https://crisprcas.i2bc.paris-saclay.fr/Home/DownloadFile?filename=spacer\\_34.zip](https://crisprcas.i2bc.paris-saclay.fr/Home/DownloadFile?filename=spacer_34.zip)).

## 7. Analyses statistiques

Les résultats obtenus à partir des différentes analyses bioinformatiques ont été importés et analysés dans RStudio en utilisant différentes librairies avec la version la plus récente au moment de l'analyse, telles que Phyloseq, stringr, UpsetR, ou tidyverse. Les figures ont été produites en utilisant les librairies ggplot2 et igrph.

## 8. Système modèle sur cultures de souches pures

Afin de mettre en évidence le marquage indirect au  $^{13}\text{C}$  des ADN viraux *via* l'utilisation d'un substrat marqué au  $^{13}\text{C}$  pour enrichir l'hôte du virus, un modèle biologique constitué d'une souche pure d'*Escherichia coli* B (DSM 613) et de son bactériophage T4 (DSM 4505) a été utilisé. Le substrat choisi était le glucose. Ces expériences ont été réalisées dans le cadre du stage de Master 2 de Maximilien Sotomski.

### 8.1. Mise en culture d'*E. coli* et infection virale par le bactériophage T4

Tout d'abord, deux étapes de précultures des cellules d'*E. coli* dans du milieu LB (LB broth, Fisher) ont été effectuées pour éviter des effets aléatoires liés à la décongélation. Pour cela, 5  $\mu\text{L}$  du stock de cellules *E. coli* conservé à  $-80^\circ\text{C}$  ont été prélevés pour ensemercer 5 mL de milieu LB, ensuite incubés à  $37^\circ\text{C}$  sous agitation à 400 rpm pendant une journée. Puis, la seconde culture a été obtenue par l'incubation de 5  $\mu\text{L}$  prélevés dans la première préculture, ajoutés à 5 mL de milieu LB, pendant 8h et sous les mêmes conditions expérimentales.

Ensuite, 20  $\mu\text{L}$  de la seconde préculture d'*E. coli* à une concentration cellulaire  $\approx 10^9$  CFU (*colony-forming units* : unités formant colonie)/mL et 8  $\mu\text{L}$  de solution virale de T4 correspondant à un total d'environ 156 PFU (*plaque-forming units* : unité formant plaque de lyse) ont été ajoutés simultanément dans 20 mL de milieu minimal M9 composé de  $^{12}\text{C}$  ou  $^{13}\text{C}$ -glucose 1% (m/V) comme unique source de carbone, de sels minimaux M9 (M9, Minimal Salts, 5X, Sigma-Aldrich) pour assurer la croissance bactérienne, ainsi que de  $\text{MgSO}_4$  1 mM (Sigma-Aldrich) et  $\text{CaCl}_2$  0,3 mM (Sigma-Aldrich) pour favoriser l'infection virale. Afin d'évaluer la résolution de l'approche de SIP virale, des cultures avec différents taux de  $^{13}\text{C}$ -glucose (0%, 20%, 40%, 60%, 80%, 100%) ont été effectuées. Après incubation pendant 30 h à  $37^\circ\text{C}$ , sous agitation à 400 rpm, les cultures ont été centrifugées pendant 15 minutes à 5 000 g, à  $10^\circ\text{C}$ . Les surnageants ont été collectés pour la préparation des particules virales

Parallèlement, des cultures de cellules d'*E. coli* seules (sans bactériophage), dans ce même

milieu culture ont été réalisées avec des taux de  $^{13}\text{C}$ -glucose variant de 0% à 35% afin de vérifier l'efficacité du marquage au  $^{13}\text{C}$ . Après incubation, les cultures ont été centrifugées à 7 200 g pendant 20 minutes à 17°C, et les culots contenant les cellules ont été lavés avec du tampon phosphate saline PBS (phosphate buffered saline) (Sigma-Aldrich) puis suspendues dans 15  $\mu\text{L}$  de ce tampon pour l'analyse de leur composition isotopique par IRMS.

## 8.2. Préparation des particules virales

Les surnageants collectés issus d'expériences d'infection virale ont été filtrés avec des filtres à seringue de porosité 0,22  $\mu\text{m}$  (filtres PES, 0,22  $\mu\text{m}$ , FisherBrand) afin d'éliminer les cellules résiduelles et les gros débris cellulaires. Les solutions de particules virales ainsi obtenues ont été stockées à 4°C en attente de titration ou d'extraction de l'ADN.

## 8.3. Titration sur tapis bactérien

La titration des solutions ou des surnageants de culture contenant le bactériophage T4 a été effectuée en boîte de Petri par dépôt de gouttelettes sur tapis de cellules d'*E. coli*. Pour cela, une couche de gélose molle (« *Top agar* ») a été déposée sur une boîte de Petri carrée de LB agar de composition classique agar (Sigma-Aldrich) :

15g/L). Ce « *Top agar* » était constitué de 6 mL de LB+agar (agar : 7,5g/L), et de 600  $\mu\text{L}$  d'une préculture d'*E.coli* obtenue comme indiqué précédemment par culture dans du milieu LB. Des gouttes de 5  $\mu\text{L}$  issues de dilutions sérielles (d'un facteur 10 ou 15) de la solution de bactériophage T4 à titrer ont été déposées sur la gélose molle solidifiée, en triplicats, à l'aide d'une micropipette-multicanaux. Après une brève phase de séchage des boîtes sous la hotte à flux laminaire, une incubation à 37°C pendant 24 h en conditions statiques a été réalisée. Les plages de lyse ont alors été comptées. Le titre initial en PFU/mL a été déterminé en tenant compte des facteurs de dilution appliqués.

## 8.4. Extraction et dosage d'ADN

L'extraction d'ADN d'*E coli* a été effectuée avec le kit « DNeasy UltraClean Microbial Kit » (QIAGEN) selon les instructions du fabricant.

Concernant l'extraction d'ADN du phage T4, les particules virales obtenues ont été d'abord concentrées soit en utilisant les tubes Amicon Ultra-15 (centrifugal Filter units, Meck Millipore) selon les instructions du fabricant, soit par centrifugation à 20 000 g pendant 4 h à 4°C, suivie par la reprise du culot dans 1 mL de milieu M9 initial. Ensuite, ces solutions virales ont été incubées avec 10 µl de DNase (DNase I THERMO 1Unité/µL) à température ambiante pendant 20 minutes, pour éliminer les ADN contaminants présents dans la solution. L'activité de DNase a ensuite été inactivée par incubation à 75°C pendant 5 minutes. L'ADN viral a été extrait avec le kit « Phage DNA Isolation Kit » (Norgen) en suivant les instructions du fabricant, avec des modifications mineures.

Le dosage d'ADN a été réalisé par le kit Qubit dsDNA HS assay (Invitrogen) (Voir section 5.3).

#### 8.1. SIP ADN

Voir section 5.4

#### 8.2. Analyses IRMS

Environ 20 µL de la suspension de cellules d'*E. coli* dans le tampon PBS, tel que décrit dans la section 8.1, ont été placés dans des capsules en étain (8 mm x H et 5 mm x L de Thermo FisherScientific) pesées au préalable. Les capsules ont été par la suite conservées à 55°C jusqu'au lendemain, afin de sécher les culots cellulaires. Le lendemain matin, les capsules ont été pesées afin de déterminer la masse de cellules puis gardées au dessiccateur en attendant leur analyse par IRMS, permettant de déterminer le ratio  $^{12}\text{C}/^{13}\text{C}$  du carbone total des échantillons solides. Ces analyses ont été réalisées avec le spectromètre décrit dans la section 4.7, avec adaptation pour l'étape de combustion des échantillons.

# Chapitre 1 : Focus sur les virus d'archées méthanogènes

## 1. Introduction

Afin d'appliquer notre stratégie de couplage du SIP avec les analyses métagénomiques, d'un substrat, il était tout d'abord nécessaire d'identifier un substrat qui favorise le processus de digestion anaérobie tout en activant des voies métaboliques spécifiques, permettant ainsi de sélectionner des communautés microbiennes moins complexes. Trois substrats, non-marqués à ce stade, ont été évalués : l'acétate, le formate et le méthanol. Ces trois substrats sont des composés à faible nombre de carbones (C1 et C2), ils peuvent donc enrichir la communauté microbienne active principalement lors des dernières étapes de la DA : l'acétogenèse et méthanogenèse.

En outre, il était également souhaitable d'ajuster certains paramètres opératoires (voir section suivante), pour notre système expérimental. Une série d'expériences test d'incubation en microcosmes a donc été menée.

Suite à cette première série d'expériences, le formate a été choisi pour l'étude de la diversité virale, en se concentrant sur les virus des archées méthanogènes. Les résultats de cette série d'expériences test sont présentés succinctement ci-dessous.

## 1. Expériences test pour guider le choix d'un substrat d'intérêt

### 1.1. Plan expérimental

En plus de la nature du substrat, deux paramètres opératoires ont également été évalués. Il s'agissait de la quantité d'inoculum utilisée et de l'ajout de particules virales collectées pendant la préparation de l'inoculum. Concernant le premier paramètre, une trop grande quantité d'inoculum pouvait aboutir à la dilution du  $^{13}\text{C}$  et limiter la croissance de la biomasse, tandis qu'une trop faible quantité pouvait conduire à des biais d'échantillonnage. Le deuxième paramètre a été évalué afin d'éviter la perte possible de diversité virale dans l'inoculum. Le plan expérimental complet est détaillé dans le tableau 10.

Tableau 10. Plan expérimental des incubations de test de différents paramètres opératoires

	Témoin négatif	Témoin négatif (+virus)	Effet du substrat			Effet d'ajout de virus		Effet de la quantité d'inoculum	
			Acétate	Formate	Méthanol	Acétate	Formate	Acétate	Formate
<b>Substrat</b>	-	-	Acétate	Formate	Méthanol	Acétate	Formate	Acétate	Formate
<b>Quantité d'inoculum ajouté (g/100 mL de phase liquide)</b>	0,5	0,5	0,5	0,5	0,5	0,5	0,5	1	1
<b>Concentration du substrat (en mM de carbone)</b>	0	0	200	200	200	200	200	200	200
<b>Ajout de virus</b>	-	+	-	-	-	+	+	-	-
<b>Répliquas</b>	<b>1</b>	<b>1</b>	<b>3</b>	<b>3</b>	<b>3</b>	<b>3</b>	<b>3</b>	<b>3</b>	<b>3</b>
<b>Volume du microcosme (mL)</b>	250								
<b>Volume de la phase liquide (mL)</b>	100								

## 1.2. Résultats

1

### a. Test des paramètres opératoires

2

La production cumulée de biogaz a été mesurée pour évaluer l'effet des différents paramètres testés (Figure 38). Concernant le fait d'ajouter des particules virales, aucun impact significatif n'a été observé. Quant au test sur la quantité d'inoculum ajouté, la production de biogaz des microcosmes avec une quantité d'inoculum plus élevée (1g par 100 mL de phase liquide) débutait légèrement plus tôt et avec une meilleure reproductibilité, mais les quantités totales de biogaz produites dans les deux cas étaient similaires.

3

4

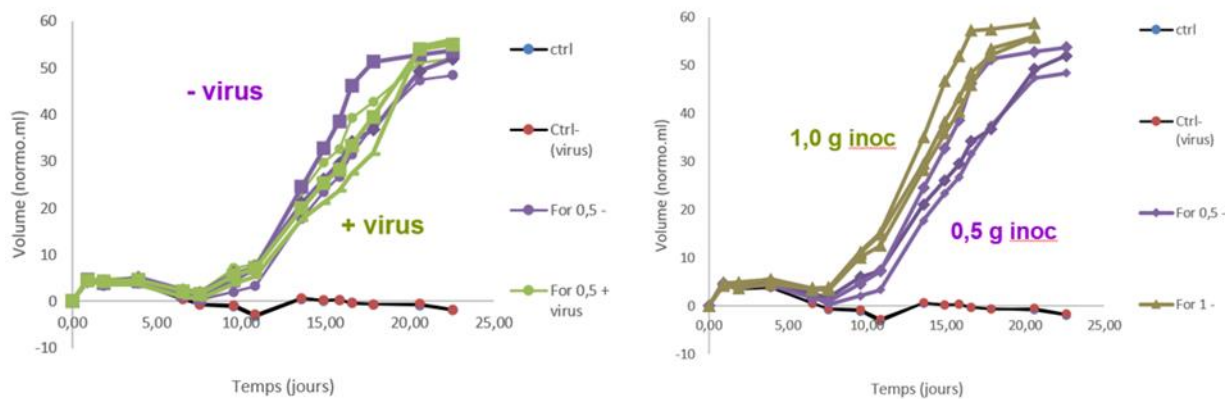
5

6

7

8

9



10

Figure 38. Productions cumulées de biogaz lors des tests de différents paramètres opératoires

11

### b. Test des substrats

12

Nous avons observé que malgré de plus grandes quantités de biogaz cumulé produites, les microcosmes contenant du méthanol comme source unique de carbone présentait une grande variabilité dans la dynamique de production du biogaz (Figure 39). Parmi les trois substrats testés, le formate était associé à une meilleure reproductibilité de la production du biogaz, ainsi qu'à une cinétique de production plus brève.

13

14

15

16

17

18



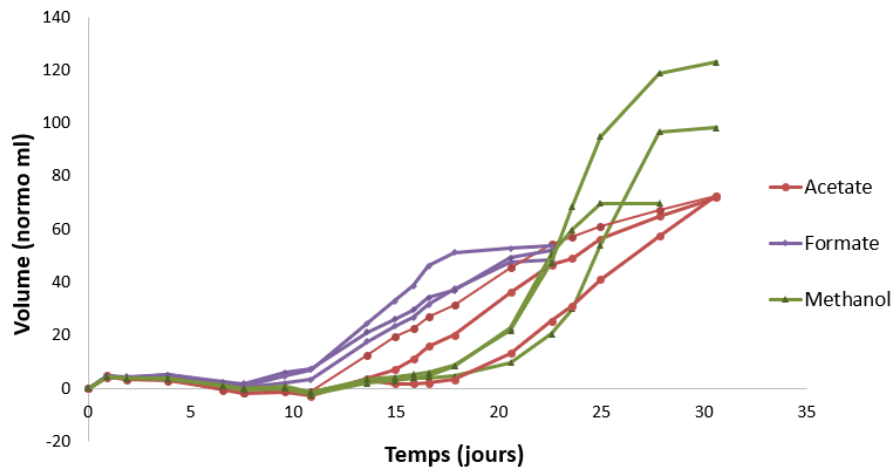


Figure 39. Productions cumulées de biogaz obtenues lors du test de différents substrats

19

20

Les analyses de metabarcoding d'ADNr 16S ont montré une augmentation de la proportion d'archées dans tous les microcosmes, à la fin de l'incubation (Figure 40). Cela montre que la communauté microbienne s'est enrichie en archées. Les proportions d'archées les plus élevées ont été observées dans les microcosmes contenant du formate. De façon intéressante, l'utilisation de différents substrats a par ailleurs activé différentes populations d'archées méthanogènes. Ainsi, tandis que l'acétate et le méthanol ont abouti à la sélection d'archées de l'ordre Methanosarcinales, l'emploi de formate a abouti à la sélection d'archées de l'ordre Methanobacteriales.

21

22

23

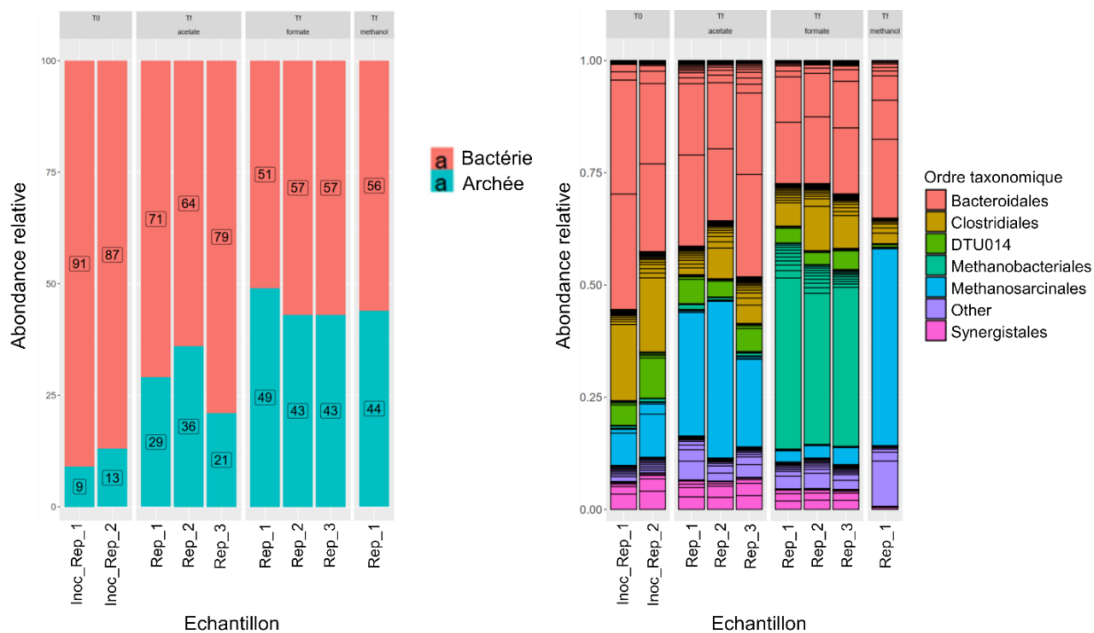
24

25

26

27

28



29

Figure 40. Composition des communautés microbiennes lors des incubations de test de différents substrats (metabarcoding d'ADNr 16S)

30

31

1.3. Conclusions	32
Sur la base des résultats présentés ci-dessus, cette série d'expériences test nous a permis de définir les conditions expérimentales à utiliser pour appliquer l'approche de SIP.	33 34
Concernant au substrat utilisé pour nos séries d'expériences, le formate a été choisi en raison de la bonne reproductibilité de production de biogaz obtenu pendant l'incubation avec ce substrat, ainsi que de son pouvoir d'activer certains groupes d'archées méthanogènes spécifiques.	35 36 37 38
Concernant aux paramètres opératoires, la quantité d'inoculum utilisé sera fixée de 1g par 100 mL de phase liquide et l'ajout des particules virales ne sera pas pris en compte.	39 40
 2. Présentation du travail décrit dans ce chapitre	 41
Dans ce chapitre, nous présentons notre première mise en œuvre du couplage entre SIP et analyses métagénomiques, pour identifier des virus infectant des archées méthanogènes impliquées dans la bioconversion du formate. Ce dernier était utilisé comme source unique de carbone dans les microcosmes de digestion anaérobie en batch. Ces microcosmes étaient dominés par des archées méthanogènes appartenant à l'ordre Methanobacterales (genres <i>Methanobacterium</i> et <i>Methanobrevibacter</i> ). Les analyses métagénomiques ont abouti à l'identification plusieurs nouveaux génomes viraux, y compris un génome complet d'un virus tête-queue (classe <i>Caudoviricetes</i> , hôte <i>Methanobacterium</i> ) et des génomes partiels de virus fusiformes dont le morphotype est spécifique aux virus d'archées. De plus, les résultats de métagénomique sont cohérents avec les observations de microscopie électronique en transmission, qui ont révélé la prédominance de particules virales de type tête-queue et fusiformes. Cette étude a permis d'élargir nos connaissances sur la diversité des virus d'archées méthanogènes. Dans ce chapitre, elle est présentée sous la forme d'un article qui a été publié en Juillet 2022 dans <i>Environmental microbiology</i> .	42 43 44 45 46 47 48 49 50 51 52 53 54 55 56
	57

**Diversity of novel archaeal viruses infecting methanogens discovered through coupling of stable isotope probing and metagenomics**

Hoang Ngo<sup>1</sup>, François Enault<sup>2</sup>, Cédric Midoux<sup>1,3,4</sup>, Mahendra Mariadassou<sup>3,4</sup>, Olivier Chapleur<sup>1</sup>, Laurent Mazéas<sup>1</sup>, Valentin Loux<sup>3,4</sup>, Théodore Bouchez<sup>1</sup>, Mart Krupovic<sup>5</sup>, Ariane Bize<sup>1\*</sup>

<sup>1</sup>Université Paris-Saclay, INRAE, PRocédés biOtechnologiques au Service de l'Environnement, 92761, Antony, France

<sup>2</sup>Université Clermont Auvergne, CNRS, LMGE, F-63000 Clermont-Ferrand, France

<sup>3</sup>Université Paris-Saclay, INRAE, MalAGE, 78350, Jouy-en-Josas, France

<sup>4</sup>Université Paris-Saclay, INRAE, BioinfOmics, MIGALE bioinformatics facility, 78350, Jouy-en-Josas, France

<sup>5</sup>Institut Pasteur, Université de Paris, CNRS UMR6047, Archaeal Virology Unit, F-75015 Paris, France

\* Correspondence: [ariane.bize@inrae.fr](mailto:ariane.bize@inrae.fr); Tel.: +33(0)1 40 96 60 89

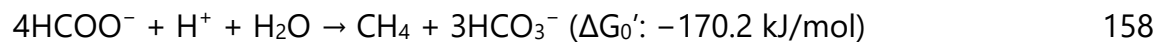
**Summary:** Diversity of viruses infecting non-extremophilic archaea has been grossly understudied. This is particularly the case for viruses infecting methanogenic archaea, key players in the global carbon biogeochemical cycle. Only a dozen of methanogenic archaeal viruses have been isolated so far. In the present study, we implemented an original coupling between stable isotope probing and complementary shotgun metagenomic analyses to identify viruses of methanogens involved in the bioconversion of formate, which was used as the sole carbon source in batch anaerobic digestion microcosms. Under our experimental conditions, the microcosms were dominated by methanogens belonging to the order Methanobacteriales (Methanobacterium and Methanobrevibacter genera). Metagenomic analyses yielded several previously uncharacterized viral genomes, including a complete genome of a head-tailed virus (class Caudoviricetes, Methanobacterium host) and several near-complete genomes of spindle-shaped viruses. The two groups of viruses are predicted to infect methanogens of the Methanobacterium and Methanosarcina genera and represent two new virus families. The metagenomics results are in good agreement with the electron microscopy observations, which revealed the dominance of head-tailed virus-like particles and the presence of spindle-shaped particles. The present study significantly expands the knowledge on the viral diversity of viruses of methanogens.

<b>Introduction</b>	97
Viruses infecting archaea represent one of the most unique parts of the global virosphere (Krupovic et al., 2018). Despite the limited number of archaeal viruses described so far compared to bacterial viruses, archaeal viruses show a great diversity of gene content and morphological properties. In particular, several morphotypes are specific to archaeal viruses, showing no similarity to viruses infecting bacteria and eukaryotes, such as bottle-shaped ( <i>Ampullaviridae</i> ), coil-shaped ( <i>Spiraviridae</i> ), or spindle-shaped ( <i>Bicaudaviridae</i> , <i>Fuselloviridae</i> , <i>Halspiviridae</i> , <i>Thaspiviridae</i> ) ones. Archaeal viruses are currently classified into 33 families (Baquero et al., 2020; Liu et al., 2021), including cosmopolitan icosahedral viruses and archaea-specific viruses.	98 99 100 101 102 103 104 105 106
Methanogenic archaea play a major role in carbon cycling at the global scale, through methanogenesis. Known methanogenic archaea are currently grouped into 8 different orders in the phyla Euryarchaeota and Halobacterota, and a few candidatus taxa in the Euryarchaeota, Halobacterota, and in the TACK group (Lyu et al., 2018; Evans et al., 2019). These were isolated from very diverse natural ecosystems such as wetlands, termite, human and livestock digestive tracts, rice fields, and deep-sea hydrothermal vents, and also from anaerobic digester (AD) (Lyu et al., 2018). Indeed, their unique metabolic features are exploited in AD processes (Ahring, 2003) for valorization of organic waste and effluents into methane-rich biogas, a renewable energy source. Among archaeal viruses, 10 have been reported to infect methanogenic archaea (methanogens) (Wood et al., 1989; Meile et al., 1989; Nölling et al., 1993; Wolf et al., 2019; Thiroux et al., 2021; Weidenbach et al., 2021; Molnár et al., 2020; Pfister et al., 1998; Weidenbach et al., 2017; Krupovič, Gribaldo, et al., 2010b). In addition, several proviruses integrated in the genomes of diverse methanogens have been described (Krupovič and Bamford, 2008; Krupovič, Forterre, et al., 2010b). Almost all of the viruses of methanogens described so far have been isolated from anaerobic digester samples (Table S1). A total of 5 head-tailed viruses or proviruses originate from thermophilic anaerobic digesters, all infecting Methanobacteriales hosts: $\Psi$ M1 and $\Psi$ M2 (family Leisingerviridae; siphovirus morphology) (Meile et al., 1989; Pfister	107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124

et al., 1998; Liu et al., 2021) and related defective provirus  $\Psi$ M100 (Luo et al., 2001) infect 125  
Methanothermobacter strains, whereas  $\Phi$ F1 (unclassified) and  $\Phi$ F3 (unclassified; 126  
siphovirus morphology) (Nölling et al., 1993) infect Methanobacterium species. Moreover, 127  
4 viruses or virus-like particles have been isolated from mesophilic anaerobic digesters: 128  
Methanobacterium-infecting virus Drs3 (family Anaerodiviridae; siphovirus morphology) 129  
(Wolf et al., 2019; Liu et al., 2021), Blf4 (unclassified; siphovirus morphology), infecting 130  
Methanoculleus strains (Methanomicrobiales) (Weidenbach et al., 2021), MetSV 131  
(unclassified) (Weidenbach et al., 2017), a spherical virus infecting Methanosarcina strains 132  
(Methanosarcinales), and finally the oblate or spindle-shaped Methanococcus 133  
(Methanococcales)-infecting A3 virus-like particles (VLPs) (Wood et al., 1989). MFTV1, an 134  
unclassified temperate head-tailed virus (siphovirus morphology), has been induced from 135  
a Methanocaldococcus fervens AG86 strain (Methanococcales) isolated from deep-sea 136  
hydrothermal vents (Thiroux et al., 2021); it is the first characterized virus infecting 137  
hyperthermophilic methanogens. In addition, MetMV (unclassified) (Molnár et al., 2020), 138  
has been suggested to infect mesophilic Methanosarcina strains, but this still needs to be 139  
confirmed. For the other four known orders of methanogens (Evans et al., 2019), no viruses 140  
have been isolated so far. Thus, the diversity of viruses infecting methanogenic archaea 141  
remains largely unexplored. 142

Metagenomics can provide a less biased view on the diversity of viruses infecting 143  
methanogens, by circumventing the challenge of cultivating some of the methanogenic 144  
archaeal strains, as well as biases associated with virus isolation. In such context, AD 145  
ecosystems prove to be particularly well suited as they are relatively easy to access to, and 146  
to establish in laboratory reactors. Moreover, they encompass methanogenic archaea from 147  
several orders, such as Methanobacteriales, Methanomicrobiales and Methanosarcinales 148  
(Lin et al., 2016; Evans et al., 2019). Yet, identifying viruses infecting methanogens is 149  
challenging in AD metagenomes due to the complexity of the catalytic microbial 150  
communities, dominated by diverse bacteria, and due to usually low proportions of 151  
methanogenic archaea in AD processes. 152

In the present study, an original experimental approach was applied with the aim of favoring the enrichment of methanogens in AD microbial communities to help the discovery of their viruses in metagenomes. To this end, AD microcosms were fed with <sup>13</sup>C-labeled formate, one of the known substrates for methanogenesis through the following equation (Sun et al., 2021):



In addition to favoring methanogens, such experimental approach has the advantage of preserving the AD process and a certain level of microbial diversity. Another benefit was the possibility to identify viruses targeting microorganisms that actively assimilated the substrate. Indeed, within complex microbial communities, not all of the microorganisms are active or involved in the degradation of a specific substrate. Thus, identifying active microorganisms, and their associated viruses, are key issues in linking viral diversity with functional aspects. Hence, we coupled stable isotope probing (SIP) (Radajewski et al., 2000), applied to the total cellular DNA, to in silico host prediction of the viral contigs. Host prediction was applied by using the cellular metagenomes obtained from the heavy (<sup>13</sup>C-enriched) and light (not enriched in <sup>13</sup>C) DNA sequences, which were used to build host databases. Whenever the predicted host was identified in the heavy cellular DNA fraction, we considered it as evidence that the virus infected active microorganisms, assimilating the <sup>13</sup>C-formate.

This original coupling (SIP and bioinformatic analyses for host prediction) led to the discovery of several previously uncharacterized genomes of viruses of methanogens. In particular, they included two contigs likely representing new families of spindle-shaped viruses, thereby expanding the knowledge on the diversity of viruses infecting methanogens. One of these families was associated with the active hydrogenotrophic methanogenic archaea, while the other one seemed to target methanogens that were not assimilating the formate in the studied microcosms, which could only be evidenced thanks to the SIP analysis.

<b>Experimental Procedures</b>	181
<b><i>AD microcosm experiments and monitoring procedure</i></b>	182
AD batch microcosms consisted in glass plasma bottles with a total volume of 500 mL.	183
Either <sup>13</sup> C-labeled-formate or unlabeled formate was employed as the sole carbon source.	184
A mineral solution suitable for AD (standard NF EN ISO 11734) was also added, as well as	185
an inoculum prepared from a lab-scale digester fed with biowaste. To obtain the inoculum,	186
the sampled anaerobic sludge was incubated in anaerobic conditions at 35 °C for 20 days,	187
for the majority of fermentable compounds to decompose. It was then centrifuged	188
(10000 g, 20 min, 15 °C), aliquoted and stored at -80 °C until further use. The microcosms	189
were hermetically sealed with an aluminum screw cap and a rubber septum. The	190
headspaces were flushed with nitrogen gas (Linde). For each treatment (13C-labeled or	191
unlabeled formate), microcosms were prepared in triplicates with a working volume of	192
300 mL, using the mineral solution, 3 g of inoculum, and either 0.2 M <sup>13</sup> C-sodium formate	193
(Cortecnet) or 0.2 M unlabeled sodium formate (Sigma Aldrich) (equivalent to 4.08 g).	194
During incubation, the following physico-chemical parameters were monitored according	195
to the protocols described in (Puig-Castellví et al., 2022): biogas production, biogas	196
composition, pH, isotopic composition of the biogas, volatile fatty acids (VFAs), total	197
inorganic carbon (TIC), total organic carbon (TOC), and chemical oxygen demand (COD).	198
To characterize the dynamics of microbial communities (archaea, bacteria and in particular	199
their viruses), a pair of microcosms (one fed with <sup>13</sup> C-formate, one with unlabeled formate)	200
was sacrificed at each of three different time points (days 8, 13 and 17), for cellular and	201
viral DNA extraction. For each sacrificed microcosm, the liquid phase was centrifuged	202
(8000 g, 20 min, 15 °C). The pellet was stored at -80 °C for subsequent cellular DNA	203
extraction, and the supernatant (containing virus-like particles) was collected and stored	204
at 4 °C, for virion preparation.	205
<b><i>Preparation and observation of VLPs</i></b>	206
The supernatant resulting from the previous step was centrifuged once more (6000 g,	207



20 min, 4 °C) to further remove cells. The newly obtained supernatant was subjected to a three-step filtration with 1.2 µm and 0.8 µm polyethersulfone filters (Sartorius), and finally with 0.2 µm acetate cellulose filters. The filtrations were realized in 500 mL filtration units (Thermo Scientific) under vacuum. The final filtrate was centrifuged (40 000 g, 3 h, 4 °C) to pellet virus-like particles. The final pellet was suspended in 2 mL SM buffer (0.1 M NaCl, 0.1 M MgSO<sub>4</sub>, 0.05 M Tris-HCl, pH 7.5) and stored at 4 °C until subsequent analysis.

Virus-like particles were observed by transmission electron microscopy (TEM) at MIMA2 MET - GABI, INRAE, AgroParisTech (78352 Jouy-en-Josas, France). Virion-containing solutions were 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 – France), and images were acquired with a charge-coupled device camera (AMT).

### ***DNA extraction and quantification***

Total cellular DNA was extracted with Qiagen DNeasy PowerSoil Kit, according to the manufacturer's instructions.

For virus-like particles, 360 µL of virion solution were treated with 2 U of Turbo DNase (2 U µl<sup>-1</sup>, Ambion), and 22 µl of RNaseA (1 mg/mL, Ambion) for 1 h at 37 °C, to remove contaminant non-encapsidated DNA. Inactivation buffer from the Ambion kit was added to stop the DNase activity. DNA extraction was then based on a classical Phenol-Chloroform method (Pickard, 2009), using 1.5 mL Phase lock gel, light (VWR). The detailed procedure for viral DNA extraction is provided in the Supplementary information (Section 1.1).

Concentrations of viral DNA were determined with Qubit dsDNA HS assay (Invitrogen), according to the manufacturer's instructions.

### ***DNA ultracentrifugation***

Ultracentrifugation on a CsCl gradient was applied to the total cellular DNAs to separate

them according to their density, as described in (Chapleur et al., 2016). Around 1 µg of DNA was employed for each sample. For each collected fraction, the DNA quantification was performed with Qubit HS dsDNA. Based on DNA density profiles (Figure 1D), the fractions were pooled into 3 main groups (referred to as "fraction pools" thereafter) and DNA was purified by Glycogen-PEG 6000 precipitation, according to (Neufeld, Vohra, et al., 2007).

The DNA of each fraction pool was amplified with RTG GenomiPhi™ V3 DNA Amplification Kit (illustra), according to the manufacturer's instructions. It was then purified by Glycogen-PEG 6000 precipitation, and the concentration of the collected DNA was determined with Qubit dsDNA HS assay (Invitrogen).

### ***16S rRNA gene metabarcoding***

Archaeal and bacterial hypervariable regions V4-V5 of the 16S rRNA gene were amplified and then sequenced according to the protocol described in (Poirier, Desmond-Le Quéméner, et al., 2016; Madigou et al., 2019), with some modifications. The targeted region was amplified by PCR with fusion primers 515F (5'- Ion A adapter-Barcode-GTGYCAGCMGCCGCGGTA-3') (Wang et al., 2007) and 928R (5'-Ion trP1 adapter-CCCCGYCAATTCMTTTRAGT-3') (Wang and Qian, 2009). The detailed protocol for library preparation is provided in the Supplementary Information (section 1.2). Sequencing was performed on an Ion Torrent Personal Genome Machine using Ion 316 Chip V2 (Life Technologies) and Ion PGM Hi-Q View Sequencing Kit (Life Technologies) according to the manufacturer's instructions. Sequencing data were processed with the Torrent Suite Software. All diversity analyses were performed with the R phyloseq package.

### ***Shotgun metagenomic sequencing***

Illumina sequencing was performed at the I2BC sequencing platform (University of Paris-Saclay, Gif-sur-Yvette, France). 250-500 ng of genomic DNA were fragmented (400 bp mean size) on a Covaris S220 sonicator. DNA fragments were end-repaired and dA-tailed (NEB#E7595), Illumina TruSeq adapters were ligated (NEB#E6040), and the PCR-free library fragments were purified using AMPure XP beads (Beckman Coulter). Final library quality

was assessed on an Agilent Bioanalyzer 2100, using an Agilent High Sensitivity DNA Kit. 263  
Libraries were pooled in equimolar proportions and sequenced using paired-end 2x150 264  
pb runs, on an Illumina NextSeq500 instrument, using NextSeq 500 High Output 300 cycles 265  
kit. Demultiplexing was performed with bcl2fastq2 v2.18.12. Adapters were trimmed with 266  
Cutadapt v1.15, and only reads longer than 10 b were kept for further analysis. 267

### **Metagenomic pipe-line**

 268

The most generic steps of our pipeline were scripted as a snakemake (Köster and 269  
Rahmann, 2012) workflow 270  
([https://forgemia.inra.fr/cedric.midoux/workflow\\_metagenomics/-/tree/v21.04](https://forgemia.inra.fr/cedric.midoux/workflow_metagenomics/-/tree/v21.04)), and 271  
applied to both cellular and viral metagenomes. After a pre-processing step (adapter 272  
removal, and trimming according to quality scores and length with fastp), reads were 273  
assembled with metaSPADES (Nurk *et al.*, 2016) (individual assembly by sample for cellular 274  
metagenomes and coassembly for metaviromes). Coding regions were predicted with 275  
Prodigal (Hyatt *et al.*, 2010). Taxonomic affiliations of contigs and of their predicted genes 276  
were respectively obtained with CAT (von Meijenfeldt *et al.*, 2019b) and kaiju (Menzel *et* 277  
*al.*, 2016), against NCBI nr database. Genes were annotated using comparison to NCBI nr 278  
database with Diamond (Buchfink *et al.*, 2015). For each dataset, the cleaned reads were 279  
mapped to the assembled contigs using Bowtie2 (Langmead and Salzberg, 2012) and 280  
counted with samtools (H. Li *et al.*, 2009a). Details of versions and parameters for the 281  
snakemake workflow are available in the GitLab repository. 282

Several steps specifically dedicated to viral contig analysis and to host prediction were 283  
performed using homemade bash and python scripts. Metagenome-assembled genomes 284  
(MAGs) were constructed from assembled cellular metagenomic contigs with Metabat2 285  
(Kang *et al.*, 2019b). MAG quality was improved with RefineM (Parks *et al.*, 2017) and 286  
controlled with CheckM (Parks *et al.*, 2015). Only MAGs with more than 60% completeness 287  
and less than 5% contamination were selected. Functional annotations of the cellular 288  
predicted genes were obtained with ghostKoala against KEGG database (Kanehisa *et al.*, 289  
2016). Concerning the viral metagenomes, their quality was assessed with ViromeQC 290

(Zolfo *et al.*, 2019) based on the trimmed reads. Viral genome detection was performed 291  
with VIBRANT (Kieft *et al.*, 2020b) and VirSorter2 (Guo *et al.*, 2021b), and the quality of viral 292  
contigs was assessed with CheckV (Nayfach *et al.*, 2021). The predicted genes were further 293  
analyzed by using HH-suite (Steinegger *et al.*, 2019) against PHROGs (Terzian *et al.*, 2021), 294  
a database dedicated to prokaryotic viruses. 295

To predict potential host for each viral contigs, two different methods were used. First, 296  
CRISPR spacers were detected in cellular metagenome contigs with CRISPRdetect (Biswas 297  
*et al.*, 2016) and CRISPRCasFinder (Couvin *et al.*, 2018). A non-redundant spacer database 298  
was built from the obtained spacer sequences. The viral contigs were subsequently aligned 299  
with BLASTn and SpacePHARER (R. Zhang *et al.*, 2021) against this homemade database 300  
and a public spacer database (CRISPRCasdb spacer: [https://crisprcas.i2bc.paris- 301](https://crisprcas.i2bc.paris-saclay.fr/Home/DownloadFile?filename=spacer_34.zip)  
[saclay.fr/Home/DownloadFile?filename=spacer\\_34.zip](https://crisprcas.i2bc.paris-saclay.fr/Home/DownloadFile?filename=spacer_34.zip)), enabling host prediction based 302  
on an alignment-dependent method. Host were also predicted using the WIsH software 303  
(Galiez *et al.*, 2017) that compared the genomic signatures of viral contigs to the one of 304  
cellular MAGs. 305

After selecting viral contigs of interest as described in the result section, gene annotation 306  
was refined on HHpred server (Zimmermann *et al.*, 2018) with default parameters, against 307  
four structural/domain databases: Pfam-A\_v35, PDB\_mmCIF70\_12\_Ot\_2021, 308  
NCBI\_Conserved\_Domains(CD)\_v3.18 and UniProt-SwissProt-viral70\_3\_Nov\_2021. 309

The detailed parameters and versions for the python and bash scripts presented in 310  
Supplementary Tables S2 and S3. 311

312

Bipartite network of viral contigs and protein orthologous groups 313

A bipartite network was built, where the two node classes were viral contigs or genomes 314  
on the one hand, and protein orthologous groups (OGs) on the other. Regarding “viral 315  
nodes”, most archaeal (pro)virus genomes available in public databases (RefSeqVirus and 316  
nr, 7<sup>th</sup> of October, 2021) or described in previous studies (Krupovič, Gribaldo, *et al.*, 2010a; 317  
Philosof *et al.*, 2017) were included, in addition to the viral contigs of interest. The final 318  
dataset consisted in 172 viral sequences, in total. All the proteins encoded in these 319  
sequences were categorized into OGs by a two-step procedure as described in (Olo Ndela 320  
*et al.*, 2021) and in the Supplementary Information (section 1.2.2). 321

These OGs were functionally annotated by comparing their HMM profiles to those of the 322  
PHROGs database (Terzian *et al.*, 2021) (annotation release v2), which contains well- 323  
annotated protein clusters of prokaryotic viruses. In addition, protein sequences were 324  
compared to RefSeqVirus using MMseqs (bit score  $\geq 50$ ), retaining only the best hit for 325  
each protein of an OG. These OGs were used as “protein nodes” to construct the bipartite 326  
networks. The shared OGs and their functional prediction are listed in the Supplementary 327  
Table S4. All networks were computed using the igraph package of R (Csardi and Nepusz, 328  
2005). 329

The data for this study have been deposited in the European Nucleotide Archive (ENA) at 330  
EMBL-EBI under accession number PRJEB46489 331  
(<https://www.ebi.ac.uk/ena/browser/view/PRJEB46489>) (16S metabarcoding raw reads, 332  
shotgun metagenomics raw reads, annotated sequences of contigs C158, C889, C1359, 333  
C1697). 334

## **Results and Discussion** 335

### ***A significant proportion of formate converted to methane through hydrogenotrophic methanogenesis*** 336 337

A total of 6 batch AD microcosms were established and monitored over time during a 338  
maximum of 17 days. They contained, as sole carbon source, either unlabeled formate (3 339

replicates) or  $^{13}\text{C}$ -formate (3 replicates). At each of days 8, 13 and 17, one pair of 340  
microcosms was sacrificed for metavirome analysis, one with unlabeled formate and one 341  
with  $^{13}\text{C}$ -formate. 342

Chemical analyses (VFA, TIC, TOC) showed that both unlabeled and  $^{13}\text{C}$ -labeled formate 343  
were consumed, and more than 85% of the initial quantity was metabolized after 17 days 344  
of incubation (Supplementary Figure S1). The bioconversion resulted in a pH increase from 345  
 $\sim 7.50$  to  $\sim 9.20$  on average, consistent with the methanogenesis equation mentioned 346  
above. Biogas was produced after a lag phase of about 5 days (Supplementary Figure S1), 347  
to reach final cumulated productions of  $\sim 100$  normo-ml. Furthermore, reproducible 348  
dynamics were observed in microcosms, irrespective of the formate type. Methane ( $\text{CH}_4$ ) 349  
was by far the dominant biogas component, followed by carbon dioxide ( $\text{CO}_2$ ) as well as 350  
traces of hydrogen ( $\text{H}_2$ ) and hydrogen sulfide ( $\text{H}_2\text{S}$ ). Collectively, these results support the 351  
successful establishment of the AD process in the microcosms. More precisely, the isotopic 352  
composition of the biogas demonstrated the dominance of  $^{13}\text{C}$  in the methane ( $>95\%$ ) 353  
produced by the methanogens in the microcosms fed with  $^{13}\text{C}$ -formate, and indicated the 354  
dominance of the hydrogenotrophic methanogenesis pathway (Whiticar et al., 1986) in 355  
the microcosms fed with  $^{12}\text{C}$ -formate. For unlabeled substrates, this method relies on the 356  
abundance of the stable isotope  $^{13}\text{C}$  in nature (1.1%) and on the difference in reaction rate 357  
between  $^{12}\text{C}$ - and  $^{13}\text{C}$ -containing substrate molecules. The results and detailed calculation 358  
of isotopic signatures are provided in the Supplementary Information (section 2.1). 359

### ***Enrichment in methanogenic archaea, especially Methanobacterium, over time*** 360

To identify the methanogenic archaea and more broadly the microbial community 361  
composition, we applied 16S rRNA gene metabarcoding (Figure 1A). The relative 362  
abundance of the dominant archaea (Methanobacteriaceae, in green) increased from 1- 363  
3% at day 0 to 26-30% at day 17. Consistently, a functional analysis based on the cellular 364  
shotgun metagenomic data (Supplementary Figure S2) showed the dominance of 365  
methane metabolism, in relative abundance, at day 17. Among the microbial groups 366  
involved in methane metabolism (Figure 1C), archaea were the main actors 367

(Methanobacteria in cyan and Methanomicrobia in blue). Similar proportions of archaea 368  
 were observed in the metagenomic dataset (Supplementary Figure S3). These results 369  
 confirmed that the microbial communities were enriched in methanogenic archaea during 370  
 the incubation. 371

The detected archaea were mostly methanogens (Figure 1B). At day 0, *Methanosarcina* 372  
 (Methanosarcinales order) was dominant. However, at the end of the incubation, at day 373  
 17, the genus *Methanobacterium* became dominant, followed by *Methanobrevibacter* 374  
 (both from the order Methanobacteriales). It can be assumed that Methanobacteriales 375  
 members were selected during the incubation since they are able to grow on formate 376  
 through the hydrogenotrophic pathway, and since they likely outcompeted 377  
 Methanosarcinales methanogens, due to their faster growth rate: the doubling time is 378  
 generally lower than 1h in Methanobacteriales, compared to more than 10h in 379  
 Methanosarcinales (Thauer et al., 2008). 380

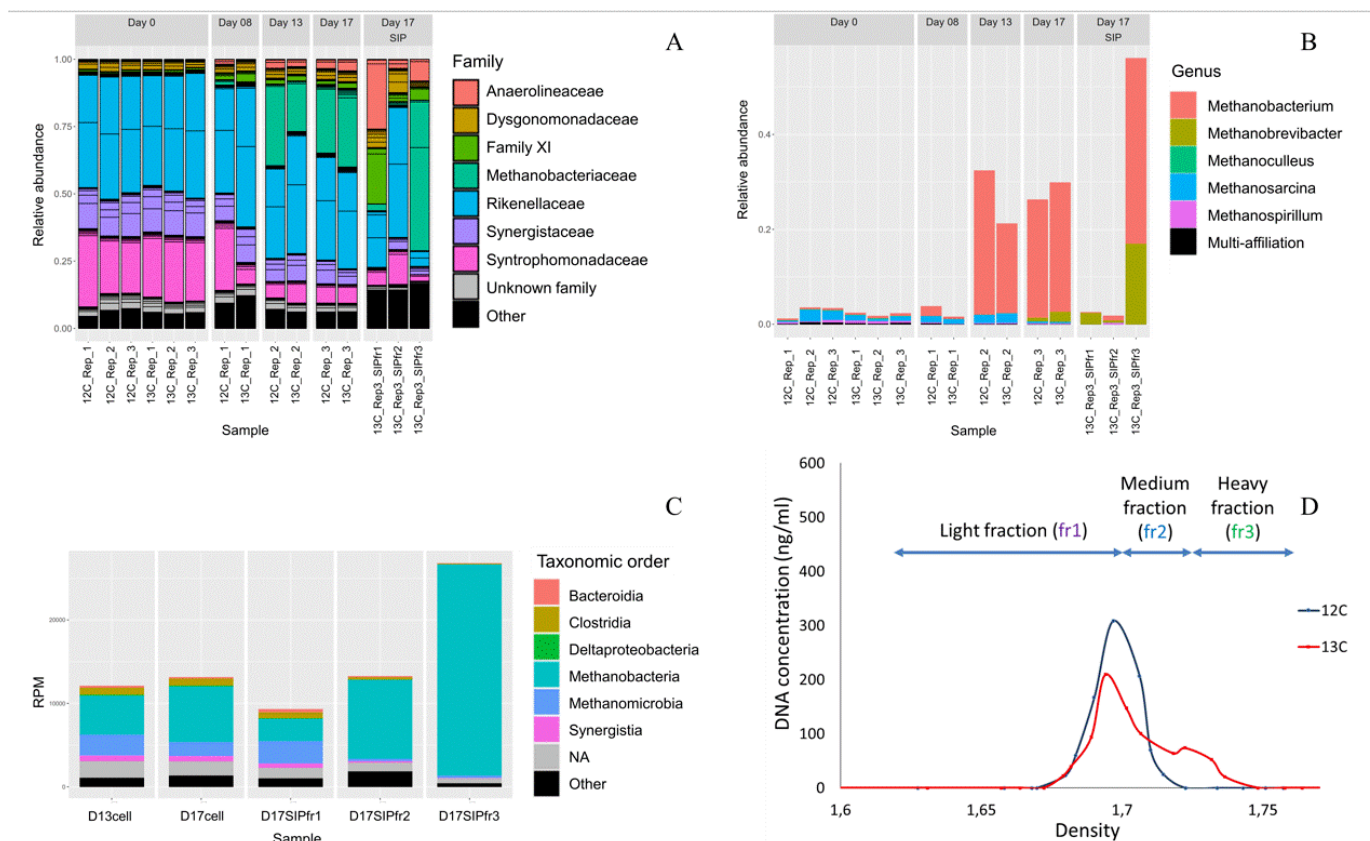


Figure 1. Microbial dynamics during AD incubation. A: Microbial community composition of archaea and 382  
 bacteria at the family level, based on 16S rRNA gene metabarcoding. B: Composition of archaea only, at 383

the genus level, based on 16S rRNA gene metabarcoding. C: Taxonomic assignment (Order level) of the 384  
microorganisms involved in methane metabolism, based on the functional analysis of the cellular shotgun 385  
metagenomes. D: Density profiles of the cellular DNAs obtained from microcosms fed with unlabeled for- 386  
mate (blue) or with  $^{13}\text{C}$ -formate (red) at day 17. The arrows labeled as fr1, fr2 and fr3 indicate how the 387  
DNA fractions were pooled in the case of labeled formate. fr1 contained only unlabeled DNA whereas fr3 388  
contained only  $^{13}\text{C}$ -labeled DNA. fr2 likely contained a mix of  $^{13}\text{C}$ -labeled and unlabeled DNA. 389

Concerning bacterial families present in these systems, a notable decrease in relative 390  
abundance was observed over time for *Synergistaceae* (phylum Synergistetes, from ~15% 391  
down to ~7%) and *Syntrophomonadaceae* (phylum Firmicutes, from ~24% down to ~7%) 392  
(Figure 1A). *Synergistaceae* members generally consume amino acids to generate short- 393  
chain fatty acid (He *et al.*, 2018), whereas *Syntrophomonadaceae* members are acetogens 394  
(Si *et al.*, 2016). For both families, the decrease is understandable due to the lack of 395  
adequate substrate (proteins and short-chain fatty acids). A notable increase in relative 396  
abundance was observed at day 17 for members of *Anaerolineaceae* (phylum Chloroflexi, 397  
from >1% up to ~4%), which are generally described as fermentatives or acetogens (Liang 398  
*et al.*, 2015; Si *et al.*, 2016), and some of them were previously shown to form syntrophic 399  
associations with methanogens (Lei *et al.*, 2018). In the present experiment, 400  
*Anaerolineaceae* bacteria may degrade formate and/or play a role in electron transfer 401  
(Wang *et al.*, 2021), in partnership with hydrogenotrophic methanogens. 402

For the six microcosms, the profiles of cellular DNA concentrations in function of their 403  
mass density were established (Supplementary Figure S4). For microcosms fed with  $^{13}\text{C}$ - 404  
formate at day 17 (Figure 1D, red line), a second peak, of denser DNA, was visible in the 405  
profile, indicating that the labeled substrate was assimilated by some of the 406  
microorganisms. The separation of DNA in CsCl gradient depends not only on the mass of 407  
the isotope, but also on the GC content of the DNA (Eason and Campbell, 1978). Hence, 408  
we pooled the DNA into 3 distinct fractions, according to their density. The first fraction 409  
(fr1, density < 1.705) corresponded exclusively to non-labeled DNA. The second fraction 410  
(fr2, 1.705 < density < 1.738) possibly contained a mix of unlabeled and  $^{13}\text{C}$ -labeled DNA. 411  
Finally, the third fraction (fr3, density > 1.738) contained exclusively  $^{13}\text{C}$ -labeled DNA. The 412



16S rRNA gene metabarcoding applied to these fractions showed that archaea reached 413 their highest proportions (56%) in fraction fr3 at day 17 (Figure 1B), whereas their 414 proportions in fractions fr1 and fr2 were lower than 3%. Consistently, the highest 415 abundance of genes involved in methane metabolism was in fraction fr3 at day 17, based 416 on metabolic pathway analysis from the shotgun metagenomics data (Figure 1C). These 417 observations confirmed that the <sup>13</sup>C-formate incorporated into the microbial biomass was 418 consumed mostly by methanogenic archaea of the genera *Methanobacterium* and 419 *Methanobrevibacter*. 420

**Presence of icosahedral and spindle-shaped VLPs, highlighted by TEM observations** 421

TEM observations revealed a great diversity of virus-like particles in the microcosms 422 (Figure 2), especially after 13 and 17 days of incubation (Figure 2A-C). Besides 423 cosmopolitan morphotypes common to the domains Bacteria and Archaea, uncommon 424 and especially archaea-specific morphotypes were also observed. 425

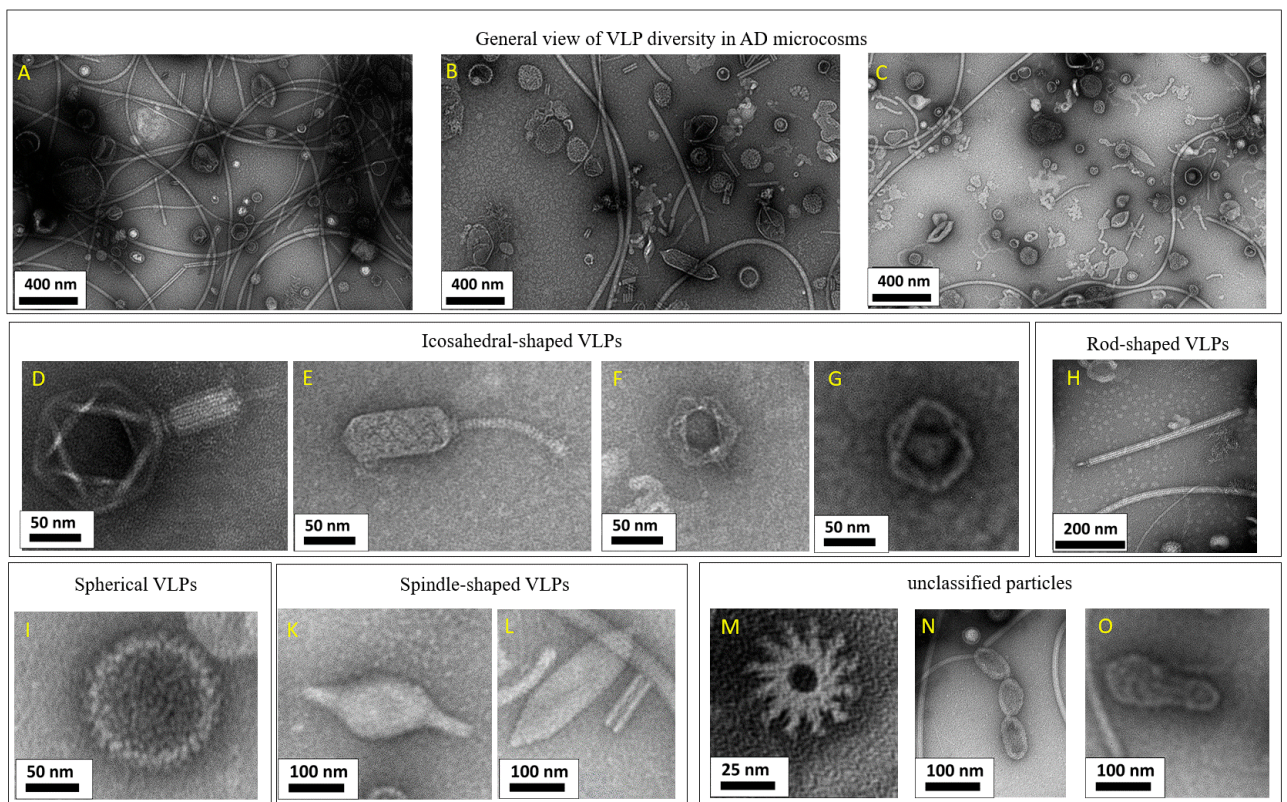


Figure 2. Morphotypic diversity of VLPs in different microcosms, observed by TEM. One representative 427 sample is shown for each incubation time point (A: day 8, B: day 13, C: day 17). 428

Cosmopolitan morphotypes included VLPs with a head-tailed morphology, typical of the class *Caudoviricetes*, as well as icosahedral tailless particles (Figure 2G). The latter could originate from tailless icosahedral viruses or from head-tailed viruses with a broken tail. In both samples, these VLPs were the most abundant and presented a large diversity: myovirus-like (*i.e.*, icosahedral capsids with contractile tails; Figure 2D), siphovirus-like (*i.e.*, icosahedral capsids with long non-contractile tails; Figure 2E) and also podovirus-like (*i.e.*, icosahedral capsids with short tails; Figure 2F) morphotypes, were observed, with capsid diameters ranging from 50 to 200 nm.

Less common viral morphotypes were detected in the microcosms at days 13 and 17, such as rod-shaped (Figure 2H), spherical (Figure 2I) and spindle-shaped (Figure 2K-L). Spindle-shaped morphotypes, which are specific to archaeal viruses, have been commonly observed in extreme geothermal and hypersaline environments (Krupovic *et al.*, 2014), but have also been reported in moderate ones, such as freshwater and marine habitats (Borrel *et al.*, 2012; Kim *et al.*, 2019). Interestingly, the presence of spindle-shaped virus-like particles has also been reported in AD plants (Calusinska *et al.*, 2016), suggesting that the hosts of these viruses could possibly be involved in the AD process.

Particles with unique morphotypes were also identified (Figure 2M-O). Some of them had a chain structure with multiple of two or three monomers and their size ranged from 100 to 400 nm. Moreover, other particles appeared as round-shaped and less than 50 nm, with a hollow star-like structure (Figure 2M). The nature of these particles is unclear.

### **Overview of the shotgun metagenomics datasets and of the host prediction approaches**

Shotgun metagenomic sequencing was applied to 7 selected DNA extracts. Four of the extracts were from the total cellular DNA and total viral DNA from samples collected at days 13 and 17 (D13cell, D13vir, D17cell, D17vir), selected due to their enrichment in methanogens observed in the metabarcoding analysis (Figure 1) and the presence of interesting VLP morphotypes (Figure 2). The three cellular DNA fractions obtained after density gradient centrifugation at day 17 were also sequenced (D17SIPfr1, D17SIPfr2,

D17SIPfr3) (Figure 1D). The purpose was primarily to analyze the metaviromes, and to use 457  
the cellular metagenomes for building specific databases for host prediction. 458

Based on classical metrics (Table 1), the sequencing data and the assemblies were of 459  
satisfactory quality. In the case of metaviromes, ViromeQC analysis indicated moderate 460  
contamination rates of 7.6% and 8.8% for D13vir and D17vir, respectively. Among 23,016 461  
assembled contigs longer than 1kb, 1,927 were exclusively detected in D17vir, suggesting 462  
that most of the corresponding viruses were selected during the late stages of the 463  
incubation. Among the 87.59% of contigs that obtained a taxonomic annotation, 98.43% 464  
were affiliated to prokaryotes, which can be explained primarily by database biases: 465  
prokaryotes were more represented than their viruses in the NCBI nr database used for 466  
the taxonomic assignment, and many viral genes could therefore have their best match in 467  
prokaryotic genomes, in particular in proviruses. Such bias is a strong limitation for the 468  
identification of viruses and their taxonomic classification. Yet, it can bring information on 469  
the possible hosts of the viral contigs. 470

Table 1. Overview of the shotgun metagenome datasets.

471

	Cellular metagenomes (individual assemblies)					Viral metagenomes (co-assembly)	
	D13cell	D17cell	D17SIPfr1	D17SIPfr2	D17SIPfr3	D13vir	D17vir
<b>Number of raw reads (millions)</b>	85.68	63.76	75.41	78.78	65.54	38.92	46.48
<b>Reads obtained after trimming (%)</b>	98.39	98.08	97.80	98.35	98.40	98.41	98.61
<b>Number of contigs (<math>\geq 1</math>kb)</b>	77,032	68,382	58,308	33,123	21,159	23,016	
<b>Number of contigs (<math>\geq 3</math>kb)</b>	18,093	15,244	14,728	9,432	5,108	5,571	
<b>Max contig length (b)</b>	622,910	622,910	449,342	388,850	389,221	372,273	
<b>Contigs with taxonomic affiliation (%)</b>	96.62	96.67	96.93	97.32	97.22	87.59	
<b>Reads mapped to contigs (%)</b>	89.35	87.74	89.03	93.17	93.73	87.20	79.85
<b>Number of MAGs</b>	81	67	75	55	28	-	
<b>Number of archaeal MAGs</b>	8	10	8	7	5	-	
<b>Number of selected MAGs (<math>\geq 60\%</math> completeness &amp; <math>\leq 5\%</math> contamination)</b>	81 (including 17 from archaea)						

472

Viral genome detection was performed with CheckV, VIBRANT and VirSorter2. All these tools rely on protein similarity profiles and both latter detect protein profiles by machine learning. CheckV, VirSorter2 and VIBRANT identified 3898, 3968 and 3904 viral genomes (6484 distinct genomes in total), respectively, with most of the genomes being linear and predicted to belong to virulent viruses. These numbers were low compared to the number of contigs longer than 1kb, likely reflecting the fact that shortest contigs contain insufficient information for confident virus identification.

For cellular metagenomes, a taxonomic affiliation was obtained for more than 96% of the contigs, suggesting that they would constitute a good-quality database of host sequences. Metagenome-assembled genomes (MAGs) were reconstructed by binning of the cellular contigs: in total, 81 MAGs (completeness >60% and contamination <5%) were selected as reference hosts, including 17 assigned to the domain Archaea.

For host prediction, two complementary methods were used. The first one relied on matching of CRISPR spacers to the viral contigs. Spacers are short sequences (26 to 72 bp) (Makarova *et al.*, 2011) of plasmid or viral origin, that are integrated into the host genomes by CRISPR-Cas (Nasko *et al.*, 2019), adaptive immunity systems identified in 85% of Archaea and 40% of Bacteria (Makarova *et al.*, 2020). The second approach relied on signatures, as most prokaryotic viruses have a genome k-mer composition similar to the one of their hosts, due to their co-evolution. Similar genomic signatures were searched between viral contigs and the 81 microbial MAGs as potential hosts, using WIsH (Galiez *et al.*, 2017). For contigs longer than 3 kb, the spacer-based method revealed a total of 375 contigs with a predicted host, whereas the signature-based method against our home-made MAG database predicted an acceptable host (p-value  $\leq 0.05$ ) for 3239 contigs. Coupling of these methods resulted in a total of 3466 contigs with a predicted host, showing their complementarity to improve the capability and accuracy of in silico host prediction.

**39 contigs likely originating from viruses of methanogens, dominated by *Caudoviricetes*** 500  
501

As our aim was to identify contigs of viruses infecting methanogenic archaea, we applied 502  
successive stringent filters and manual curation, relying on the integration of results from 503  
complementary bioinformatic analyses (Figure 3). A first filter based on contig length was 504  
applied to eliminate contigs which would contain only limited information. The second 505  
filter aimed at removing an important fraction of the cellular contaminants. In the third 506  
step, contigs possibly originating from archaea or archaeal viruses were selected. Finally, 507  
the contigs obtained at this stage were analyzed manually, to remove most remaining 508  
contaminant contigs (typically originating either from cells or from bacterial viruses). 509

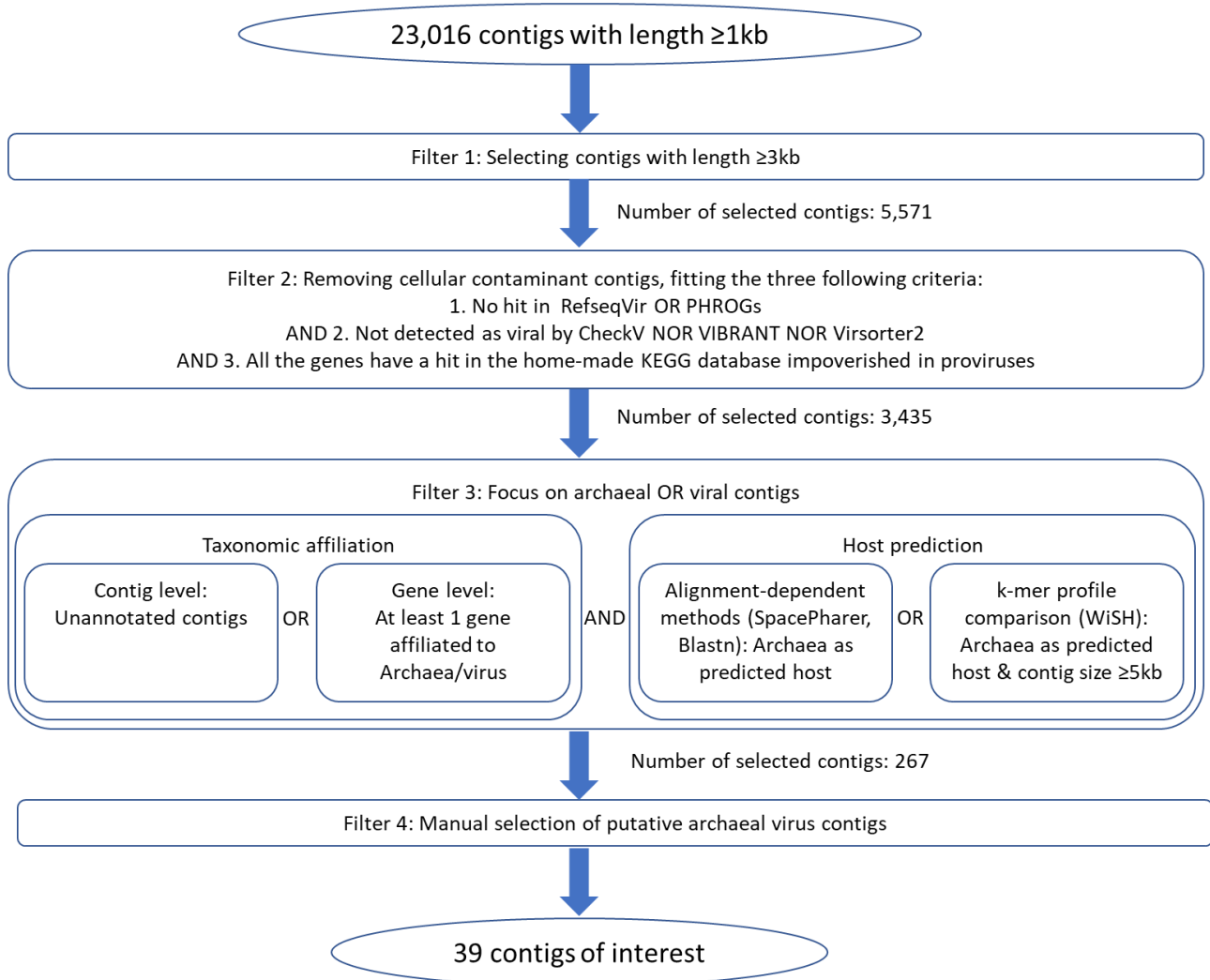


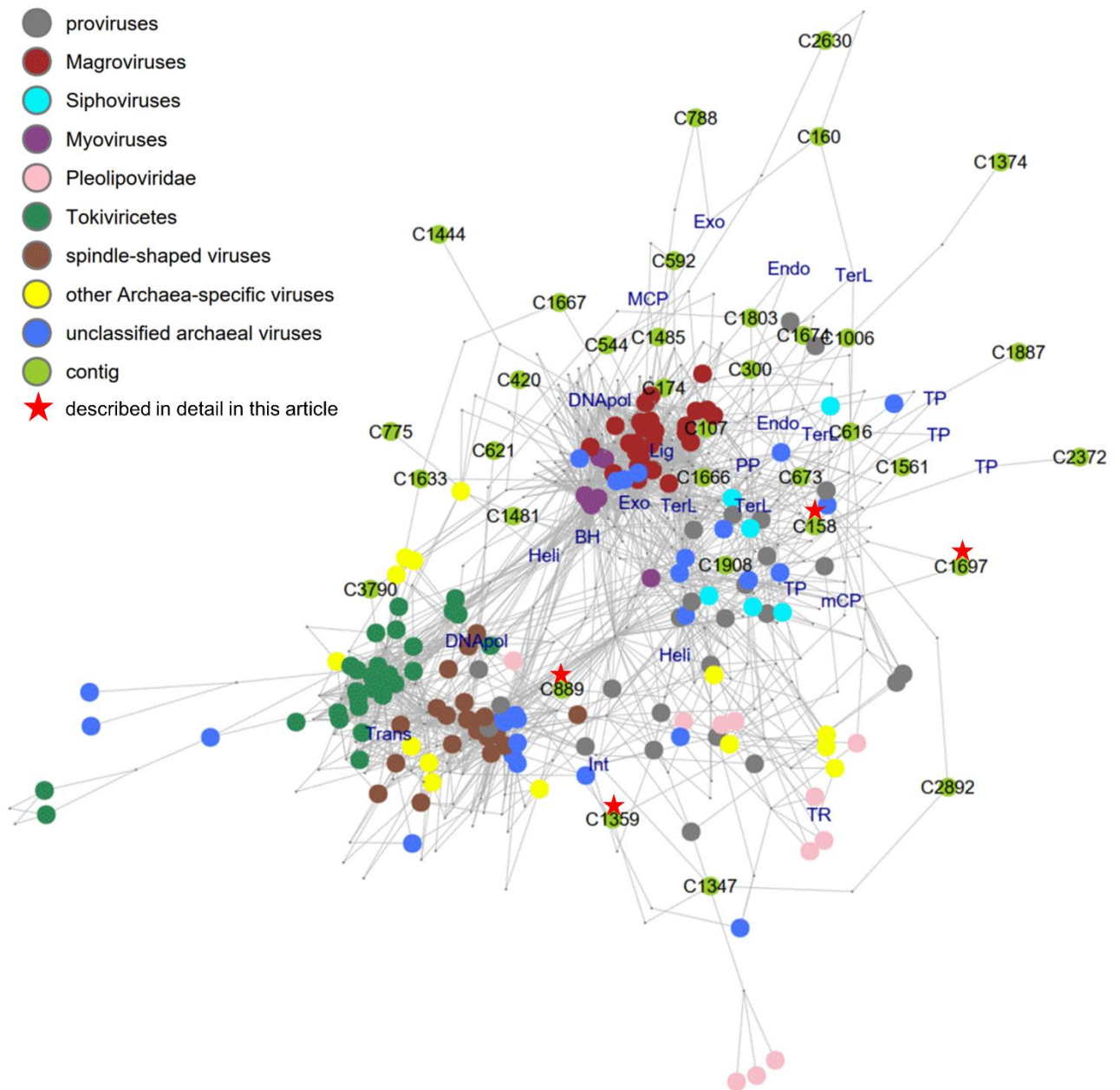
Figure 3. Strategy for the selection of contigs of interest, likely originating from viruses of methanogens 510  
511

512 A total of 39 contigs were thereby selected as contigs of interest for further analysis, as  
513 putatively originating from archaeal viruses. Due to the limited accuracy of some host  
514 prediction tools (e.g. of the order of 70-80% at the phylum level (Galiez *et al.*, 2017)),  
515 this selection of contigs still possibly contained a few contaminants. The selected  
516 contigs ranged in lengths from 4.2 kb to 53 kb (median: 9.4 kb) and the reads per  
517 kilobase per million mapped reads (RPKM) values from 0.07 to 468.27 for D13vir, and  
518 from 0.20 to 229.40 for D17vir. Nine of these contigs could not be assigned to any  
519 known taxon (either cellular or viral), suggesting a great source of previously  
520 uncharacterized viral genomes. Two contigs, C1661 and C1697, had best-predicted  
521 hosts as MAGs from the unlabeled metagenome D17SIPfr1, suggesting that they  
522 represent viruses infecting inactive and minor methanogens in our microcosms.  
523 Besides, 34 were predicted as virulent by VIBRANT and only one, C1485, was predicted  
524 as temperate (Supplementary Figure S5, Supplementary Table S5). Genome scaffold  
525 quality results were obtained through 3 different bioinformatic tools: VIBRANT, CheckV  
526 and VirSorter2: 35 contigs had low or medium quality. However, given that all of these  
527 tools were developed based on the viral databases overwhelmingly dominated by  
528 bacterial *Caudoviricetes*, their accuracy on datasets including novel archaeal viruses,  
529 especially those with small or medium-sized genomes, is not to be expected. All the  
530 results from the bioinformatics tools are available for this set of 39 contigs, in  
531 Supplementary Table S5.

532 A bipartite network was built (Figure 4) to evaluate the similarity between the contigs  
533 of interest and the archaeal (pro)viruses described in the literature. Such networks  
534 enable to represent complex systems comprising two distinct classes of components  
535 (nodes) (Iranzo, Krupovic, *et al.*, 2016). Here, the two classes of node correspond to viral  
536 contigs or genomes and protein orthologous groups (OG), respectively, *i.e.*, two viral  
537 nodes can be connected only indirectly, through shared OGs. Previously-published  
538 archaeal pro(viral) genomes were labeled according to their taxonomic affiliation, when  
539 established. The top half of the network contained viral genomes from (pro)viruses with  
540 a cosmopolitan head-tailed morphotype (siphovirus, myovirus, magrovirus, provirus).  
541 In the bottom half, the presence of various viral families specific to archaea was  
542 observed. Such a network topology is consistent with the spatial distribution in archaeal  
543 virus networks previously described in the literature (Iranzo, Koonin, *et al.*, 2016).  
544 Moreover, head-tail viruses (class *Caudoviricetes*) are known to be highly mosaic  
545 (Krupovič, Forterre, *et al.*, 2010b), hence generating tightly interconnected networks.

546 Out of 39 contigs of interest (green), 24 shared as least one OG with (pro)viruses  
547 belonging to the class *Caudoviricetes*, including virion morphogenesis proteins, such  
548 as terminase large subunit (TerL), various tail proteins, major/minor capsid proteins  
549 (MCP/mCP), and baseplate hub proteins (BH).





550

551 Figure 4. Bipartite network of known archaeal viruses and of the 39 contigs of interest, and of protein  
 552 orthologous groups (OGs). Most of the contigs of interest originate from archaeal viruses. The ge-  
 553 nomes and contigs are represented as circles colored according to the legend shown in the figure, and  
 554 OGs are denoted by the intersections of edges. Some shared OGs with a functional annotation are rep-  
 555 resented. TerL: Terminase large subunit, MCP: major capsid protein, mCP: minor capsid protein, Exo:  
 556 exonuclease, TP: tail protein, PP: portal protein, Lig: ligase, DNAPol, DNA polymerase, BH: baseplate  
 557 hub protein, Heli: helicase, Trans: transposase, TR transcription regulator.

558

559 To further explore the organization of the network and relationships among the 39  
560 contigs of interest and archaeal (pro)viruses, they were clustered according to the  
561 presence/absence of shared genes (see Experimental Procedures). Consistent with the  
562 network analysis, the two main viral clusters (VC) were related to the class  
563 *Caudoviricetes* (Supplementary Figure S6). The first cluster, VC4, included 4 contigs,  
564 C1803, C174, C300 and C616, and was held together through OGs related to the capsid  
565 formation and packaging module, and/or DNA, RNA and nucleotide metabolism  
566 (Supplementary Figure S6). Moreover, in the longest contig of VC4, C174, a sheath  
567 protein was detected, also suggesting its affiliation to viruses with contractile tails  
568 (myovirus-like morphology) (Fokine and Rossmann, 2014). For all the contigs in this  
569 cluster, only WIsH predicted an archaeal host, whereas the other tools employed  
570 showed either a bacterial host or no identified one. To ascertain the host assignment,  
571 the taxonomic and functional annotations of each gene from the cluster were  
572 examined: no single gene was assigned to Archaea or annotated as an archaeal protein,  
573 while many were related to Bacteria, suggesting that bacterial hosts were more  
574 probable. It highlights the importance of relying on different complementary tools for  
575 more accurate host prediction.

576 The second cluster, VC10, included 5 contigs, C1666, C1908, C158, C673 and C1006.  
577 Several OGs shared among members of the cluster were related to various viral  
578 functional modules: virion morphogenesis (including capsid and tail formation, and  
579 genome packaging), DNA/RNA and nucleotide metabolism, as well as integration and  
580 excision. In particular, C158 seem to be closely related to C673 and C1908 as they share  
581 respectively 6 and 2 OGs. VC10 was located near the siphovirus nodes (cyan) in the  
582 network. Importantly, the taxonomic affiliation of these contigs and their host  
583 prediction by different tools were consistent with each other and confirmed that they  
584 originated from archaeal viruses, unlike contigs from VC4.

585 Contigs located on the periphery of the network shared very few OGs with other known

586 viruses (only one or two OGs per contig). Moreover, those shared OGs were often  
587 uncharacterized or hypothetical proteins. Several contigs shared annotated OGs  
588 belonging to viral families specific to archaea, and they could have originated from  
589 previously uncharacterized viruses of methanogens. For example, C775 shared an OG  
590 annotated as helicase with eight genomes belonging to the family *Lipothrixviridae*.  
591 However, only the presence of signature genes, *i.e.* characteristic of particular virus  
592 groups, such as those encoding major structural proteins, can provide a reliable  
593 taxonomic affiliation of contigs, as it is the case for VC4 and VC10 contigs. For contigs  
594 outside of these two viral clusters, such signature genes were identified only for two  
595 contigs (C1359 and C1697). Indeed, a structural protein typical of archaeal spindle-  
596 shaped viruses (Krupovic *et al.*, 2014) was identified in these two contigs, suggesting  
597 that they represent new families of spindle-shaped viruses, as described in more detail  
598 in the next section.

### 599 ***Viruses infecting active methanogens during the incubation***

600 Among the 39 viral contigs of interest, several had predicted hosts corresponding to  
601 the dominant and active methanogens in the studied microcosms, namely, those from  
602 the order Methanobacteriales (*Methanobacterium* and *Methanobrevibacter* genera).

603 In particular, contig C158 (belonging to VC10) was predicted to infect  
604 *Methanobacterium* species, according to its taxonomic affiliation and the signature-  
605 based host prediction. *Methanobacterium* was the most abundant archaeal genus in  
606 the studied microcosms. C158 has a length of 42,490 bp and contains 49 predicted  
607 genes. Consistent with the abundance of the predicted host, C158 was the most  
608 abundant of the 39 contigs in metaviromes at days 13 and 17, with RPKM values of 468  
609 and 229, respectively. Furthermore, C158 is a complete dsDNA circular genome, from  
610 a head-tailed virus. All functional modules typical for the class *Caudoviricetes* were  
611 identified in this contig (Figure 5), such as those required for virion morphogenesis  
612 (HK97-like MCP, mCP, tail proteins, terminase subunits, capsid maturation protease)  
613 and several enzymes necessary for the life cycle, such as an intracellular proteinase

614 inhibitor and an integrase, this latter suggesting a temperate life style. According to its  
615 proximity to siphoviruses in the network (Figure 4) and VIRFAM analysis (Lopes *et al.*,  
616 2014), C158 is likely to have a siphovirus-like morphology, *i.e.*, long non-contractile tail.  
617 Considering the lack of very strong similarity with previously characterized viruses, it  
618 likely represents a new family of archaeal viruses within the *Caudoviricetes*.

619 The second-most abundant contig of interest was C1359, one of the two contigs  
620 predicted to have a spindle-shaped morphology, with RPKM values at days 13 and 17  
621 of 285 and 183, respectively. Surprisingly, this 9936 bp-long contig (17 detected genes)  
622 was predicted to infect *Methanoculleus* sp. (order Methanomicrobiales), which was  
623 present at very low abundances over the incubation time (less than 0.5% of the archaea  
624 based on shotgun metagenomic data). This host prediction was based on spacer  
625 alignment (SpacePHARER) against a public spacer database. By contrast, WISH and  
626 spacer alignment with blastn did not yield any high-confidence predicted host. Based  
627 on the archaeal community composition (see section 3.1.2), this is likely a false  
628 prediction, despite a highly significant p-value ( $2.03e-07$ ). Nevertheless, C1359 gene  
629 content confirmed the probable archaeal nature of its host, and its high abundance  
630 suggests that it infected a dominant methanogen (order Methanobacteriales). Indeed,  
631 it encoded a protein showing a significant similarity with the major capsid proteins of  
632 two spindle-shaped archaeal viruses, the halophilic *Haloarcula hispanica* virus His1  
633 (Bath and Dyall-Smith, 1998) and the hyperthermophilic *Sulfolobus shibatae* virus SSV1  
634 (Palm *et al.*, 1991) (probability: 98.84% and 98.36% respectively). Moreover, spindle-  
635 shaped viruses are specific to archaea (Pina *et al.*, 2011; Snyder *et al.*, 2015; Prangishvili  
636 *et al.*, 2017; Krupovic *et al.*, 2018), excluding the possibility that the virus infects bacteria.  
637 This predicted morphotype is consistent with the presence of spindle-shaped VLPs  
638 observed by TEM (Figure 2K-L) and reported in AD ecosystems (Calusinska *et al.*, 2016).  
639 Due to its limited similarity with previously characterized viruses, C1359 likely  
640 represents a new viral family, and corresponds to the first spindle-shaped virus  
641 reported in association with the order Methanobacteriales.



## 658 ***Viruses infecting minor methanogens in the formate incubation microcosms***

659 As mentioned above, methanogens of the orders Methanomicrobiales and  
660 Methanosarcinales were present in the studied microcosms at a low abundance (less  
661 than 5%). Nevertheless, previously uncharacterized viral genomes possibly infecting  
662 these methanogens were identified.

663 Contig C889 has a length of 14,120 bp and contains 26 predicted genes. Its RPKM  
664 values at days 13 and 17 were of 2.3 and 3.3 respectively. Based on its taxonomic  
665 affiliation, this contig seemed to originate from a virus infecting a Methanomicrobiales  
666 host. In this contig, a protein similar to replicative minichromosome maintenance  
667 (MCM) helicases was detected. This is not a structural protein but it has been identified  
668 in several families of archaeal viruses and archaeal plasmids (Krupovič, Gribaldo, *et al.*,  
669 2010a), strongly supporting an archaeal host. No structural protein could be identified  
670 in C889, suggesting either a partial genome, or a new type of virus.

671 Contig C1697 has a length of 8,207 bp, contains 11 predicted genes, and had RPKM  
672 values of 4.1 at day 13 and 8.4 at day 17. Based on its taxonomic affiliation and on host  
673 prediction with WIsH, this contig originates from a *Methanosarcina* virus. It encodes  
674 one protein showing significant profile similarity with the major capsid protein of SSV1  
675 (*Fuselloviridae*) (probability: 94.92%), suggesting a spindle-shaped virion morphology.  
676 C1697 did not show pronounced similarity with C1359 (which also had a predicted  
677 spindle-shaped morphotype) or with known spindle-shaped viruses, suggesting that it  
678 could represent a second new family of spindle-shaped archaeal viruses. Notably,  
679 however, similar to spindle-shaped halspivirus His1 (Bath and Dyll-Smith, 1998) and  
680 thaspivirus *Nitrosopumilus spindle-shaped virus 1* NSV1 (Kim *et al.*, 2019), C1697  
681 encodes a protein-primed family B DNA polymerase, suggesting that the genome of  
682 this virus is linear.

## 683 **Concluding remarks**

684 Predicting the hosts of viruses is a major bottleneck in microbial ecology. In recent

685 years, beside the classic culture-dependent approaches, several promising methods  
686 have been developed to identify the hosts for viruses within complex microbial  
687 communities, such as PhageFISH (Barrero-Canosa and Moraru, 2019), Meta3C  
688 (Marbouty *et al.*, 2014), epicPCR (Spencer *et al.*, 2016; Sakowski *et al.*, 2021) or viral  
689 tagging (Deng *et al.*, 2014). These single cell level methods generated important new  
690 knowledge, but they also suffer from limitations, either due to the requirement of a  
691 preexisting knowledge on the viral genome sequence, of the capability to cultivate the  
692 host, or due to their complexity. Several purely bioinformatic approaches of host  
693 prediction have been also developed (Edwards *et al.*, 2016; Coclet and Roux, 2021).  
694 However, the accuracy of *in silico* methods is limited, in most cases, by the lack of  
695 microbial genomes closely related to that of the true host. Viral host prediction  
696 therefore remains challenging for metagenomics studies.

697 The original experimental approach developed in the present study is not strictly  
698 speaking a method for the identification of viral hosts but it still presents some  
699 advantages in this perspective. Relying on SIP enabled us to discriminate between the  
700 active microorganisms involved in formate-metabolism and the other ones. Although  
701 this substrate was not consumed exclusively by methanogens, the use of formate  
702 strongly enriched the community in methanogens and their associated viruses,  
703 resulting in a simplified microbial community to study and, possibly, an improved MAG  
704 assembly. Coupling the results from different bioinformatic methods (taxonomic  
705 assignation of the contigs, alignment- and signature-based host prediction), we were  
706 thus able to detect several previously-uncharacterized genomes of viruses infecting  
707 methanogens, with high accuracy, and to determine whether they targeted  
708 methanogens actively involved in the formate bioconversion. It was especially possible  
709 thanks to the establishment of specific host databases, through the shotgun  
710 sequencing of cellular metagenomes. Interestingly, one of them likely corresponds to  
711 a new archaeal virus family within the *Caudoviricetes*, and two others to putative new  
712 families of spindle-shaped archaeal viruses. These results significantly expand the

713 knowledge on the diversity of viruses of methanogens, since no spindle-shaped virus  
714 has been reported until now for the orders Methanobacteriales and Methanosarcinales.  
715 It illustrates the complementarity between SIP, metagenomics and specific  
716 bioinformatic tools for host-virus analysis in complex microbial communities. Our  
717 approach can be viewed as complementary to the available experimental methods for  
718 host identification, and could likely be combined with most of them.

719 To sum up, we proposed an original coupling between SIP, an experimental method,  
720 and metagenomic analyses with complementary bioinformatic tools to identify viruses  
721 of methanogenic archaea within anaerobic digestion microcosms. It enabled successful  
722 enrichment of the microbial community in methanogenic archaea, and their labeling  
723 with  $^{13}\text{C}$ . The *in silico* approach we developed led to the identification of several dozens  
724 of contigs predicted to originate from viruses infecting methanogenic archaea. Their  
725 analyses through gene-sharing network and comparative genomics highlighted the  
726 dominance of the *Caudoviricetes* class, with the discovery of a previously  
727 uncharacterized siphovirus infecting Methanobacteriales hosts. It also led to the  
728 discovery of two spindle-shaped viruses representing two new putative families, not  
729 previously reported for the orders Methanobacteriales and Methanosarcinales. Our  
730 results significantly expand the knowledge on the diversity of viruses of methanogens  
731 and reinforce the notion of the wide environmental and phylogenetic distribution of  
732 spindle-shaped viruses in Archaea (Krupovic *et al.*, 2020). Our original experimental  
733 approach enables to identify viruses infecting key functional groups contributing to  
734 biogeochemical fluxes in communities of uncultured microbes. It can be invaluable for  
735 the study of viruses infecting metabolically active microorganisms in virtually any type  
736 of complex microbial community.

### 737 **Acknowledgments**

738 We are grateful to Chrystelle Bureau for her technical support on 16S rDNA  
739 metabarcoding, and to Angeline Guenne and Nadine Derlet for their technical



740 assistance in analytical chemistry, in INRAE-PROSE. We also would like to acknowledge  
741 Christine Longin of the INRAE-MIMA2, platform for amazing TEM observations.

742 This work has benefited from the facilities and expertise of MIMA2 MET – GABI, INRA,  
743 Agroparistech, 78352 Jouy-en-Josas, France. We also acknowledge the High-  
744 throughput sequencing facility of I2BC (University of Paris-Saclay, Gif-sur-Yvette,  
745 France) for its sequencing.

746 This work was supported by Agence Nationale de la Recherche, France (ANR-17-CE05-  
747 0011, project VIRAME). MK was supported by Agence Nationale de la Recherche grant  
748 ANR-20-CE20-0009.

749 The authors declared to have no conflict of interest.

## 750 **References**

- 751 Ahring, B.K. (2003). Perspectives for Anaerobic Digestion, in: Ahring, B.K., Angelidaki, I.,  
752 de Macario, E.C., Gavala, H.N., Hofman-Bang, J., Macario, A.J.L., Elferink,  
753 S.J.W.H.O., Raskin, L., Stams, A.J.M., Westermann, P., Zheng, D. (Eds.),  
754 Biomethanation I, Advances in Biochemical Engineering/Biotechnology.  
755 Springer, Berlin, Heidelberg, pp. 1–30.
- 756 Baquero, D.P., Liu, Y., Wang, F., Egelman, E.H., Prangishvili, D., Krupovic, M. (2020).  
757 Chapter Four - Structure and assembly of archaeal viruses, in: Kielian, M.,  
758 Mettenleiter, T.C., Roossinck, M.J. (Eds.), Advances in Virus Research, Virus  
759 Assembly and Exit Pathways. Academic Press, pp. 127–164.
- 760 Barrero-Canosa, J., Moraru, C. (2019). PhageFISH for Monitoring Phage Infections at  
761 Single Cell Level. *Methods Mol Biol* 1898, 1–26.
- 762 Bath, C., Dyll-Smith, M.L. (1998). His1, an Archaeal Virus of the Fuselloviridae Family  
763 That Infects *Haloarcula hispanica*. *Journal of Virology*.
- 764 Biswas, A., Staals, R.H.J., Morales, S.E., Fineran, P.C., Brown, C.M. (2016). CRISPRDetect:  
765 A flexible algorithm to define CRISPR arrays. *BMC Genomics* 17, 356.
- 766 Borrel, G., Colombet, J., Robin, A., Lehours, A.-C., Prangishvili, D., Sime-Ngando, T.  
767 (2012). Unexpected and novel putative viruses in the sediments of a deep-  
768 dark permanently anoxic freshwater habitat. *ISME J* 6, 2119–2127.
- 769 Buchfink, B., Xie, C., Huson, D.H. (2015). Fast and sensitive protein alignment using  
770 DIAMOND. *Nat Methods* 12, 59–60.

- 771 Calusinska, M., Marynowska, M., Goux, X., Lentzen, E., Delfosse, P. (2016). Analysis of  
772 dsDNA and RNA viromes in methanogenic digesters reveals novel viral  
773 genetic diversity. *Environ. Microbiol.* 18, 1162–1175.
- 774 Chapleur, O., Mazeas, L., Godon, J.-J., Bouchez, T. (2016). Asymmetrical response of  
775 anaerobic digestion microbiota to temperature changes. *Appl Microbiol  
776 Biotechnol* 100, 1445–1457.
- 777 Coclet, C., Roux, S. (2021). Global overview and major challenges of host prediction  
778 methods for uncultivated phages. *Current Opinion in Virology* 49, 117–126.
- 779 Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B., Rocha,  
780 E.P.C., Vergnaud, G., Gautheret, D., Pourcel, C. (2018). CRISPRCasFinder, an  
781 update of CRISRFinder, includes a portable version, enhanced performance  
782 and integrates search for Cas proteins. *Nucleic Acids Research* 46, W246–  
783 W251.
- 784 Csardi, G., Nepusz, T. (2005). The Igraph Software Package for Complex Network  
785 Research. *InterJournal Complex Systems*, 1695.
- 786 Deng, L., Ignacio-Espinoza, J.C., Gregory, A.C., Poulos, B.T., Weitz, J.S., Hugenholtz, P.,  
787 Sullivan, M.B. (2014). Viral tagging reveals discrete populations in  
788 *Synechococcus* viral genome sequence space. *Nature* 513, 242–245.
- 789 Eason, R., Campbell, A.M. (1978). 8 - Analytical Ultracentrifugation, in: Birnie, G.D.,  
790 Rickwood, D. (Eds.), *Centrifugal Separations in Molecular and Cell Biology*.  
791 Butterworth-Heinemann, pp. 251–287.
- 792 Edwards, R.A., McNair, K., Faust, K., Raes, J., Dutilh, B.E. (2016). Computational  
793 approaches to predict bacteriophage–host relationships. *FEMS Microbiology  
794 Reviews* 40, 258–272.
- 795 Evans, P.N., Boyd, J.A., Leu, A.O., Woodcroft, B.J., Parks, D.H., Hugenholtz, P., Tyson, G.W.  
796 (2019). An evolving view of methane metabolism in the Archaea. *Nature  
797 Reviews Microbiology* 1.
- 798 Fokine, A., Rossmann, M.G. (2014). Molecular architecture of tailed double-stranded  
799 DNA phages. *Bacteriophage* 4, e28281.
- 800 Galiez, C., Siebert, M., Enault, F., Vincent, J., Söding, J. (2017). WISH: who is the host?  
801 Predicting prokaryotic hosts from metagenomic phage contigs.  
802 *Bioinformatics* 33, 3113–3114.
- 803 Guo, J., Bolduc, B., Zayed, A.A., Varsani, A., Dominguez-Huerta, G., Delmont, T.O.,  
804 Pratama, A.A., Gazitúa, M.C., Vik, D., Sullivan, M.B., Roux, S. (2021). VirSorter2:  
805 a multi-classifier, expert-guided approach to detect diverse DNA and RNA  
806 viruses. *Microbiome* 9, 37.
- 807 He, J., Wang, X., Yin, X., Li, Q., Li, X., Zhang, Y., Deng, Y. (2018). Insights into biomethane

808 production and microbial community succession during semi-continuous  
809 anaerobic digestion of waste cooking oil under different organic loading  
810 rates. *AMB Express* 8, 92.

811 Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J. (2010).  
812 Prodigal: prokaryotic gene recognition and translation initiation site  
813 identification. *BMC Bioinformatics* 11, 119.

814 Iranzo, J., Koonin, E.V., Prangishvili, D., Krupovic, M. (2016a). Bipartite Network Analysis  
815 of the Archaeal Virosphere: Evolutionary Connections between Viruses and  
816 Capsidless Mobile Elements. *J Virol* 90, 11043–11055.

817 Iranzo, J., Krupovic, M., Koonin, E.V. (2016b). The Double-Stranded DNA Virosphere as  
818 a Modular Hierarchical Network of Gene Sharing. *mBio* 7, e00978-16.

819 Kanehisa, M., Sato, Y., Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools  
820 for Functional Characterization of Genome and Metagenome Sequences.  
821 *Journal of Molecular Biology, Computation Resources for Molecular Biology*  
822 428, 726–731.

823 Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., Wang, Z. (2019). MetaBAT 2: an  
824 adaptive binning algorithm for robust and efficient genome reconstruction  
825 from metagenome assemblies. *PeerJ* 7, e7359.

826 Kieft, K., Zhou, Z., Anantharaman, K. (2020). VIBRANT: automated recovery, annotation  
827 and curation of microbial viruses, and evaluation of viral community function  
828 from genomic sequences. *Microbiome* 8, 90.

829 Kim, J.-G., Kim, S.-J., Cvirkaite-Krupovic, V., Yu, W.-J., Gwak, J.-H., López-Pérez, M.,  
830 Rodriguez-Valera, F., Krupovic, M., Cho, J.-C., Rhee, S.-K. (2019). Spindle-  
831 shaped viruses infect marine ammonia-oxidizing thaumarchaea. *PNAS* 116,  
832 15645–15650.

833 Köster, J., Rahmann, S. (2012). Snakemake—a scalable bioinformatics workflow engine.  
834 *Bioinformatics* 28, 2520–2522.

835 Krupovič, M., Bamford, D.H. (2008). Archaeal proviruses TKV4 and MVV extend the  
836 PRD1-adenovirus lineage to the phylum Euryarchaeota. *Virology* 375, 292–  
837 300.

838 Krupovic, M., Cvirkaite-Krupovic, V., Iranzo, J., Prangishvili, D., Koonin, E.V. (2018).  
839 Viruses of archaea: Structural, functional, environmental and evolutionary  
840 genomics. *Virus Res* 244, 181–193.

841 Krupovic, M., Dolja, V.V., Koonin, E.V. (2020.) The LUCA and its complex virome. *Nat Rev*  
842 *Microbiol* 18, 661–670.

843 Krupovič, M., Forterre, P., Bamford, D.H. (2010a). Comparative Analysis of the Mosaic  
844 Genomes of Tailed Archaeal Viruses and Proviruses Suggests Common

- 845 Themes for Virion Architecture and Assembly with Tailed Viruses of Bacteria.  
846 *Journal of Molecular Biology* 397, 144–160.
- 847 Krupovič, M., Gribaldo, S., Bamford, D.H., Forterre, P. (2010b). The Evolutionary History  
848 of Archaeal MCM Helicases: A Case Study of Vertical Evolution Combined  
849 with Hitchhiking of Mobile Genetic Elements. *Molecular Biology and*  
850 *Evolution* 27, 2716–2732.
- 851 Krupovič, M., Gribaldo, S., Bamford, D.H., Forterre, P. (2010c). The Evolutionary History  
852 of Archaeal MCM Helicases: A Case Study of Vertical Evolution Combined  
853 with Hitchhiking of Mobile Genetic Elements. *Molecular Biology and*  
854 *Evolution* 27, 2716–2732.
- 855 Krupovic, M., Quemin, E.R.J., Bamford, D.H., Forterre, P., Prangishvili, D. (2014).  
856 Unification of the Globally Distributed Spindle-Shaped Viruses of the  
857 Archaea. *Journal of Virology*.
- 858 Langmead, B., Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat*  
859 *Methods* 9, 357–359.
- 860 Lei, Y., Wei, L., Liu, T., Xiao, Y., Dang, Y., Sun, D., Holmes, D.E. (2018). Magnetite enhances  
861 anaerobic digestion and methanogenesis of fresh leachate from a municipal  
862 solid waste incineration plant. *Chemical Engineering Journal* 348, 992–999.
- 863 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis,  
864 G., Durbin, R., 1000 Genome Project Data Processing Subgroup. (2009). The  
865 Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–  
866 2079.
- 867 Liang, B., Wang, L.-Y., Mbadinga, S.M., Liu, J.-F., Yang, S.-Z., Gu, J.-D., Mu, B.-Z. (2015).  
868 Anaerolineaceae and Methanosaeta turned to be the dominant  
869 microorganisms in alkanes-dependent methanogenic culture after long-  
870 term of incubation. *AMB Express* 5, 37.
- 871 Lin, Q., De Vrieze, J., Li, J., Li, X. (2016). Temperature affects microbial abundance, activity  
872 and interactions in anaerobic digestion. *Bioresource Technology* 209, 228–  
873 236.
- 874 Liu, Y., Demina, T.A., Roux, S., Aiewsakun, P., Kazlauskas, D., Simmonds, P., Prangishvili,  
875 D., Oksanen, H.M., Krupovic, M. (2021). Diversity, taxonomy, and evolution of  
876 archaeal viruses of the class Caudoviricetes. *PLOS Biology* 19, e3001442.
- 877 Lopes, A., Tavares, P., Petit, M.-A., Guérois, R., Zinn-Justin, S. (2014). Automated  
878 classification of tailed bacteriophages according to their neck organization.  
879 *BMC Genomics* 15, 1027.
- 880 Luo, Y., Pfister, P., Leisinger, T., Wasserfallen, A. (2001). The Genome of Archaeal  
881 Prophage  $\Psi$ M100 Encodes the Lytic Enzyme Responsible for Autolysis of

- 882 Methanothermobacter wolfeii. Journal of Bacteriology.
- 883 Lyu, Z., Shao, N., Akinyemi, T., Whitman, W.B. (2018). Methanogenesis. Current Biology  
884 28, R727–R732.
- 885 Madigou, C., Lê Cao, K.-A., Bureau, C., Mazéas, L., Déjean, S., Chapleur, O. (2019).  
886 Ecological consequences of abrupt temperature changes in anaerobic  
887 digesters. Chemical Engineering Journal 361, 266–277.
- 888 Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J.J., Charpentier, E., Horvath, P.,  
889 Moineau, S., Mojica, F.J.M., Wolf, Y.I., Yakunin, A.F., van der Oost, J., Koonin,  
890 E.V. (2011). Evolution and classification of the CRISPR-Cas systems. Nat Rev  
891 Microbiol 9, 467–477.
- 892 Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J.,  
893 Charpentier, E., Cheng, D., Haft, D.H., Horvath, P., Moineau, S., Mojica, F.J.M.,  
894 Scott, D., Shah, S.A., Siksnys, V., Terns, M.P., Venclovas, Č., White, M.F.,  
895 Yakunin, A.F., Yan, W., Zhang, F., Garrett, R.A., Backofen, R., van der Oost, J.,  
896 Barrangou, R., Koonin, E.V. (2020). Evolutionary classification of CRISPR-Cas  
897 systems: a burst of class 2 and derived variants. Nat Rev Microbiol 18, 67–83.
- 898 Marbouty, M., Cournac, A., Flot, J.-F., Marie-Nelly, H., Mozziconacci, J., Koszul, R. (2014).  
899 Metagenomic chromosome conformation capture (meta3C) unveils the  
900 diversity of chromosome organization in microorganisms. Elife 3, e03318.
- 901 Meile, L., Jenal, U., Studer, D., Jordan, M., Leisinger, T. (1989). Characterization of  $\psi$ M1,  
902 a virulent phage of Methanobacterium thermoautotrophicum Marburg.  
903 Arch. Microbiol. 152, 105–110.
- 904 Menzel, P., Ng, K.L., Krogh, A. (2016). Fast and sensitive taxonomic classification for  
905 metagenomics with Kaiju. Nature Communications 7, 11257.
- 906 Molnár, J., Magyar, B., Schneider, G., Laczi, K., Valappil, S.K., Kovács, Á.L., Nagy, I.K.,  
907 Rákhely, G., Kovács, T. (2020). Identification of a novel archaea virus, detected  
908 in hydrocarbon polluted Hungarian and Canadian samples. PLOS ONE 15,  
909 e0231864.
- 910 Nasko, D.J., Ferrell, B.D., Moore, R.M., Bhavsar, J.D., Polson, S.W., Wommack, K.E. (2019).  
911 CRISPR Spacers Indicate Preferential Matching of Specific Virioplankton  
912 Genes. mBio.
- 913 Nayfach, S., Camargo, A.P., Eloë-Fadrosh, E., Roux, S., Kyrpides, N. (2021). CheckV:  
914 assessing the quality of metagenome-assembled viral genomes. Nat  
915 Biotechnol 39, 578–585.
- 916 Neufeld, J.D., Vohra, J., Dumont, M.G., Lueders, T., Manefield, M., Friedrich, M.W.,  
917 Murrell, J.C. (2007). DNA stable-isotope probing. Nat Protoc 2, 860–866.
- 918 Nölling, J., Groffen, A., de Vos, W.M. (1993).  $\phi$  F1 and  $\phi$ F3, two novel virulent, archaeal

- 919 phages infecting different thermophilic strains of the genus  
920 *Methanobacterium*. *Microbiology* 139, 2511–2516.
- 921 Nurk, S., Meleshko, D., Korobeynikov, A., Pevzner, P. (2016). metaSPAdes: a new  
922 versatile de novo metagenomics assembler. arXiv:1604.03071 [q-bio].
- 923 Olo Ndela, E., Enault, F., Toussaint, A. (2021). Transposable Prophages in *Leptospira*: An  
924 Ancient, Now Diverse, Group Predominant in Causative Agents of Weil's  
925 Disease. *International Journal of Molecular Sciences* 22, 13434.
- 926 Palm, P., Schleper, C., Grampp, B., Yeats, S., McWilliam, P., Reiter, W.-D., Zillig, W. (1991).  
927 Complete nucleotide sequence of the virus SSV1 of the archaeobacterium  
928 *Sulfolobus shibatae*. *Virology* 185, 242–250.
- 929 Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W. (2015). CheckM:  
930 assessing the quality of microbial genomes recovered from isolates, single  
931 cells, and metagenomes. *Genome Res.* 25, 1043–1055.
- 932 Parks, D.H., Rinke, C., Chuvochina, M., Chaumeil, P.-A., Woodcroft, B.J., Evans, P.N.,  
933 Hugenholtz, P., Tyson, G.W. (2017). Recovery of nearly 8,000 metagenome-  
934 assembled genomes substantially expands the tree of life. *Nat Microbiol* 2,  
935 1533–1542.
- 936 Pfister, P., Wasserfallen, A., Stettler, R., Leisinger, T. (1998). Molecular analysis of  
937 *Methanobacterium* phage  $\Psi$ M2. *Molecular Microbiology* 30, 233–244.
- 938 Filosof, A., Yutin, N., Flores-Urbe, J., Sharon, I., Koonin, E.V., Béjà, O. (2017). Novel  
939 Abundant Oceanic Viruses of Uncultured Marine Group II Euryarchaeota.  
940 *Curr Biol* 27, 1362–1368.
- 941 Pickard, D.J.J. (2009). Preparation of bacteriophage lysates and pure DNA. *Methods Mol*  
942 *Biol* 502, 3–9.
- 943 Pina, M., Bize, A., Forterre, P., Prangishvili, D. (2011). The archeoviruses. *FEMS Microbiol.*  
944 *Rev.* 35, 1035–1054.
- 945 Poirier, S., Desmond-Le Quéméner, E., Madigou, C., Bouchez, T., Chapleur, O. (2016).  
946 Anaerobic digestion of biowaste under extreme ammonia concentration:  
947 Identification of key microbial phylotypes. *Bioresour Technol* 207, 92–101.
- 948 Prangishvili, D., Bamford, D.H., Forterre, P., Irazzo, J., Koonin, E.V., Krupovic, M. (2017).  
949 The enigmatic archaeal virosphere. *Nature Reviews Microbiology* 15, 724–  
950 739.
- 951 Puig-Castellví, F., Midoux, C., Guenne, A., Conteau, D., Franchi, O., Bureau, C., Madigou,  
952 C., Jouan-Rimbaud Bouveresse, D., Kroff, P., Mazéas, L., Rutledge, D.N., Gaval,  
953 G., Chapleur, O. (2022). Metataxonomics, metagenomics and metabolomics  
954 analysis of the influence of temperature modification in full-scale anaerobic  
955 digesters. *Bioresource Technology* 346, 126612.

- 956 Radajewski, S., Ineson, P., Parekh, N.R., Murrell, J.C. (2000). Stable-isotope probing as a  
957 tool in microbial ecology. *Nature* 403, 646–649.
- 958 Sakowski, E.G., Arora-Williams, K., Tian, F., Zayed, A.A., Zablocki, O., Sullivan, M.B.,  
959 Preheim, S.P. (2021). Interaction dynamics and virus–host range for estuarine  
960 actinophages captured by epicPCR. *Nat Microbiol* 6, 630–642.
- 961 Si, B., Liu, Z., Zhang, Y., Li, J., Shen, R., Zhu, Z., Xing, X. (2016). Towards biohythane  
962 production from biomass: Influence of operational stage on anaerobic  
963 fermentation and microbial community. *International Journal of Hydrogen*  
964 *Energy* 41, 4429–4438.
- 965 Snyder, J.C., Bolduc, B., Young, M.J. (2015). 40 Years of archaeal virology: Expanding  
966 viral diversity. *Virology*, 60th Anniversary Issue 479–480, 369–378.
- 967 Spencer, S.J., Tamminen, M.V., Preheim, S.P., Guo, M.T., Briggs, A.W., Brito, I.L., Weitz,  
968 D.A., Pitkänen, L.K., Vigneault, F., Virta, M.Pj., Alm, E.J. (2016). Massively  
969 parallel sequencing of single cells by epicPCR links functional genes with  
970 phylogenetic markers. *The ISME Journal* 10, 427–436.
- 971 Steinegger, M., Meier, M., Mirdita, M., Vöhringer, H., Haunsberger, S.J., Söding, J. (2019).  
972 HH-suite3 for fast remote homology detection and deep protein annotation.  
973 *BMC Bioinformatics* 20, 1–15.
- 974 Sun, H., Yang, Z., Shi, G., Arhin, S.G., Papadakis, V.G., Goula, M.A., Zhou, L., Zhang, Y.,  
975 Liu, G., Wang, W. (2021). Methane production from acetate, formate and  
976 H<sub>2</sub>/CO<sub>2</sub> under high ammonia level: Modified ADM1 simulation and  
977 microbial characterization. *Science of The Total Environment* 783, 147581.
- 978 Terzian, P., Olo Ndela, E., Galiez, C., Lossouarn, J., Pérez Bucio, R.E., Mom, R., Toussaint,  
979 A., Petit, M.-A., Enault, F. (2021). PHROG: families of prokaryotic virus proteins  
980 clustered using remote homology. *NAR Genomics and Bioinformatics* 3,  
981 lqab067.
- 982 Thauer, R.K., Kaster, A.-K., Seedorf, H., Buckel, W., Hedderich, R., (2008). Methanogenic  
983 archaea: ecologically relevant differences in energy conservation. *Nat Rev*  
984 *Microbiol* 6, 579–591.
- 985 Thiroux, S., Dupont, S., Nesbø, C.L., Bienvenu, N., Krupovic, M., L’Haridon, S., Marie, D.,  
986 Forterre, P., Godfroy, A., Geslin, C. (2021). The first head-tailed virus, MFTV1,  
987 infecting hyperthermophilic methanogenic deep-sea archaea.  
988 *Environmental Microbiology* 23, 3614–3626.
- 989 von Meijenfeldt, F.A.B., Arkhipova, K., Cambuy, D.D., Coutinho, F.H., Dutilh, B.E. (2019).  
990 Robust taxonomic classification of uncharted microbial sequences and bins  
991 with CAT and BAT. *Genome Biology* 20, 217.
- 992 Wang, J., Zhao, Z., Zhang, Y. (2021). Enhancing anaerobic digestion of kitchen wastes

- 993 with biochar: Link between different properties and critical mechanisms of  
994 promoting interspecies electron transfer. *Renewable Energy* 167, 791–799.
- 995 Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R. (2007). Naïve Bayesian Classifier for Rapid  
996 Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Appl*  
997 *Environ Microbiol* 73, 5261–5267.
- 998 Wang, Y., Qian, P.-Y. (2009). Conservative fragments in bacterial 16S rRNA genes and  
999 primer design for 16S ribosomal DNA amplicons in metagenomic studies.  
1000 *PLoS One* 4, e7401.
- 1001 Weidenbach, K., Nickel, L., Neve, H., Alkhnbashi, O.S., Künzel, S., Kupczok, A.,  
1002 Bauersachs, T., Cassidy, L., Tholey, A., Backofen, R., Schmitz, R.A. (2017).  
1003 *Methanosarcina* Spherical Virus, a Novel Archaeal Lytic Virus Targeting  
1004 *Methanosarcina* Strains. *Journal of Virology* 91, e00955-17.
- 1005 Weidenbach, K., Wolf, S., Kupczok, A., Kern, T., Fischer, M.A., Reetz, J., Urbańska, N.,  
1006 Künzel, S., Schmitz, R.A., Rother, M. (2021). Characterization of Blf4, an  
1007 Archaeal Lytic Virus Targeting a Member of the Methanomicrobiales. *Viruses*  
1008 13, 1934.
- 1009 Whiticar, M.J., Faber, E., Schoell, M. (1986). Biogenic methane formation in marine and  
1010 freshwater environments: CO<sub>2</sub> reduction vs. acetate fermentation—Isotope  
1011 evidence. *Geochimica et Cosmochimica Acta* 50, 693–709.
- 1012 Wolf, S., Fischer, M.A., Kupczok, A., Reetz, J., Kern, T., Schmitz, R.A., Rother, M. (2019).  
1013 Characterization of the lytic archaeal virus Drs3 infecting *Methanobacterium*  
1014 *formicicum*. *Arch. Virol.* 164, 667–674.
- 1015 Wood, A.G., Whitman, W.B., Konisky, J. (1989). Isolation and characterization of an  
1016 archaeobacterial viruslike particle from *Methanococcus voltae* A3. *Journal of*  
1017 *Bacteriology* 171, 93–98.
- 1018 Zhang, R., Mirdita, M., Levy Karin, E., Norroy, C., Galiez, C., Söding, J. (2021).  
1019 SpacePHARER: sensitive identification of phages from CRISPR spacers in  
1020 prokaryotic hosts. *Bioinformatics*.
- 1021 Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F.,  
1022 Söding, J., Lupas, A.N., Alva, V. (2018). A Completely Reimplemented MPI  
1023 Bioinformatics Toolkit with a New HHpred Server at its Core. *Journal of*  
1024 *Molecular Biology, Computation Resources for Molecular Biology* 430, 2237–  
1025 2243.
- 1026 Zolfo, M., Pinto, F., Asnicar, F., Manghi, P., Tett, A., Bushman, F.D., Segata, N. (2019).  
1027 Detecting contamination in viromes using ViromeQC. *Nat Biotechnol* 37,  
1028 1408–1412.
- 1029



1030 **Supplementary text**

1031 **1. Materials and methods**

1032 *1.1. Viral DNA extraction with Phenol-Chloroform method*

1033 Capsid digestion was performed by adding 10 µL of SDS 20% and 4 µL of proteinase K  
1034 (10 mg/mL, Ambion), followed by incubation at 37 °C for 30 min. After digestion, the  
1035 obtained solution was transferred into a 1.5 mL Phase lock gel, light (VWR). For viral  
1036 DNA extraction 0.5 mL of phenol:chloroform:isoamyl alcohol (25:24:1) (Sigma) were  
1037 added and a centrifugation at 1,500 g for 5 min was performed. This step was repeated  
1038 once. Subsequently, 0.5 mL of chloroform: IAA (24:1) (Sigma) was added, before  
1039 centrifuging at 6,000 g for 5 min.

1040 DNA was precipitated at room temperature for 20 min by adding 45 µL of 3 M sodium  
1041 acetate (Sigma) and 500 µL of 100 % isopropanol (Thermo). The DNA pellet was  
1042 washed twice with 70 % ethanol and finally suspended in 100 µL of Tris-HCl 10 mM  
1043 buffer (Thermo).

1044 *1.2. 16S rRNA gene metabarcoding*

1045 The PCR mixture contained 1X SuperFi Buffer, 0.2 mM each dNTP, 0.5 µM each primer,  
1046 1 U Platinum SuperFi DNA Polymerase (Invitrogen), and 10–20 ng template DNA in a  
1047 50 µL reaction volume. Amplification was performed as follows: 30 sec at 98 °C, 30  
1048 cycles of 10 sec at 98 °C, 10 sec at 50 °C, 30 sec at 72 °C, followed by a final extension  
1049 of 5 min at 72 °C.

1050 PCR products were purified using Solid Phase Reversible Immobilization (SPRI)  
1051 magnetic beads (Mag-Bind TotalPure NGS magnetic beads, Omega Bio-Tek) according  
1052 to the manufacturer's instructions, with a bead: amplicon ratio of 1.2; they were eluted  
1053 in 45 µL TE Buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). Purified amplicons were  
1054 quantified using D1000 ScreenTape and 2200 TapeStation or DNA 1000 Kit and 2100  
1055 Bioanalyzer (Agilent Technologies), following the manufacturer's instructions. Then, all

1056 amplicons were pre-diluted at 500 pM in molecular grade water and equimolarly  
1057 pooled. The pool was then diluted at 100 pM for sequencing.

1058 In order to prepare template-positive Ion Sphere Particles (ISPs) containing DNA  
1059 clonally amplified by emulsion PCR, the library was diluted to 26 pM and set up on the  
1060 Ion OneTouch 2 Instrument (Life Technologies) using the Ion PGM Hi-Q View OT2 Kit  
1061 (Life Technologies) and following the manufacturer's instructions. These templated ISPs  
1062 were then purified on the Ion OneTouch ES (Life Technologies) according to the  
1063 manufacturer's instructions.

1064 *1.3. Bipartite network analysis: Orthologous Groups (OGs) formation and viral ge-*  
1065 *nome clustering*

1066 All the proteins encoded in these sequences were categorized into orthologous groups  
1067 (OGs) by a two-step procedure: (i) all proteins were compared to each other using  
1068 MMseqs (Steinegger and Söding, 2017) (bit score  $\geq 50$  and reciprocal coverage  $> 80\%$ )  
1069 and clustered with MCL (Enright *et al.*, 2002) (inflation 2.0); and (ii) for each protein  
1070 cluster, a multiple alignment was built using ClustalOmega (Sievers *et al.*, 2011),  
1071 transformed into an HMM profile and all profiles were compared to each other using  
1072 HHsearch of the HHSuite toolkit (Söding, 2005) (version 2.0.16, cut-offs: probability  
1073  $\geq 90\%$  and coverage  $\geq 50\%$ ). Singleton proteins were also individually compared to all  
1074 these HMM profiles. Using significant similarities, singletons and protein clusters were  
1075 then once more clustered with MCL (inflation 2.0), in order to group clusters that were  
1076 similar, yet not already gathered due to the entangled nature of similarity networks  
1077 and to the fact that larger and more diverse clusters enable the detection of more  
1078 distant homologies.

1079 First, a bipartite network was determined using these OGs and viruses (contigs and  
1080 genomes) (Figure 4). To reduce the complexity of this network, viruses were also  
1081 grouped into viral clusters. To this end, the number of OGs shared by a pair of viruses  
1082 was computed and only pairs sharing more than 10 % of their OGs and at least two

1083 OGs were used to cluster viruses with MCL (Enright et al. 2002) (inflation 2.0). A second  
1084 bipartite network was thus defined using OGs and these viral clusters (Supplementary  
1085 Figure S6)

## 1086 **2. Results**

### 1087 *2.1. Isotopic analysis*

1088 In order to identify the active methanogenic pathway(s), the isotopic composition of  
1089 the biogas was analyzed by the coupling of a gas chromatograph (Trace GC Ultra,  
1090 Thermo Fisher Scientific) with an Isotope-Ratio Mass Spectrometer (IRMS) (Delta V  
1091 PLUS, Thermo Fisher Scientific). Delta expression  $\delta^{13}\text{C}$  (‰) of methane was used to  
1092 identify the active methanogenesis pathway(s) (Whiticar *et al.*, 1986). This index was  
1093 calculated using the following formula:

$$1094 \quad \delta^{13}\text{C} = \frac{R_{\text{sample}} - R_{\text{standard}}}{R_{\text{standard}}} \times 1000 \text{ [‰]}$$

1095 where R was the ratio  $^{13}\text{C}/^{12}\text{C}$  and  $R_{\text{standard}}$  was equal to 0.0112372.

1096 In the microcosms fed with unlabeled formate, the  $\delta^{13}\text{C}$  (‰) of methane was always  
1097 below 62‰, indicating the dominance of the hydrogenotrophic pathway.

## Supplementary tables

Table S1. Synthetic view on the viruses of methanogenic archaea isolated so far

Name	Class	Family	Morphology	Host order	Host genus	Origin	Reference
<b>ΨM1</b>	<i>Caudoviricetes</i>	<i>Leisingerviridae</i>	siphovirus	Methanobacteriales	<i>Methanothermobacter</i>	thermophilic AD	(Meile <i>et al.</i> , 1989; Liu <i>et al.</i> , 2021)
<b>ΨM2</b>	<i>Caudoviricetes</i>	<i>Leisingerviridae</i>	siphovirus	Methanobacteriales	<i>Methanothermobacter</i>	thermophilic AD	(Meile <i>et al.</i> , 1989; Liu <i>et al.</i> , 2021)
<b>ΨM100</b>	<i>Caudoviricetes</i>	-	-	Methanobacteriales	<i>Methanothermobacter</i>	thermophilic AD	(Luo <i>et al.</i> , 2001)
<b>ΦF1</b>	unclassified	unclassified	siphovirus	Methanobacteriales	<i>Methanobacterium</i>	thermophilic AD	(Nölling <i>et al.</i> , 1993)
<b>ΦF3</b>	unclassified	unclassified	siphovirus	Methanobacteriales	<i>Methanobacterium</i>	thermophilic AD	(Nölling <i>et al.</i> , 1993)
<b>Drs3</b>	<i>Caudoviricetes</i>	<i>Anaerodiviridae</i>	siphovirus	Methanobacteriales	<i>Methanobacterium</i>	mesophilic AD	(Wolf <i>et al.</i> , 2019; Liu <i>et al.</i> , 2021)
<b>BIf4</b>	unclassified	unclassified	siphovirus	Methanomicrobiales	<i>Methanoculleus</i>	mesophilic AD	(Weidenbach <i>et al.</i> , 2021)
<b>MetSV</b>	unclassified	unclassified	spherical	Methanosarcinales	<i>Methanosarcina</i>	mesophilic AD	(Weidenbach <i>et al.</i> , 2017)
<b>A3-VLP</b>	unclassified	unclassified	oblate/spindle	Methanococcales	<i>Methanococcus</i>	mesophilic AD	(Wood <i>et al.</i> , 1989)
<b>MFTV1</b>	unclassified	unclassified	siphovirus	Methanococcales	<i>Methanocaldococcus</i>	deep-sea hydrothermal vents	(Thiroux <i>et al.</i> , 2021)

Table S2. Bioinformatic tool parameters and database versions

Tool	Version	Options	Employed database	Reference website
<b>fastp</b>	0.20.0	--length_required 50	NA	<a href="https://github.com/OpenGene/fastp">https://github.com/OpenGene/fastp</a>
<b>MetaSPADES</b>	3.14.0	--meta	NA	<a href="https://github.com/ablab/spades">https://github.com/ablab/spades</a>
<b>Prodigal</b>	2.6.3	-f gff -g 11 -p meta	NA	<a href="https://github.com/hyattpd/Prodigal">https://github.com/hyattpd/Prodigal</a>
<b>CAT</b>	5.2.3	default parameter	CAT database 07/01/2021	<a href="https://github.com/dutilh/CAT">https://github.com/dutilh/CAT</a>
<b>Kaiju</b>	1.7.3	default parameter	nr eukaryote 07/2021	<a href="https://github.com/bioinformatics-centre/kaiju">https://github.com/bioinformatics-centre/kaiju</a>
<b>Diamond</b>	0.9.26	--sensitive --max-target-seqs 1 --outfmt 100 --salltitles -- sallseqid --verbose	nr eukaryote 07/2021	<a href="https://github.com/bbuchfink/diamond">https://github.com/bbuchfink/diamond</a>
<b>Bowtie2</b>	2.4.1	default parameter	NA	<a href="https://github.com/BenLangmead/bowtie2">https://github.com/BenLangmead/bowtie2</a>
<b>samtools</b>	1.9	default parameter	NA	<a href="https://github.com/samtools/samtools">https://github.com/samtools/samtools</a>
<b>CRISPRdetect</b>	2.2	-check_direction 1 - array_quality_score_cutoff 3 -T	NA	<a href="http://crispr.otago.ac.nz/CRISPRDetect/predict_crispr_array.html">http://crispr.otago.ac.nz/CRISPRDetect/predict_crispr_array.html</a>
<b>CRISPRCasFinder</b>	4.2.20	-log -metagenome	NA	<a href="https://github.com/dcouverin/CRISPRCasFinder">https://github.com/dcouverin/CRISPRCasFinder</a>
<b>Metabat2</b>	2-2.12.1	--minContig 5000 --unbinned - -minClsSize 200000	NA	<a href="https://bitbucket.org/berkeleylab/metabat/">https://bitbucket.org/berkeleylab/metabat/</a>
<b>RefineM</b>	0.1.2	Functions: refinem taxon_filter & refinem filter_bins	Genome Taxonomy Database: gtdb_r95_protein_db.2020- 07-30	<a href="https://github.com/dparks1134/RefineM">https://github.com/dparks1134/RefineM</a>

<b>CheckM</b>	1.0.18	lineage_wf -x fa	NA	<a href="https://github.com/Ecogenomics/CheckM">https://github.com/Ecogenomics/CheckM</a>
<b>WISH*</b>	1.1	default parameter	Homemade MAG database	<a href="https://github.com/soedinglab/WISH">https://github.com/soedinglab/WISH</a>
<b>CheckV</b>	0.7.0	checkv_end_to_end	checkv-db-v1.0	<a href="https://bitbucket.org/berkeleylab/checkv/">https://bitbucket.org/berkeleylab/checkv/</a>
<b>VIBRANT</b>	1.2.1	-f nucl -virome	NA	<a href="https://github.com/AnantharamanLab/VIBRANT">https://github.com/AnantharamanLab/VIBRANT</a>
<b>VirSorter2</b>	2-2.2.3	default parameter	NA	<a href="https://github.com/jiarong/VirSorter2">https://github.com/jiarong/VirSorter2</a>
<b>ViromeQC</b>	1.0	default parameter	NA	<a href="https://github.com/SegataLab/viromeqc">https://github.com/SegataLab/viromeqc</a>
<b>SpacePharer</b>	2.8.12	easy-predict	spacers from local MAG databases and CRISPRCasdb (Cf.Mat&Met)	<a href="https://github.com/soedinglab/spacepharer">https://github.com/soedinglab/spacepharer</a>
<b>blastn</b>	2.9.0	-task blastn-short -evaluate 0.003 -word_size 6 -gapopen 10 -gapextend 2 -penalty -1 -max_target_seqs 1000	spacers from local MAG databases and public spacer database (Cf.Mat&Met)	<a href="https://blast.ncbi.nlm.nih.gov/">https://blast.ncbi.nlm.nih.gov/</a>

\* For WISH, the models for the putative hosts were created using the 81 selected MAGs and the default options.

The null parameters were then obtained by using 11944 viral genomes downloaded from NCBI on the 5th and 6th of December, 2020 (Supplementary Table S3), and the R script provided by WISH authors: <https://raw.githubusercontent.com/soedinglab/WISH/master/computeNullParameters.R>

Finally, the prediction on the viral contigs were run as advised by WISH authors and using the file nullParameters.tsv generated at the previous step.

## Supplementary figures

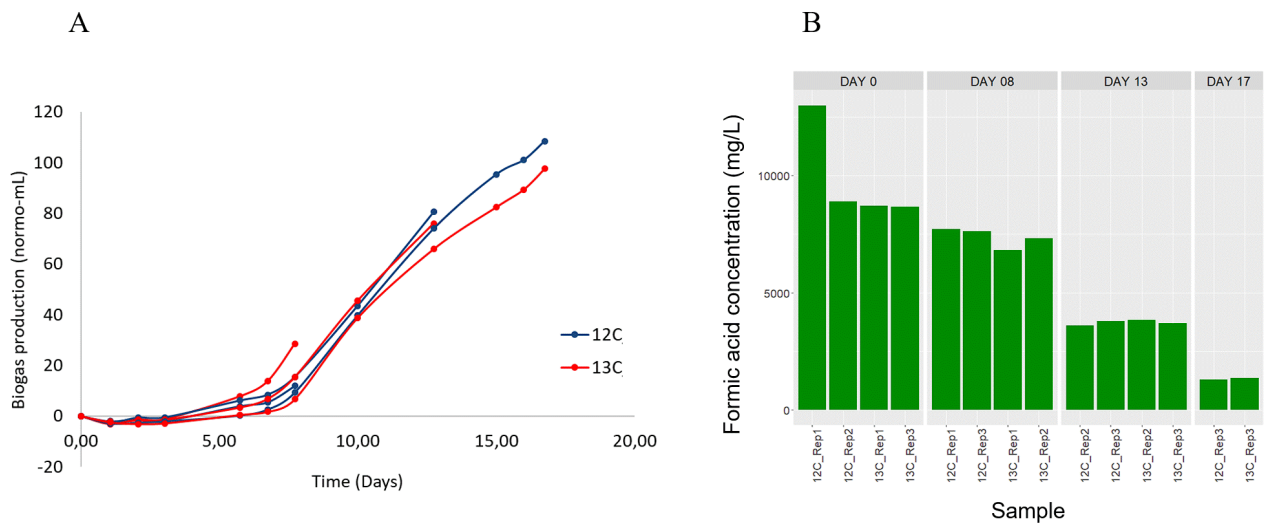


Figure S1. Dynamics of AD process. A: Biogas production during incubation of the six experimental batch AD microcosms, including three fed with unlabeled formate (blue line) and three fed with  $^{13}\text{C}$ -formate (red line).

B: Substrate consumption during incubation.

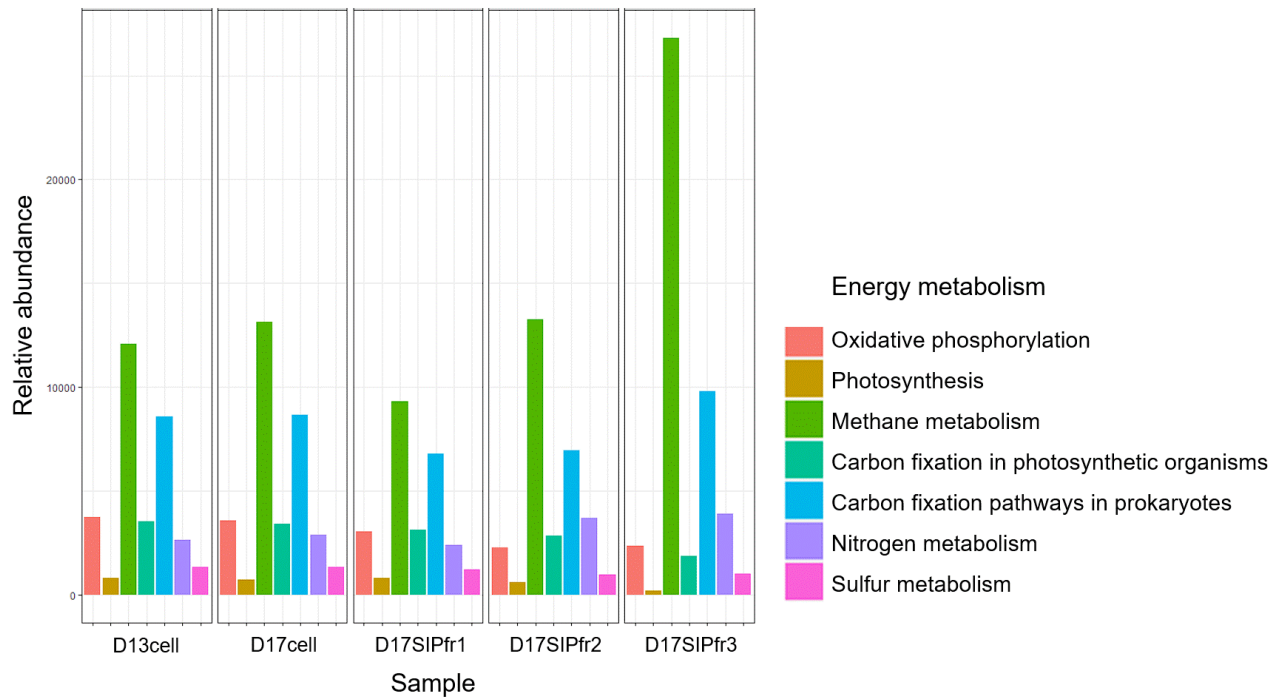


Figure S2. Metabolic pathway analyses of cellular metagenomic data for the category energy metabolism.

The relative abundance is expressed in read counts per million reads (RPM).



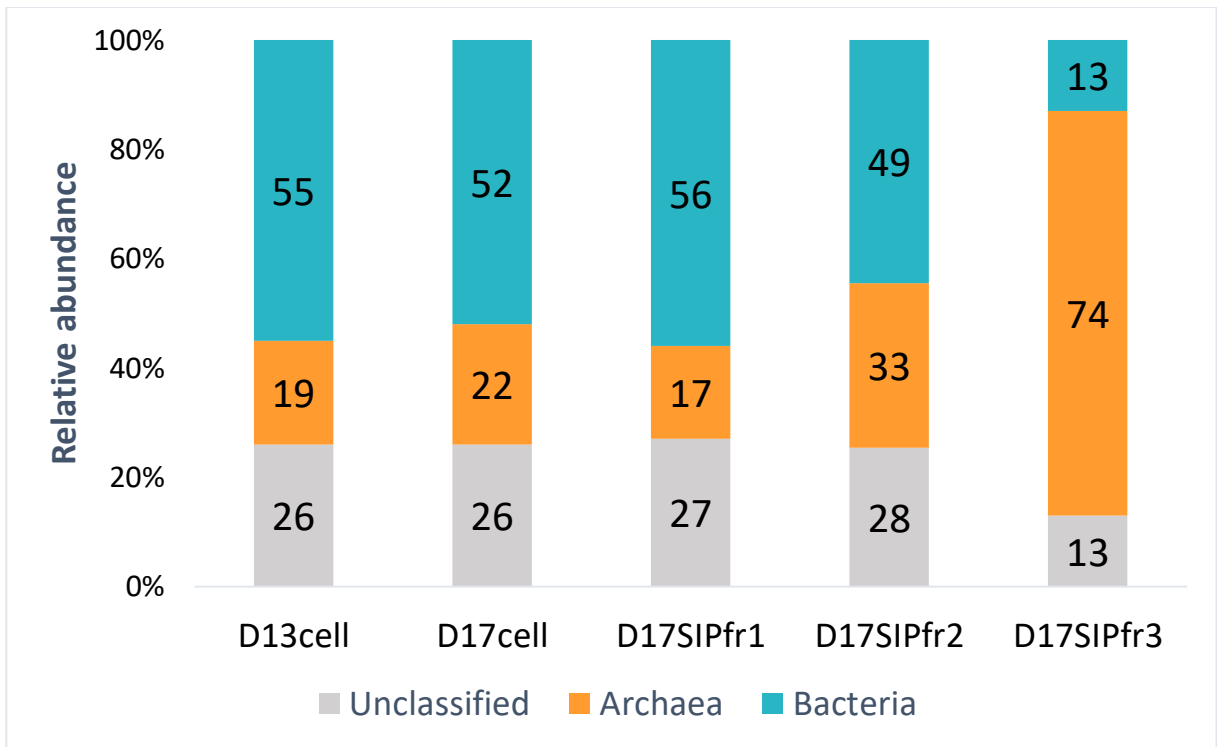


Figure S3. Taxonomic affiliation of the reads from the cellular metagenome dataset

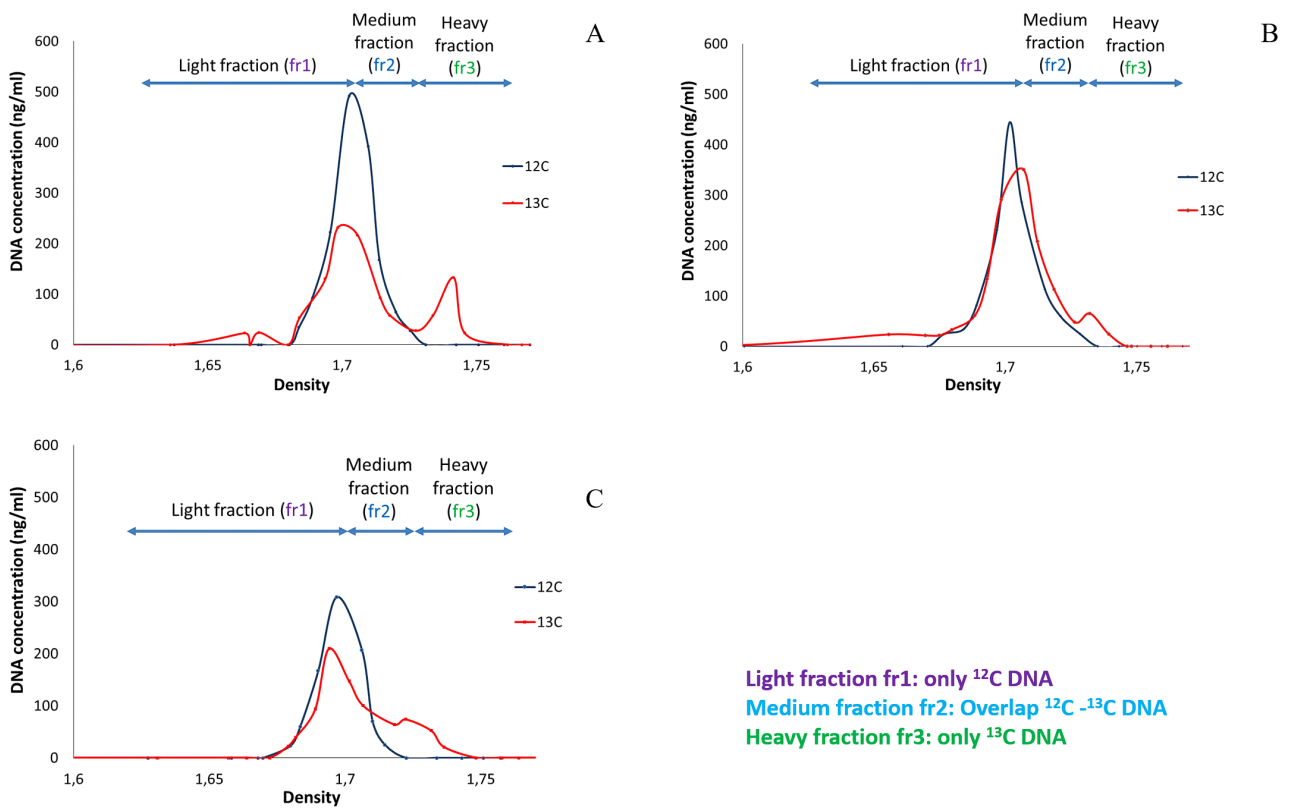


Figure S4. Density profiles of the cellular DNAs collected from microcosms fed either with unlabeled formate (blue) or with <sup>13</sup>C-formate (red), at three different time points. A: day 8, B: day 13 and C: day 17. In the case of <sup>13</sup>C-formate, the DNAs were pooled into 3 fractions based on their mass density in CsCl gradient. Light fraction (fr1) contained only unlabeled DNA whereas heavy fraction (fr3) contained only <sup>13</sup>C-labeled DNA. The medium density fraction (fr2) was a mix of <sup>13</sup>C-labeled and unlabeled DNA.

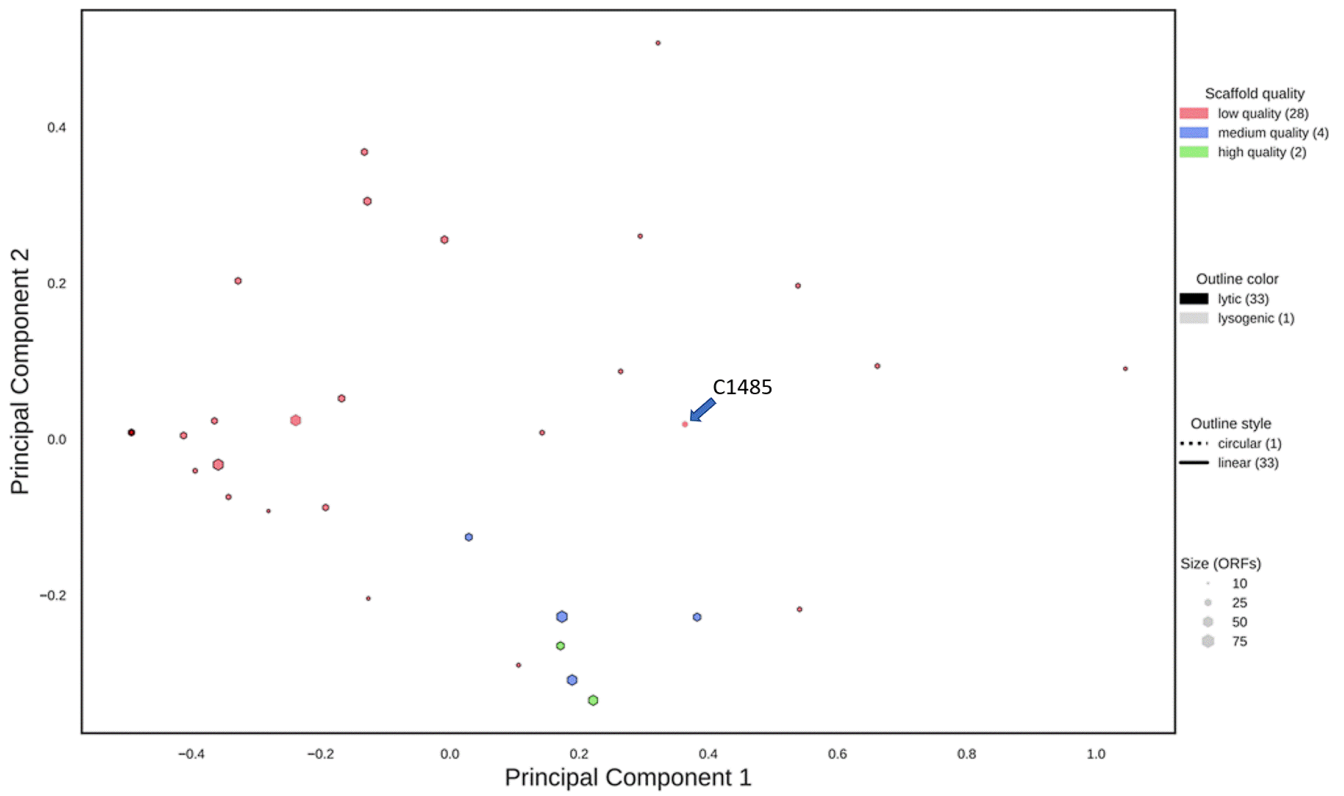


Figure S5. Principal component analysis of the 39 contigs of interest obtained with VIBRANT. Contigs are represented by circles, whose size vary according to their number of Open reading frames (ORFs). The dot color represents the scaffold quality whereas the outline color represents the virus lifestyle.

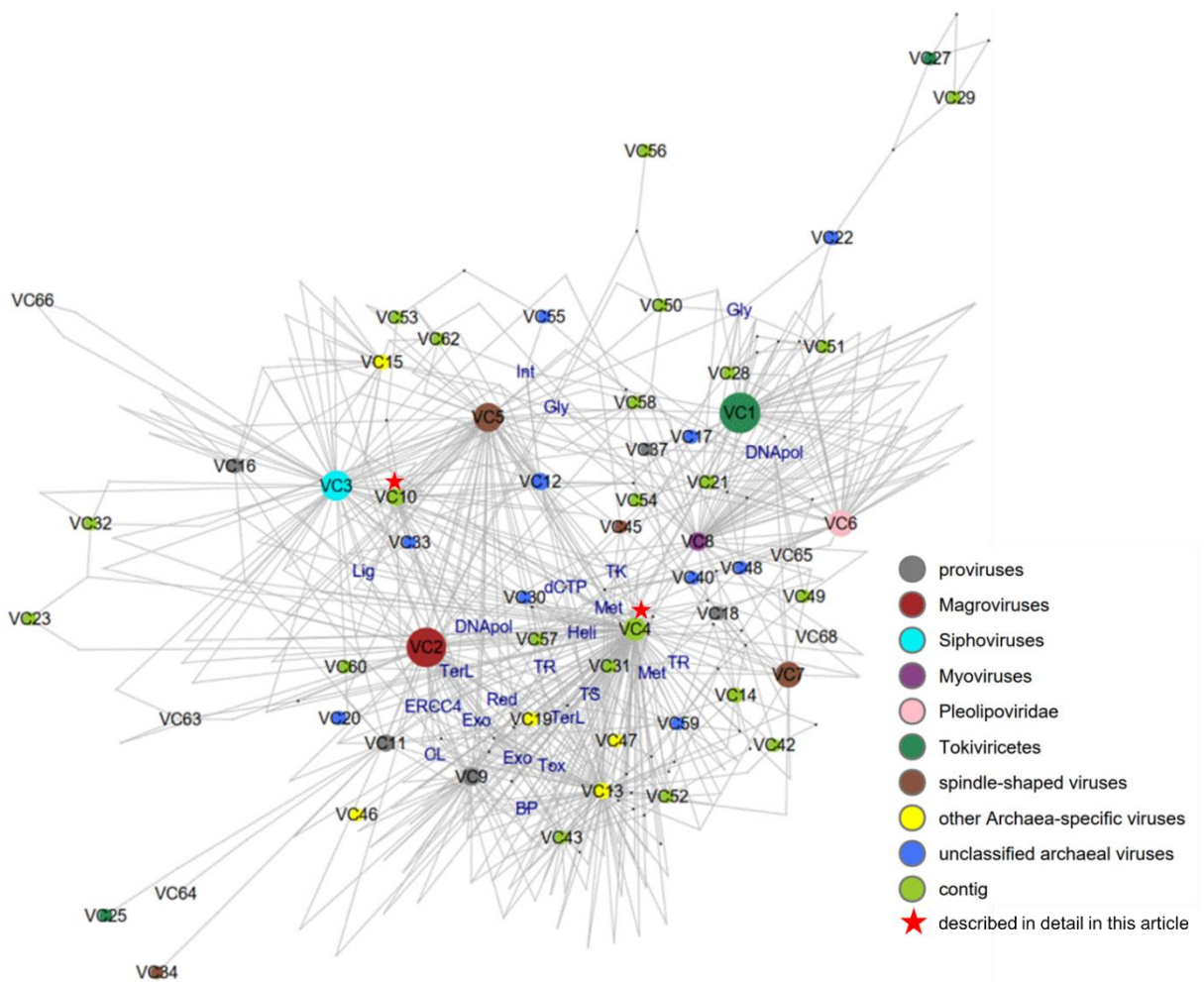


Figure S6. Bipartite network of known archaeal viruses and 39 contigs of interest, grouped into viral clusters (VC), and of protein orthologous groups (OGs). The VCs are represented as circles colored according to the legend shown in the figure, and OGs are denoted by the intersections of edges. The size of each circle is related to the number of members in the corresponding VC. Only OGs shared by at least 2 clusters in which they are conserved in more than 50% of viruses are represented with functional annotation (dCTP: dCTP deaminase, Met: DNA methyltransferase, TerL: terminase large subunit, Lig: ATP-dependent DNA ligase, Exo: DNA repair exonuclease, TR: transcriptional regulator, DNAPol: DNA polymerase, Tox: RelE-like toxin, BP: baseplate protein, DNAPol: DNA polymerase, Exo: exonuclease, Heli: DNA helicase, Met: DNA methyltransferase, Gly: glycosyltransferase, TR: transcriptional regulator, Red: ribonucleoside-diphosphate reductase large subunit, CL: clamp loader of DNA polymerase, ERCC4: ERCC4-type nuclease, TerL: terminase large subunit, TK: thymidylate kinase, Int: integrase, Gly: glycosyltransferase, TS: thymidylate synthase).

## References

- Enright, A.J., Van Dongen, S., and Ouzounis, C.A. (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 30: 1575–1584.
- Liu, Y., Demina, T.A., Roux, S., Aiewsakun, P., Kazlauskas, D., Simmonds, P., et al. (2021) Diversity, taxonomy, and evolution of archaeal viruses of the class Caudoviricetes. *PLOS Biology* 19: e3001442.
- Luo, Y., Pfister, P., Leisinger, T., and Wasserfallen, A. (2001) The Genome of Archaeal Prophage  $\Psi$ M100 Encodes the Lytic Enzyme Responsible for Autolysis of *Methanothermobacter wolfeii*. *Journal of Bacteriology*.
- Meile, L., Jenal, U., Studer, D., Jordan, M., and Leisinger, T. (1989) Characterization of  $\psi$ M1, a virulent phage of *Methanobacterium thermoautotrophicum* Marburg. *Arch Microbiol* 152: 105–110.
- Nölling, J., Groffen, A., and de Vos, W.M. (1993)  $\phi$  F1 and  $\phi$ F3, two novel virulent, archaeal phages infecting different thermophilic strains of the genus *Methanobacterium*. *Microbiology* 139: 2511–2516.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7: 539.
- Söding, J. (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics* 21: 951–960.
- Steinegger, M. and Söding, J. (2017) MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* 35: 1026–1028.
- Thiroux, S., Dupont, S., Nesbø, C.L., Biennu, N., Krupovic, M., L'Haridon, S., et al. (2021) The first head-tailed virus, MFTV1, infecting hyperthermophilic methanogenic deep-sea archaea. *Environmental Microbiology* 23: 3614–3626.
- Weidenbach, K., Nickel, L., Neve, H., Alkhnabashi, O.S., Künzel, S., Kupczok, A., et al. (2017) *Methanosarcina* Spherical Virus, a Novel Archaeal Lytic Virus Targeting *Methanosarcina* Strains. *Journal of Virology* 91: e00955-17.
- Weidenbach, K., Wolf, S., Kupczok, A., Kern, T., Fischer, M.A., Reetz, J., et al. (2021) Characterization of Blf4, an Archaeal Lytic Virus Targeting a Member of the Methanomicrobiales. *Viruses* 13: 1934.
- Whiticar, M.J., Faber, E., and Schoell, M. (1986) Biogenic methane formation in marine and freshwater environments: CO<sub>2</sub> reduction vs. acetate fermentation—Isotope evidence. *Geochimica et Cosmochimica Acta* 50: 693–709.
- Wolf, S., Fischer, M.A., Kupczok, A., Reetz, J., Kern, T., Schmitz, R.A., and Rother, M. (2019) Characterization of the lytic archaeal virus Drs3 infecting *Methanobacterium formicum*. *Arch Virol* 164: 667–674.
- Wood, A.G., Whitman, W.B., and Konisky, J. (1989) Isolation and characterization of an archaeobacterial viruslike particle from *Methanococcus voltae* A3. *Journal of Bacteriology* 171: 93–98.

---

# Chapitre 2 : Virus infectant des microorganismes impliqués dans la méthanisation du formate

Les travaux décrits dans le chapitre 1 ont montré qu'il était possible d'appliquer avec succès la stratégie imaginée, et ils ont apporté des connaissances sur la diversité des virus d'archées (morphologique et génomique) au sein des procédés de digestion, en utilisant le formate comme source unique de carbone. Des nouveaux génomes viraux ont notamment été découverts, en particulier ceux de deux virus fusiformes infectant les archées méthanogènes.

Il a donc été décidé d'élargir le cadre d'application aux bactéries, tout en conservant le formate comme substrat. En effet, comme évoqué en introduction, outre certaines archées méthanogènes, des bactéries peuvent également se développer en utilisant le formate comme source de carbone (voir Synthèse bibliographique, section 2.5). Une proportion non-négligeable de bactéries (environ 30%) a d'ailleurs été observée dans les fractions lourdes (fr3) et de densité intermédiaire (fr2), lors de l'étude présentée au chapitre 1.

En outre, les analyses bioinformatiques effectuées lors de ce premier volet ont présenté quelques limites et pouvaient être optimisées sur certains points. Ceci concernait en particulier la construction des bases de données locales (MAGs, *spacers*). En effet, comme les métagénomés cellulaires avaient été assemblés individuellement, des MAGs redondants étaient présents dans la base de données locale, ce qui rendait la prédiction moins spécifique. Dans ce second volet, la mise en œuvre de co-assemblage a permis de construire une base de données de MAGs moins redondants et de meilleure qualité. Un co-assemblage n'est pas toujours conseillé, notamment lorsque les échantillons sont issus de sources très différentes, mais il était parfaitement adapté dans le cas de mon expérience. Comme le même inoculum et les mêmes conditions avaient été utilisées pour

toutes les incubations, la diversité des communautés microbiennes sélectionnées était très similaire d'un échantillon à l'autre.

La deuxième série d'expériences avait donc pour but de caractériser les virus infectant les formatotrophes, archées ou bactéries, au sein des procédés de digestion anaérobie. Lors de cette série et contrairement à la première série d'expériences, des répliques ont été séquencés (métagénomique *shotgun*) afin de pouvoir évaluer la reproductibilité concernant l'abondance des virus identifiés. Cela a également permis d'optimiser les analyses bioinformatiques, pour une meilleure prédiction de l'hôte des virus.

A nouveau, les résultats de cette série d'expérience ont mené à la découverte de nouveaux génomes de virus infectant des bactéries et des archées impliquées dans le métabolisme du formate. Ces résultats permettent donc d'enrichir nos connaissances sur les communautés microbiennes au sein du processus de digestion anaérobie.

1 **Diversity of novel viruses in anaerobic digestion**  
2 **microcosms fed with formate, discovered through**  
3 **coupling of stable isotope probing and**  
4 **metagenomics**

5 Hoang Ngo<sup>1</sup>, Caroline Talleu<sup>1</sup>, François Enault<sup>2</sup>, Cédric Midoux<sup>1,3,4</sup>, Mahendra

6 Mariadassou<sup>3,4</sup>, Théodore Bouchez<sup>1</sup>, Mart Krupovic<sup>5</sup>, Ariane Bize<sup>1\*</sup>

7

8 <sup>1</sup>Université Paris-Saclay, INRAE, PRocédés biOtechnologiques au Service de  
9 l'Environnement, 92761, Antony, France

10 <sup>2</sup>Université Clermont Auvergne, CNRS, LMGE, F-63000 Clermont-Ferrand, France

11 <sup>3</sup>Université Paris-Saclay, INRAE, MaIAGE, 78350, Jouy-en-Josas, France

12 <sup>4</sup>Université Paris-Saclay, INRAE, BioinfOmics, MIGALE bioinformatics facility, 78350,  
13 Jouy-en-Josas, France

14 <sup>5</sup>Institut Pasteur, Université de Paris, CNRS UMR6047, Archaeal Virology Unit, F-75015  
15 Paris, France

16 \* Correspondence: [ariane.bize@inrae.fr](mailto:ariane.bize@inrae.fr); Tel.: +33(0)1 40 96 60 89

17



18

19 **Abstract**

20 Despite the great viral diversity reported in anaerobic digestion (AD) processes, the  
21 knowledge on these viruses is still limited due to the complexity of the microbial  
22 communities catalyzing the AD process. In this study, we identified viruses infecting  
23 prokaryotes involved in the bioconversion of formate to biogas, by applying an  
24 approach coupling stable isotope probing and metagenomic analyses. With an  
25 experimental system using  $^{13}\text{C}$ - or unlabeled formate as the sole carbon source, we  
26 were able to enrich the microbial communities in formatotrophs, including  
27 methanogenic archaea (Methanobacteriales order) and acetogenic bacteria, with the  
28 dominance of the order Clostridiales. Metagenomic analyses revealed several  
29 previously-uncharacterized viral genomes, leading to the discovery of novel  
30 (pro)viruses infecting members of the genera *Clostridium*, *Methanobacterium*,  
31 *Methanobrevibacter*, and especially, for the first time, a virus infecting the species  
32 *Acetomicrobium* (family Synergistaceae). These results expand greatly the knowledge  
33 on the viral diversity in AD process.

34

## 35 **Introduction**

36 Viruses are ubiquitous and are known as major drivers of several ecosystems. In  
37 anaerobic digestion (AD) plants, a natural biological process applied at the industrial  
38 scale to valorize organic waste or effluents into methane-rich biogas, the great diversity  
39 and abundance of virus-like particles (VLPs) has been reported (Chien *et al.*, 2013;  
40 Calusinska *et al.*, 2016; Willenbücher *et al.*, 2022), suggesting that viruses could have  
41 important effects on the functioning of this process. However, these viruses are still  
42 poorly characterized.

43 In order to decipher host-virus networks in AD ecosystems, we used a simple substrate,  
44 formate, to simplify the microbial community and to focus on the viruses infecting  
45 microorganisms participating to the last steps of anaerobic digestion, acetogenesis and  
46 methanogenesis. During AD, formate can be converted to acetate by  
47 chemolithotrophic acetogenic bacteria (also called homoacetogens), through the  
48 Wood-Ljungdahl pathway, or consumed by some hydrogenotrophic methanogenic  
49 archaea to produce methane (Lemaire *et al.*, 2020), also through the Wood-Ljungdahl  
50 pathway. It can also be used during the autotrophic growth of some sulfate-reducing  
51 bacteria or archaea (Brysch *et al.*, 1987) or methylotrophs through the serine pathway  
52 (Crowther *et al.*, 2008). Formate can also be assimilated by autotrophic growth through  
53 the reductive pentose phosphate pathway.

54 Among these various functional groups, viruses infecting methanogenic archaea have  
55 been the most described, with about 9 viruses isolated so far (Wood *et al.*, 1989; Meile  
56 *et al.*, 1989; Nölling *et al.*, 1993; Wolf *et al.*, 2019; Thiroux *et al.*, 2021; Weidenbach *et al.*,  
57 2021; Pfister *et al.*, 1998; Weidenbach *et al.*, 2017; Krupovič, Forterre, *et al.*, 2010a)  
58 and some proviruses identified in the genomes of methanogenic archaea (Luo *et al.*,  
59 2001; Krupovič and Bamford, 2008; Chibani *et al.*, 2022). Among these viruses, MFTV1  
60 is the sole characterized virus infecting a hyperthermophilic methanogen, isolated from  
61 a deep-sea hydrothermal vent: *Methanocaldococcus fervens* AG86 strain

62 (Methanococcales) (Thiroux *et al.*, 2021). The other viruses of methanogens have been  
63 isolated from AD plant anaerobic sludge samples. Four viruses or virus-like particles  
64 (VLPs) have been isolated from mesophilic digesters: the oblate or spindle-shaped A3  
65 VLPs infecting *Methanococcus* strains (Wood *et al.*, 1989), the spherical virus MetSV  
66 infecting *Methanosarcina* strains (Weidenbach *et al.*, 2017), and the two siphoviruses  
67 Drs3 (family *Anaerodiviridae*) (Wolf *et al.*, 2019; Liu *et al.*, 2021) and Blf4 (unclassified)  
68 (Weidenbach *et al.*, 2021), infecting *Methanobacterium* and *Methanoculleus* strains  
69 respectively. Moreover, four other siphoviruses have also been isolated from  
70 thermophilic digesters:  $\Psi$ M1,  $\Psi$ M2 (family *Leisingerviridae*) infecting  
71 *Methanothermobacter* strains (Meile *et al.*, 1989; Pfister *et al.*, 1998; Liu *et al.*, 2021) and  
72  $\Phi$ F1,  $\Phi$ F3 (unclassified) infecting *Methanobacterium* species (Nölling *et al.*, 1993).  
73 Concerning viruses infecting bacterial formatotrophs, the vast majority of identified  
74 bacteriophages (bacterial viruses) infect hosts belonging to the genus *Clostridium*,  
75 especially *C. perfringens*, *C. botulinum* and *C. difficile* due to the impact of these species  
76 on human and animal health (Hernández and Vives, 2020). The first detected virus of  
77 these species was reported in 1968 (K Inoue and Iida, 1968) and most of the identified  
78 viral genomes and viruses infecting *Clostridium* hosts are lysogenic. The complete  
79 genome of the phage  $\Phi$ CP24R, infecting *C. perfringens* was the first viral genome  
80 defined as virulent for these formatotrophs (Morales *et al.*, 2012). In the context of AD  
81 plants, despite the great viral diversity reported in the literature and the fact that several  
82 viral genomes were predicted as viruses possibly infecting formatotrophs (Tamaki *et*  
83 *al.*, 2012; Calusinska *et al.*, 2016; Willenbücher *et al.*, 2022), these viruses are still poorly  
84 characterized due to the limits of bioinformatic host prediction methods. Hence,  
85 knowledge on these viruses in AD process is still limited.

86 The present study aimed to characterize and identify viruses infecting microorganisms  
87 involved in the bioconversion of formate to biogas within the AD process. Establishing  
88 the direct link between viruses and their host in AD process is challenging due to the  
89 diversity and complexity of the catalytic microbial communities. Here, we relied on a

90 culture-independent approach, by coupling stable isotope probing (SIP), shotgun  
91 metavirome and cellular metagenome sequencing, as well as bioinformatic tools for *in*  
92 *silico* host prediction. SIP is a reference method in environmental microbiology to  
93 identify active microorganisms involved in specific metabolisms, within complex  
94 microbial communities. In our study, AD microcosms were fed with <sup>13</sup>C-labeled formate  
95 to promote the growth of formatotrophs (e.g. through acetogenesis or  
96 methanogenesis). In addition to viral DNA extraction for metavirome sequencing, the  
97 cellular DNAs from the active microorganisms (enriched in <sup>13</sup>C) were sequenced and  
98 used to build host databases, for *in silico* host prediction through alignment-  
99 dependent and signature-based approaches. Those different types of host prediction  
100 methods were employed intentionally to improve the accuracy of host prediction.

101 This original approach led to the discovery of novel genomes of viruses of  
102 formatotrophs, such as *Clostridium* and *Methanobacterium* genera. Especially, we  
103 report for the first time the complete (or nearly complete) genome of a virus infecting  
104 members of the family Synergistaceae (*Acetomicrobium* species), and that of viruses  
105 infecting members of the genus *Clostridium* present in AD process.

## 106 **Experimental Procedures**

### 107 ***AD microcosm experiments and monitoring procedure***

108 AD batch microcosms consisted in glass plasma bottles with a total volume of 2 L.  
109 Either <sup>13</sup>C-labeled-formate or unlabeled formate was employed as the sole carbon  
110 source. A mineral solution suitable for AD (standard NF EN ISO 11734) was also added,  
111 as well as an inoculum (volatile matter content: 12 % of the inoculum wet mass),  
112 prepared from a semi-continuous lab-scale digester fed with biowaste. To obtain the  
113 inoculum, the sampled anaerobic sludge was incubated in anaerobic conditions at 35°C  
114 for 20 days, for the majority of fermentable compounds to decompose. The sludge was  
115 then centrifuged (10,000 g, 20 min, 15°C), aliquoted and stored at -80°C until further  
116 use. The microcosms were hermetically sealed with an aluminum screw cap and a

117 rubber septum. The headspaces were flushed with nitrogen gas (Linde). For each  
118 treatment ( $^{13}\text{C}$ -labeled or unlabeled formate), microcosms were prepared in triplicates,  
119 with a working volume of 1.2 L, using the mineral solution, 12 g of inoculum, and either  
120 0.2 M  $^{13}\text{C}$ -sodium formate (Cortecnet) or 0.2 M unlabeled sodium formate (Sigma  
121 Aldrich) (equivalent to 16.32 g).

122 During incubation, the following physico-chemical parameters were monitored  
123 according to the protocols described in (Puig-Castellví et al., 2022): biogas production,  
124 biogas composition, pH, isotopic composition of the biogas, volatile fatty acids (VFAs),  
125 dissolved inorganic carbon (DIC), dissolved organic carbon (DOC), and chemical  
126 oxygen demand (COD). Furthermore, due to a sharp increase in pH in 2 of the reactors  
127 (13C\_Rep1 and 13\_Rep3), 6 mL and 3,4 mL of HCl 1M were added in respectively at day  
128 17 in order to mitigate the inhibition of biogas production.

129 When the biogas production reached the stationary phase, the incubation was stopped  
130 and the liquid phase was used for cellular and viral DNA extraction. For each  
131 microcosm, the liquid phase was centrifuged (8,000 g, 20 min, 15°C). The pellet was  
132 stored at -80°C for subsequent cellular DNA extraction, and the supernatant  
133 (containing virus-like particles) was collected and stored at 4°C, for virion preparation.

#### 134 ***Preparation and observation of VLPs***

135 The supernatant resulting from the previous step was centrifuged once more (6,000 g,  
136 20 min, 4°C) to further remove microbial cells and debris. The newly-obtained  
137 supernatant was subjected to a three-step filtration with 1.2  $\mu\text{m}$  and 0.8  $\mu\text{m}$  pore-size  
138 acetate cellulose membrane filters (Sartorius, 47 mm diameter), and finally with 0.2  $\mu\text{m}$   
139 polyethersulfone filters (Millipore, 47 mm diameter). The filtrations were realized in  
140 500 mL filtration units (Thermo Scientific) under vacuum. The final filtrate was  
141 centrifuged (40,000 g, 3 h, 4°C) to pellet virus-like particles. The final pellet was  
142 suspended in 2 mL SM buffer (0.1 M NaCl, 0.1 M  $\text{MgSO}_4$ , 0.05 M Tris-HCl, pH 7.5) and  
143 stored at 4°C until subsequent analysis.

144 Virus-like particles were observed by transmission electron microscopy (TEM) at the  
145 platform MIMA2 MET - GABI, INRAE, Agroparistech, 78352 Jouy-en-Josas, France. After  
146 a 10 times dilution in SM buffer, the virion-containing solutions were adsorbed onto a  
147 carbon film membrane on a 300-mesh copper grid, stained with 1% uranyl acetate,  
148 dissolved in distilled water, and dried at room temperature. Grids were examined with  
149 a Hitachi HT7700 electron microscope operated at 80kV (Elexience – France), and  
150 images were acquired with a charge-coupled device camera (AMT).

### 151 ***DNA extraction and quantification***

152 Total cellular DNA was extracted with Qiagen DNeasy PowerSoil Kit, according to the  
153 manufacturer's instructions.

154 For virus-like particles, 360  $\mu\text{L}$  of virion solution were treated with 2U of Turbo DNase  
155 (2 U. $\mu\text{l}^{-1}$ , Ambion), and 22  $\mu\text{L}$  of RNaseA (1 mg/mL, Ambion) for 1 h at 37°C, to remove  
156 contaminant DNA in the liquid phase. Inactivation buffer from the Ambion kit was  
157 added to stop the DNase activity. DNA extraction was then based on a classical Phenol-  
158 Chloroform method (Pickard, 2009), using 1.5 ml Phase lock gel, light (VWR). The  
159 detailed procedure for viral DNA extraction is provided in (Ngo et al., 2022-Submitted  
160 in Environ. Microbiol.).

161 Concentrations of viral and cellular DNA solutions were determined with Qubit dsDNA  
162 HS assay (Invitrogen), according to the manufacturer's instructions.

### 163 ***SIP-DNA***

164 Ultracentrifugation on a CsCl gradient was applied to the total cellular DNAs to  
165 separate them according to their density, as described in (Chapleur et al., 2016). Around  
166 1  $\mu\text{g}$  of DNA was employed for each sample. For each collected fraction, the DNA  
167 quantification was performed with Qubit HS dsDNA (Invitrogen). Based on DNA density  
168 profiles (Fig 1C), the collected DNA was pooled into 3 main fractions: light (fr1:  
169 containing only  $^{12}\text{C}$  DNA), heavy (fr3: containing only  $^{13}\text{C}$  DNA) and intermediate (fr2:

170 containing a mix of <sup>12</sup>C and <sup>13</sup>C DNA). The DNA from each fraction was purified by  
171 Glycogen-PEG 6000 precipitation, according to (Neufeld, Vohra, *et al.*, 2007).

172 The DNA of each heavy and intermediate fraction was amplified with RTG GenomiPhi™  
173 V3 DNA Amplification Kit (Illustra, Cytiva), according to the manufacturer's instructions.  
174 It was then purified by Glycogen-PEG 6000 precipitation, and the concentration of the  
175 collected DNA was determined with Qubit dsDNA HS assay (Invitrogen).

### 176 ***16S rRNA gene metabarcoding***

177 Archaeal and bacterial hypervariable regions V4-V5 of the 16S rRNA gene were  
178 amplified and then sequenced according to the protocol described in (Poirier,  
179 Desmond-Le Quéméner, *et al.*, 2016; Madigou *et al.*, 2019), with some modifications.  
180 The targeted region was amplified by PCR with fusion primers 515F (5'- Ion A adapter-  
181 Barcode-GTGYCAGCMGCCGCGGTA-3') (Wang *et al.*, 2007) and 928R (5'-Ion trP1  
182 adapter-CCCCGYCAATTCMTTTRAGT-3') (Wang and Qian, 2009). The detailed protocol  
183 for library preparation is provided in (Ngo *et al.*, 2022-Submitted in Environ. Microbiol.).  
184 Sequencing was performed on an Ion Torrent Personal Genome Machine using Ion 316  
185 Chip V2 (Life Technologies) and Ion PGM Hi-Q View Sequencing Kit (Life Technologies)  
186 according to the manufacturer's instructions. Sequencing data were processed with the  
187 Torrent Suite Software. All diversity analyses were performed with the R phyloseq  
188 package.

### 189 ***Shotgun metagenomic sequencing***

190 250-500 ng of genomic DNA of each sample were collected for Next generation  
191 sequencing (NGS). Illumina sequencing was performed at the Eurofins Genomics  
192 Europe Sequencing GmbH platform (Ebersberg, Germany). All steps performed (DNA  
193 quantification of the starting material, library preparation, adapter ligation,  
194 amplification and size selection, quality control) have been developed and validated by  
195 the sequencing platform and are based on long-standing experience. Sequencing was  
196 performed on the Illumina NovaSeq 6000 platform using 2x150 Sequence mode.

## 197 ***Metagenomics pipe-line***

198 The most generic steps of our pipeline were scripted as a snakemake workflow  
199 ([https://forgemia.inra.fr/cedric.midoux/workflow\\_metagenomics](https://forgemia.inra.fr/cedric.midoux/workflow_metagenomics)), and applied to both  
200 cellular and viral metagenomes. After a pre-processing step (adapter removal, and  
201 trimming according to quality scores), reads were co-assembled with megahit (Li et al.,  
202 2015). Taxonomic affiliation of the contigs was obtained with CAT (von Meijenfeldt *et*  
203 *al.*, 2019b) and kaiju (Menzel et al., 2016) against NCBI nr database. Coding regions  
204 were predicted with Prodigal (Hyatt et al., 2010). Taxonomic affiliation and functional  
205 annotation of predicted genes were assessed with kaiju (Menzel et al., 2016) and  
206 Diamond (Buchfink et al., 2015) against nr database, respectively. For each dataset, the  
207 trimmed reads were mapped to the assembled contigs and the predicted genes using  
208 Bowtie2 (Langmead and Salzberg, 2012), and then counted with htseq-count (Anders  
209 et al., 2015) and samtools (H. Li *et al.*, 2009b) respectively. Viral genome and provirus  
210 genome detection was performed with VIBRANT (Kieft *et al.*, 2020b) and VirSorter2  
211 (Guo *et al.*, 2021b), and the quality of viral sequences was assessed with CheckV  
212 (Nayfach et al., 2021).

213 Several steps specifically dedicated to viral contig analysis and to host prediction were  
214 performed using homemade bash and python scripts. For the cellular metagenomes,  
215 spacers were detected in contigs with CRISPRdetect (Biswas et al., 2016). A non-  
216 redundant spacer database was built from the obtained spacer sequences. The viral  
217 contigs were subsequently aligned with SpacePHARER (R. Zhang *et al.*, 2021) against  
218 this homemade database and a public spacer database (CRISPRCasdb spacer:  
219 [https://crisprcas.i2bc.paris-saclay.fr/Home/DownloadFile?filename=spacer\\_34.zip](https://crisprcas.i2bc.paris-saclay.fr/Home/DownloadFile?filename=spacer_34.zip)),  
220 enabling host prediction based on an alignment-dependent method. Besides,  
221 metagenome-assembled genomes (MAGs) were constructed from cellular  
222 metagenomic contigs with Metabat2 (Kang *et al.*, 2019b), to obtain a reference MAG  
223 collection for host prediction with WiSH (Galiez *et al.*, 2017), based on k-mer



224 composition (genomic signatures). The MAG quality was improved with RefineM (Parks  
225 *et al.*, 2017) and controlled with CheckM (Parks *et al.*, 2015). Only MAGs with more than  
226 60% completeness and less than 10% contamination were selected. Functional  
227 annotations of the cellular predicted genes were obtained with ghostKoala against  
228 KEGG database (Kanehisa *et al.*, 2016) (<https://www.kegg.jp/ghostkoala/>). Concerning  
229 the viral metagenomes, their quality was assessed with ViromeQC (Zolfo *et al.*, 2019)  
230 using the trimmed reads as entry. The taxonomic affiliation of the viral contigs was  
231 obtained with kaiju against nr-viruses database. After selecting viral contigs of interest  
232 as described in the result section, gene annotation was refined using hhpred on a server  
233 (<https://toolkit.tuebingen.mpg.de/tools/hhpred>) (Zimmermann *et al.*, 2018) with  
234 default parameters, against four structural/domain databases: Pfam-A\_v35,  
235 PDB\_mmCIF70\_12\_Ot\_2021, NCBI\_Conserved\_Domains (CD)\_v3.18 and UniProt-  
236 SwissProt-viral70\_3\_Nov\_2021.

237 The detailed parameters for each tool are presented in Supplementary table S1.

238

239 **Results and discussions**

240 ***Formate converted to methane mainly through hydrogenotrophic***  
241 ***methanogenesis pathway***

242 For the 6 batch AD microcosms fed with unlabeled formate (3 replicates) or <sup>13</sup>C-formate  
243 (3 replicates) as sole carbon source, the biogas production kinetics and the total  
244 cumulated biogas production showed a high level of variability (Figure 1A).  
245 Furthermore, analysis of biogas isotopic composition, which was performed for the  
246 microcosms fed with unlabeled formate, showed the dominance of the acetoclastic  
247 methanogenesis pathway in one replicate (<sup>12</sup>C\_rep1) and of hydrogenotrophic  
248 methanogenesis pathway in the others. The results and detailed calculation of isotopic  
249 signatures are provided in the Supplementary Information 1. There was possibly  
250 competition between several metabolic pathways in the reactors. VFA analysis actually  
251 confirmed the conversion of part of the formate to acetate, in all the reactor, showing  
252 the accumulation of the latter at the end of incubation (Supplementary Figure S1).

253 These results were contradictory with the dominance of the hydrogenotrophic  
254 methanogenesis pathway and the absence of acetate at the end of the incubation,  
255 determined in the previous similar study with the same inoculum but in smaller-scale  
256 microcosms (500 mL instead of 2L) (Ngo et al., 2022-Submitted in Environ. Microbiol.).  
257 This difference might be due to the operating conditions, especially the sampling effect  
258 occurring when adding the inoculum, and the fact that the long-term inoculum  
259 preservation (> 18 months), even if at -80°C, may affect its microbial community  
260 composition through cell mortality. The effect of the labeling did not appear as a  
261 convincing hypothesis considering the biogas production curves, and our previous  
262 study (Ngo et al., 2022-Submitted in Environ. Microbiol.). This lack of effect was  
263 confirmed through statistical analysis with PERMANOVA (~ labeling\_status) being  
264 applied to different relevant datasets (16S rDNA amplicons, cellular metagenomes).  
265 The influence of the labeling was never significant, confirming the lack of major effect

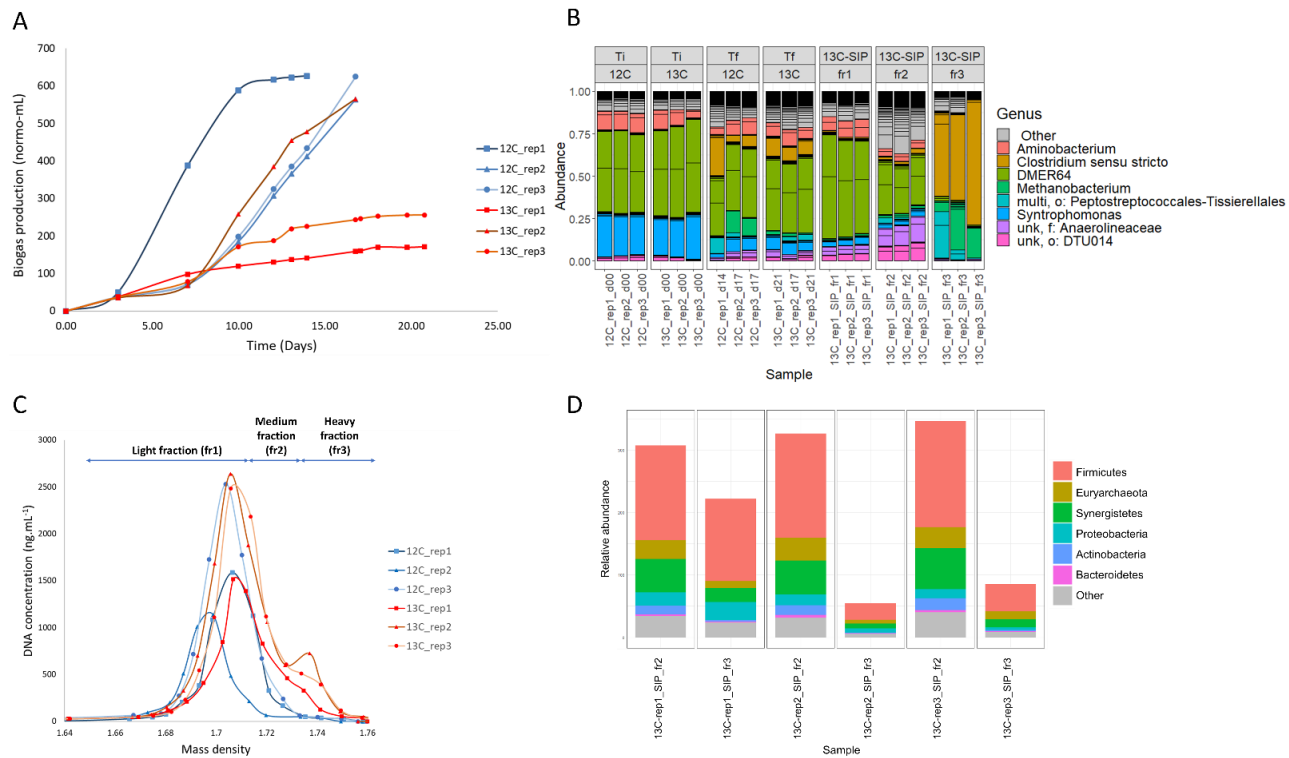
266 of employing <sup>13</sup>C-formate compared to unlabeled formate in our study, on the  
 267 composition of microbial communities. The full detail of the analyses is presented in  
 268 the Supplementary Information 2 and Supplementary Table S2.

269 Hence, the variations observed in the methanogenesis pathway among reactors could  
 270 be primarily explained by the competition between different formate-consuming  
 271 microorganisms (Table 1).

272 [Table 1. Selection of possible reactions for formate consumption in AD process](#)

	Reactions	Microorganisms
<b>Hydrogenotrophic methanogenesis pathway</b>	$4\text{HCOO}^- + \text{H}_2 + \text{H}^+ \rightarrow \text{CH}_4 + 3\text{HCO}_3^-$	Many hydrogenotrophic methanogens
<b>Acetogenesis + Acetoclastic methanogenesis pathway</b>	$\text{HCOO}^- + \text{CO}_2 + 3\text{H}_2 \rightarrow \text{CH}_3\text{COO}^- + 2\text{H}_2\text{O}$	Many acetogenic bacteria
	$\text{CH}_3\text{COO}^- + \text{H}_2\text{O} \rightarrow \text{CH}_4 + \text{HCO}_3^-$	Acetoclastic methanogens

273



274

275

276

277

278

279

280

Figure 1. Microbial dynamics during AD incubation. A: Cumulative biogas production of all microcosms. B: Microbial community composition of archaea and bacteria at the family level, based on 16S rRNA gene metabarcoding. C: Density profiles of the cellular DNAs obtained from microcosms fed with unlabeled formate (blue) or with  $^{13}\text{C}$ -formate (red) at the end of the incubation. The arrows labeled as fr1, fr2 and fr3 indicate how the collected DNA were pooled in the case of labeled formate. fr1 contained only unlabeled DNA whereas fr3 contained only  $^{13}\text{C}$ -labeled DNA. fr2 likely contained a mix of  $^{13}\text{C}$ -labeled and unlabeled DNA. D: Taxonomic assignation (Phylum level) of the microorganisms involved in the methane metabolism, based on the functional analysis of the cellular shotgun metagenomes.

281 ***Enrichment in acetogenic bacteria in competition with hydrogenotrophic***  
282 ***methanogens, over time***

283 To evaluate the dynamics of the microbial communities, 16S rRNA gene metabarcoding  
284 was performed (Figure 1B). The competition among formatotrophs was again  
285 confirmed by the trends of relative abundance of acetogens and hydrogenotrophic  
286 methanogens, especially with the two following groups: *Clostridium sensu stricto* and  
287 Methanobacteriaceae family. In fact, the relative abundance of *Clostridium sensu*  
288 *stricto* was higher when that of Methanobacteriaceae was low, and reciprocally.  
289 Moreover, an increase in proportions of other acetogens or syntrophic bacteria (Liang  
290 et al., 2015; Si et al., 2016; He et al., 2018), was also observed, namely members of the  
291 families Anaerolineaceae and Synergistaceae.

292 The SIP-DNA method was applied to the six microcosms in order to identify the  
293 microorganisms actively involved in the bioconversion of formate to biogas. The  
294 profiles of cellular DNA concentrations in function of their mass density were  
295 established for the 6 microcosms (Figure 1C). For those fed with with <sup>13</sup>C-formate (red  
296 lines) a second peak (or a bump) was observed comparing to those fed with unlabeled  
297 formate (blue lines), suggesting that the formate was consumed only by certain  
298 subpopulations of the microbial community. We therefore decided to pool the DNA  
299 into 3 distinct fractions, following the protocol described in our previous work (Ngo et  
300 al., 2022-Submitted in Environ. Microbiol.): "The first fraction (fr1, density < 1.705)  
301 corresponded exclusively to non-labeled DNA. The second fraction (fr2, 1.705 < density  
302 < 1.738) possibly contained a mix of unlabeled and <sup>13</sup>C-labeled DNA. Finally, the third  
303 fraction (fr3, density > 1.738) contained exclusively <sup>13</sup>C-labeled DNA". The 16S rRNA  
304 gene metabarcoding applied to these fractions showed the efficiency of the SIP  
305 approach with a strong selection of specific microorganisms in fractions fr2 and fr3. In  
306 particular, fr3 was dominated by members of *Clostridium sensu stricto* and  
307 Methanobacteriaceae families, supporting the enrichment of formatotrophs.

308 Chemical analysis and 16S rRNA gene metabarcoding results showed that acetogens  
309 were enriched overtime and were responsible for acetate production. However, there  
310 was no sign or a very low level of enrichment of acetoclastic methanogens such as  
311 members of Methanosarcinaceae or Methanomicrobiaceae families, which were also  
312 present at the beginning of the incubation. This phenomenon could be explained by  
313 different hypotheses. First, it can be related to the batch conditions: the bioconversion  
314 of formate resulted in an increase of pH, from  $\approx 7.50$  at the beginning of the incubation  
315 to  $\approx 9.50$  at the end. pH is known as a major factor influencing the functioning of the  
316 AD process (Wang et al., 2014; Zhai et al., 2015; Eryildiz et al., 2020). Secondly, this  
317 observation could also partly result from amplification biases.

318 In order to verify this absence of acetoclastic methanogens and also to prepare home-  
319 made specific databases for host prediction, DNA collected from "intermediate"  
320 fractions fr2 and "heavy" fractions fr3 were selected for NGS sequencing  
321 (13C\_rep1\_SIP\_fr2, 13C\_rep1\_SIP\_fr3, 13C\_rep2\_SIP\_fr2, 13C\_rep2\_SIP\_fr3,  
322 13C\_rep3\_SIP\_fr2 and 13C\_rep3\_SIP\_fr3). Taxonomic affiliation of the trimmed reads  
323 revealed the presence of acetoclastic methanogens in all fr2 fractions, and their  
324 proportion was significant among the archaea represented in these fractions ( $\approx 20$ -  
325  $40\%$ ). However, the total proportion of archaea in this fraction was low (less than  $12\%$   
326 of the total number of reads). Acetoclastic methanogens were likely active during the  
327 incubation but they were outcompeted by hydrogenotrophic methanogens, due to  
328 their slower growth rate (Thauer et al., 2008). Regarding active bacteria, members of  
329 Firmicutes phylum were highly dominant in all the heavy fractions, indicating a result  
330 similar to the 16S rRNA gene metabarcoding analysis. Moreover, metabolic pathway  
331 analyses also confirmed the dominance of Firmicutes members in all the energy  
332 metabolism categories (Supplementary figure S2), including methane metabolism  
333 (Figure 1D).

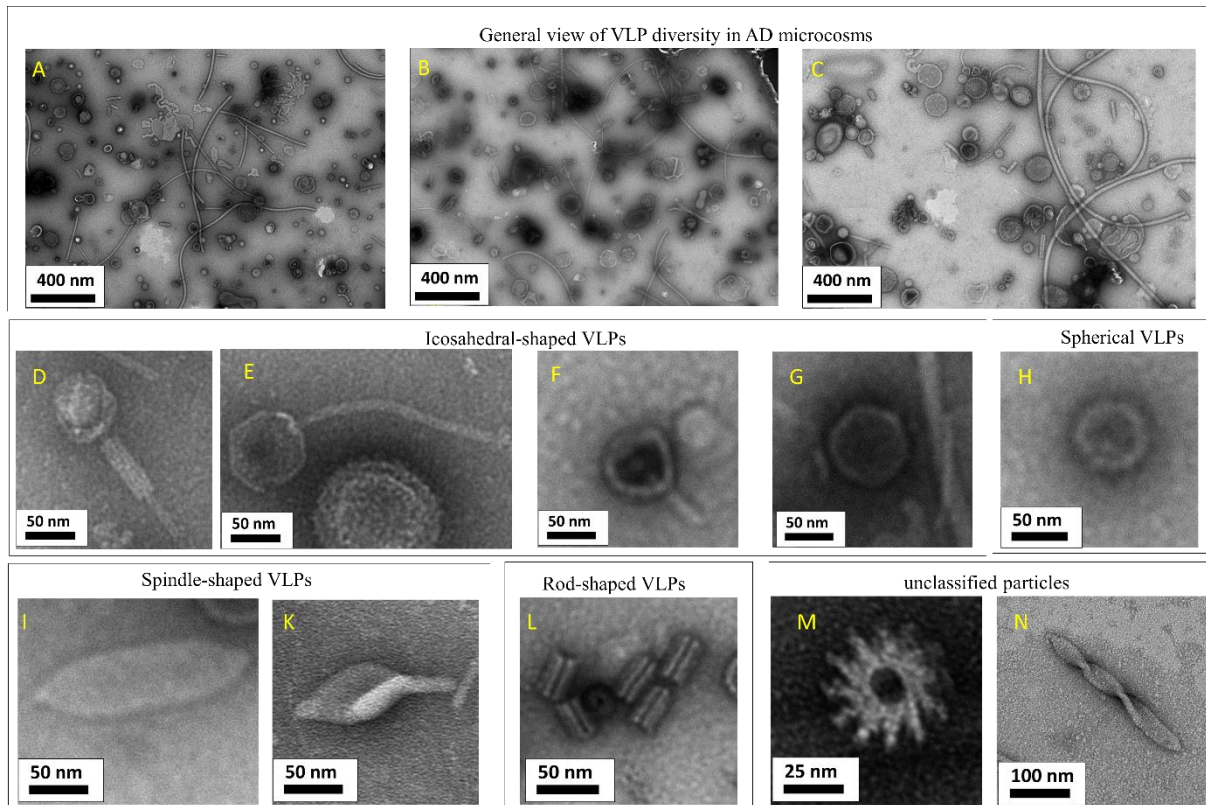
334 **Great diversity of VLPs in AD process revealed through TEM observation**

335 A great diversity of virus-like particles was observed in all the microcosms (Figure 2A-  
336 C), despite the variability of the biogas production. Cosmopolitan morphotypes  
337 common to the domains Bacteria and Archaea, including VLPs with a head-tailed  
338 morphology, were the most abundant in all samples. They presented a large diversity,  
339 with myovirus-like (i.e., icosahedral capsids with contractile tails; Figure 2D), siphovirus-  
340 like (i.e., icosahedral capsids with long non-contractile tails; Figure 2E) and also  
341 podovirus-like (i.e., icosahedral capsids with short tails; Figure 2F) morphotypes.  
342 Icosahedral tailless particles (Figure 2G) were also represented. They could originate  
343 from tailless icosahedral viruses or from head-tailed viruses with a broken tail.

344 Beside the dominance of cosmopolitan morphotype VLPs, uncommon viral  
345 morphotypes were detected, such as spherical (Figure 2H) and especially spindle-  
346 shaped (Figure 2I-K), the latter structure being specific to archaeal viruses. Spindle-  
347 shaped viruses have been reported in extreme geothermal and hypersaline  
348 environments (Krupovic et al., 2014), in freshwater, marine habitats (Borrel et al., 2012;  
349 Kim et al., 2019) and also in AD plants (Calusinska et al., 2016), suggesting that the  
350 hosts of these viruses could possibly be involved in the AD process. Interestingly, our  
351 previous study revealed the detection of two new families of spindle-shaped viruses  
352 infecting methanogens (*Methanobacterium* and *Methanosarcina* species) (Ngo et al.,  
353 2022-Submitted in Environ. Microbiol.). Moreover, rod-shaped particles were observed  
354 in all samples (Figure 2L). However, their length was very low ( $\approx 50$  nm) compared to  
355 that of known rudiviruses ( $\approx 400$ -800 nm). These latter are a family of rod-shaped  
356 viruses infecting some hyperthermophilic archaea. The observation of similar particles  
357 was also reported in several AD plants (Calusinska et al., 2016; Robledo-Mahón et al.,  
358 2019).

359 Particles with unique morphotypes were also identified (Figure 2M-N). Some of them  
360 appeared as round-shaped and less than 50 nm in diameter, with a hollow star-like

361 structure (Figure 2M) and they were also reported in our previous study (Ngo et al.,  
362 2022-Submitted in Environ. Microbiol.). Some other particles with a remarkable slightly  
363 twisted structure were detected (size of 200-300 nm) (Figure 2N). However, the nature  
364 of all these particles remains unclear.



365

366 [Figure 2. Morphotypic diversity of VLPs in different microcosms, observed by TEM.](#)

367

### 368 **Overview of the *shotgun* metagenome datasets**

369 Beside 6 cellular metagenomes mentioned in the previous section, shotgun  
370 metagenomic sequencing was also applied to the total viral DNA extracted at the end  
371 of the incubation for all the microcosms (total of 6 metaviromes: 12C\_rep1\_V,  
372 12C\_rep2\_V, 12C\_rep3\_V, 13C\_rep1\_V, 13C\_rep2\_V and 13C\_rep3\_V). The purpose was  
373 primarily to analyze the metaviromes, and to use the cellular metagenomes to build  
374 specific databases for host prediction. In addition, viral genome detection was also  
375 applied to cellular metagenomes in order to identify proviruses.



376 Based on classical metrics (Table 2), the sequencing data showed satisfactory results in  
377 terms of sequencing depth, read quality and assembly quality. For the cellular  
378 metagenomes, more than 97% of the contigs obtained a taxonomic affiliation.  
379 Furthermore, the phyla Firmicutes and Bacteroidetes dominated among these affiliated  
380 contigs (Figure 3A), showing a good agreement with the 16S metabarcoding result, for  
381 the SIP DNA fractions fr2 and fr3. These first analyses thus suggested that the cellular  
382 contigs would constitute a good-quality database of host sequences. Metagenome-  
383 assembled genomes (MAGs) were constructed by binning of these contigs: in total, 114  
384 MAGs (completeness >60% and contamination <5%) were selected as reference hosts,  
385 including 9 assigned to the domain Archaea. In order to elucidate the presence of viral  
386 sequences (e.g. proviruses) in the cellular DNA sequences, CheckV, VIBRANT and  
387 VirSorter2 were applied to the contigs and these 3 tools identified 3260, 1440 and 6043  
388 viral genomes, respectively, representing 7953 distinct viral genomes in total. All these  
389 tools rely on protein similarity profiles and Virsorter2 revealed a higher number of viral  
390 genomes thanks to its algorithm adapted to the detection of short viral genome (Guo  
391 *et al.*, 2021b).

392

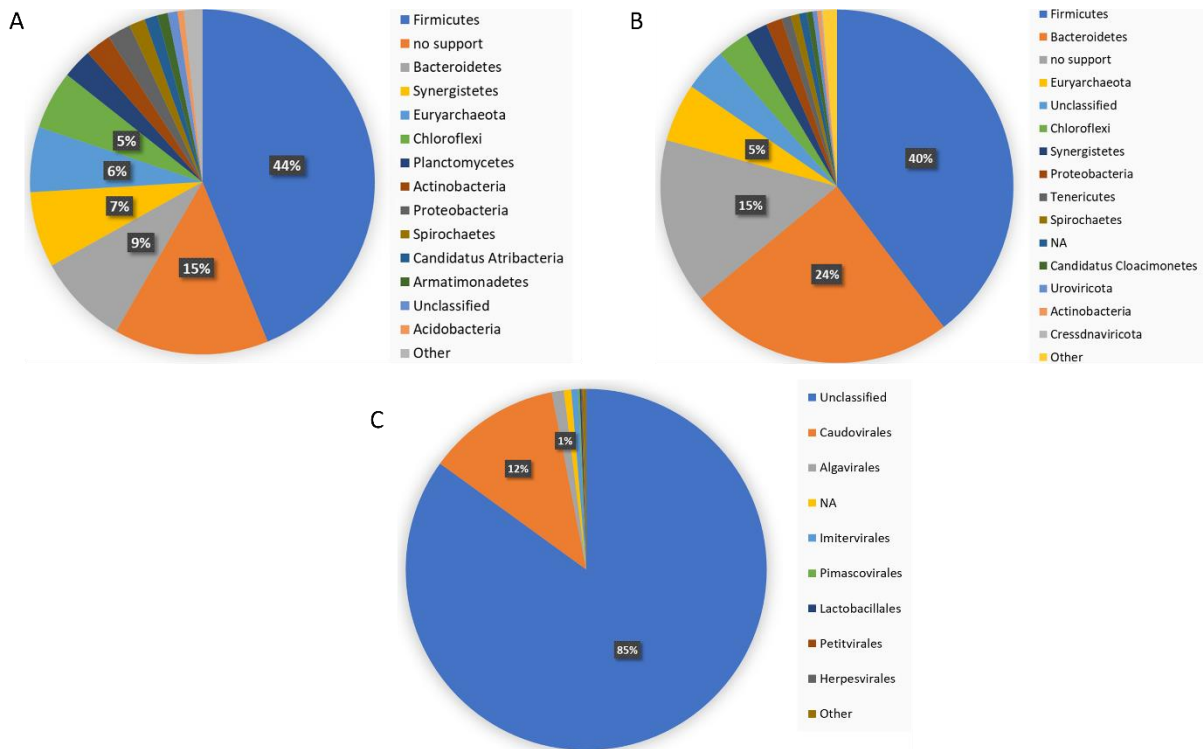
393

Table.2. General overview of the shotgun metagenome datasets.

	Cellular metagenomes (coassembly)						Viral metagenomes (coassembly)					
	13C_rep1_SIP_fr2	13C_rep1_SIP_fr3	13C_rep2_SIP_fr2	13C_rep2_SIP_fr3	13C_rep3_SIP_fr2	13C_rep3_SIP_fr3	12C_rep1_V	12C_rep2_V	12C_rep3_V	13C_rep1_V	13C_rep2_V	13C_rep3_V
Number of raw reads (millions)	82.31	79.71	83.30	97.70	122.17	101.67	54.69	56.24	41.76	45.15	55.42	58.21
Reads obtained after trimming (%)	97.74	98.37	97.98	98.28	98.48	98.41	98.71	98.20	97.92	98.18	98.45	98.55
Number of contig (≥1kb)	144832						56613					
Number of contig (≥3kb)	38658						12535					
Max contig length (b)	650103						425914					
Contigs with taxonomic affiliation (%)	97.24						92.17					
Reads mapped to contigs (%)	91.69	93.99	92.12	94.62	92.60	94.68	89.04	86.20	87.14	88.13	89.07	89.02
Number of MAGs	192						-					
Number of selected MAGs (≥60 % completeness & ≤5 % contamination)	114						-					
Number of archaeal MAGs (in selected MAGs)	9						-					

394

395 Concerning the metavirome datasets, quality check with ViromeQC showed that most  
396 of the metaviromes had a moderate contamination rate (less than 8%), except for one  
397 sample (12C\_Rep3\_V), with a contamination rate of 76%. The contamination rate  
398 provided by ViromeQC takes into account a set of 31 single-copy markers and the small  
399 and large subunits of the rRNA gene (Zolfo et al., 2019). Focusing only on rRNA gene  
400 subunits, which are generally used for prokaryote identification, the contamination rate  
401 of all the samples was low ( $1.0e-6$ ). The high contamination rate of 12C\_Rep3\_V  
402 metavirome might be explained by the presence of membrane vesicles revealed by  
403 TEM, known as a challenge for virion preparation. Among 56 613 assembled contigs  
404 longer than 1kb, more than 92% obtained a taxonomic affiliation against NCBI nr  
405 database and 95.04% were affiliated to prokaryotes whereas only 15% of contigs had  
406 a taxonomic affiliation against nr-viruses database, a variant of NCBI nr database  
407 impoverished in cellular sequences (Figure 3B-C). This observation can be explained by  
408 the dominance of prokaryotes in the NCBI nr database used for the taxonomic  
409 assignation, and by the fact that the best match of viral genes can also be found in  
410 prokaryotic genomes, in particular in proviruses. Although these biases are a hurdle for  
411 the identification and classification of viruses, hits in prokaryotic genomes can  
412 nevertheless bring information on the possible hosts of the viral contigs. Contigs  
413 affiliated using NCBI nr database actually revealed the dominance of the phyla  
414 Firmicutes and Bacteroidetes (Figure 3B), again in good agreement with the 16S rDNA  
415 metabarcoding and cellular metagenomics results. This suggested the promising  
416 discovery of viruses infecting members of these phyla. Moreover, most of the contigs  
417 annotated using nr-viruses database belonged to the class Caudoviricetes (Figure 3C),  
418 a class of viruses with a cosmopolitan head-tailed morphology, representing more than  
419 96% of the bacteriophages in general. This dominance of the Caudoviricetes class was  
420 in good agreement with the VLP observation by TEM (Figure 2).



421  
 422 Figure 3. Taxonomic affiliation of the various datasets. A: Cellular metagenome coassembled contigs  
 423 annotated by CAT against NCBI nr database. B: Metaviriome coassembled contigs annotated by CAT  
 424 against NCBI nr database. C: Metaviriome coassembled contigs annotated by Kaiju against NCBI nr-Vir  
 425 database. The colors represent the 15 most abundant phyla in A and B, and the 10 most abundant viral  
 426 orders in C.

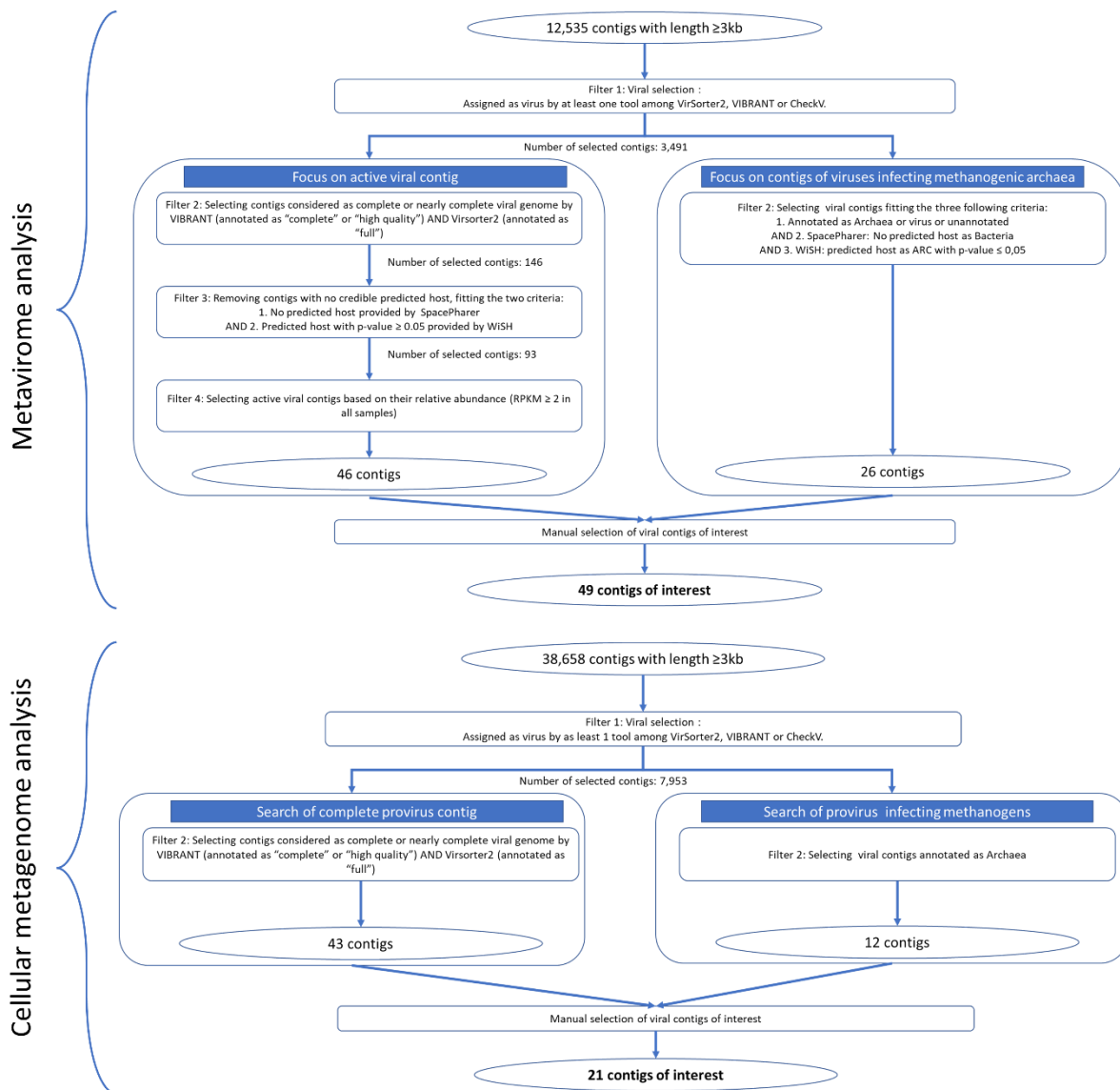
427 For host prediction, in order to improve the accuracy, two complementary approaches  
 428 (based on spacer alignment and on genome signatures) were implemented.  
 429 SpacePharer is an alignment-dependent tool based on the matching of CRISPR spacers  
 430 to the viral contigs. The CRISPR-Cas system is an adaptive immunity system identified  
 431 in 85% of Archaea and 40% of Bacteria (Makarova et al., 2020). During a viral infection,  
 432 a short sequence (26 to 72 bp) of the infecting viral genome, named "spacer" is  
 433 sometimes inserted into the CRISPR array of the host (Edwards et al., 2016). These  
 434 spacers are considered as proof the past host-virus interactions. For contigs longer  
 435 than 3 kb, this method revealed a total of 417 contigs with a predicted host. WISH  
 436 (Galiez et al., 2017) was used for the approach relying on genomic signatures: most  
 437 prokaryotic viruses and their hosts have similar oligonucleotide profile (k-mer), due to

438 their co-evolution over long time periods. Using WIsH for the viral contigs, against our  
439 home-made MAG database, yielded an acceptable host ( $p$ -value  $\leq 0.05$ ) for 8302  
440 contigs longer than 3 kb. Pooling the results of both methods, host prediction was  
441 obtained for a total of 8515 distinct contigs, which confirmed their complementarity to  
442 improve in silico host prediction.

443 ***Focus on contigs likely originating from viruses infecting formatotrophic***  
444 ***microorganisms***

445 Only contigs longer than 3kb were selected in order to eliminate short contigs  
446 containing limited information. Independent analyses were performed to focus on  
447 viruses from formatotrophic bacteria on the one hand and on viruses infecting  
448 methanogenic archaea on the other, detected among the metavirome contigs or the  
449 cellular ones (proviruses) (Figure 4).

450 Concerning the analysis of viruses infecting formatotrophic bacteria, a filter based on  
451 the scaffold quality provided by VIBRANT and VirSorter2 was applied to focus only on  
452 complete or nearly complete viral contigs. Considering that these tools were developed  
453 based on viral databases overwhelmingly dominated by bacterial *Caudoviricetes*, this  
454 strategy is expected to result mostly in the selection of caudoviruses, which are known  
455 to be dominant among bacterial viruses. For contigs from the metavirome dataset, 2  
456 additional filters were applied to refine the selection (Figure 4). The first one aimed to  
457 remove contigs with unreliable host prediction, and the subsequent one to select  
458 contigs from viruses infecting metabolically active microbial community members.



459

460

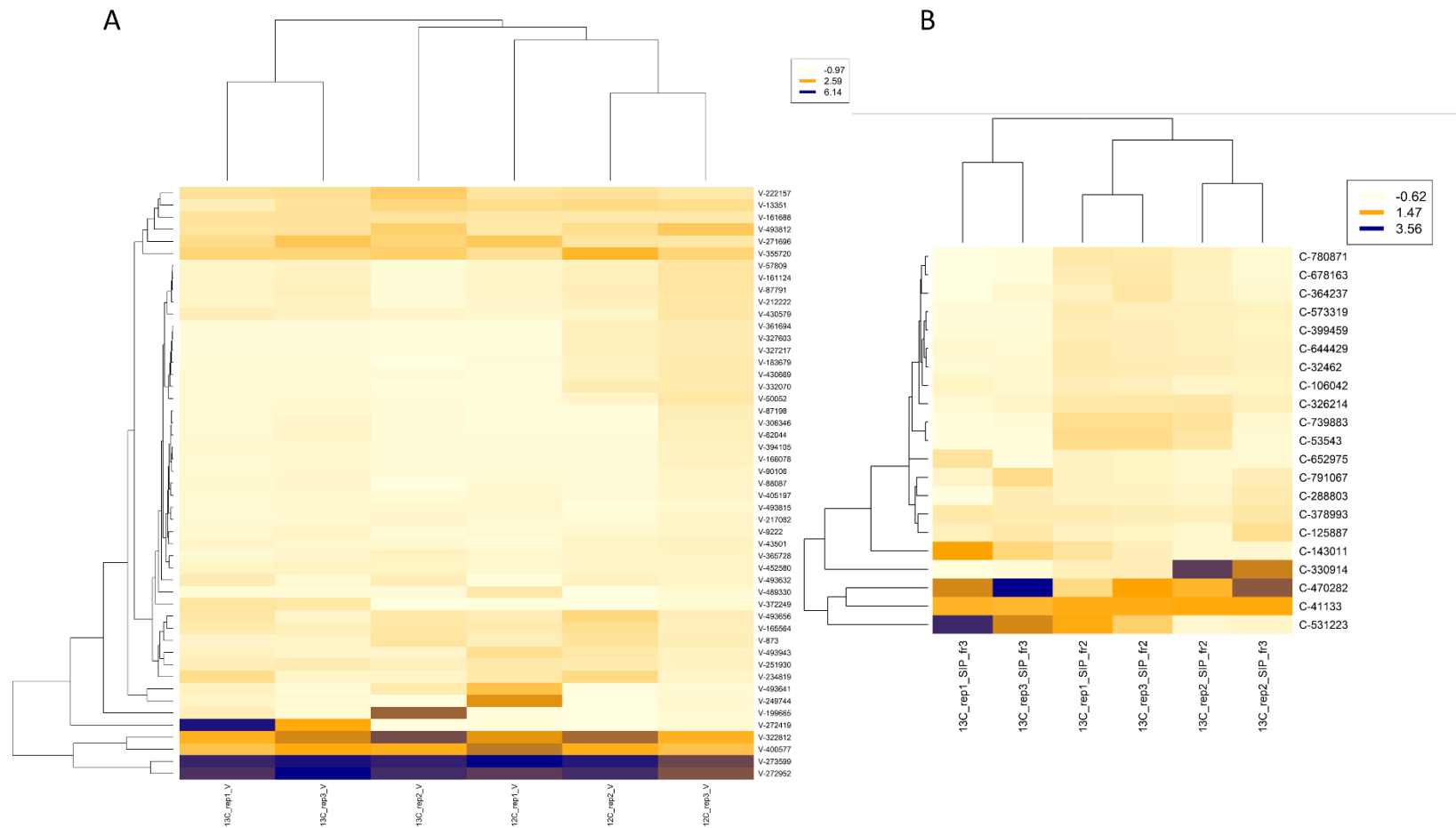
Figure 4. Strategy for the selection of contigs of interest

461 Regarding the analysis of viruses infecting methanogenic archaea, the filter based on  
 462 scaffold quality was not applied. As stated above, these tools are expected to be well-  
 463 adapted to bacterial caudoviruses, while their accuracy on datasets including novel  
 464 archaeal viruses, especially those with small or medium-sized genomes, is not to be  
 465 expected. The first filter based on the taxonomic annotation of genes aimed to remove  
 466 cellular contaminants. The second one aimed to remove contigs from bacterial viruses.  
 467 Finally, the contigs preselected through each analysis were examined manually to  
 468 retain only those with a high-confidence prediction of the host. To sum up, 49 viral

469 contigs (names starting with V-) and 21 putative proviruses (names starting with C-)  
470 were selected in the metavirome and the cellular metagenomes, respectively.  
471 For each of these contig sets, a heatmap (Figure 5) was prepared to compare the contig  
472 distribution profile of each microcosm.

473 Considering the 49 archaeal and bacterial viral contigs of interest, 4 viral contigs,  
474 forming a cluster, had a rather uniform abundance across the 6 metaviromes and  
475 showed among the highest abundance values: V\_322812, V\_400577, V\_273599 and  
476 V\_272952 (Figure 5A). Six viral contigs also had a uniform abundance across the 6  
477 metaviromes, but with more intermediate abundance values: V\_222157, V\_13351,  
478 V\_161688, V\_493812, V\_271696, V\_355720. Most other contigs had overall lower  
479 abundances and a rather even distribution across the metaviromes. Finally, a few viral  
480 contigs clearly showed uneven distributions, being more abundant in one (V\_493641,  
481 V\_249744, V\_199665) or two (V\_272419) metaviromes compared to the others. For the  
482 21 putative proviruses (Figure 5B), the abundances were overall moderate and variable.  
483 These observations may be related with the uneven distribution of microorganisms  
484 highlighted with the 16S rRNA gene and cellular metagenomic analyses.

485 Among these two sets of 49 and 21 viral contigs, some had a predicted host from the  
486 order Methanobacteriales (*Methanobacterium* and *Methanobrevibacter* genera,  
487 hydrogenotrophic methanogens). Concerning bacterial viral genomes, most of the  
488 predicted hosts belonged to one of the three phyla Firmicutes, Bacteroidetes and  
489 Synergistetes. In the next section, we focused more specifically on a selection of 6 viral  
490 contigs whose host was predicted with a very high confidence (showing agreement  
491 between the different prediction tools) (Figure 6). Three were predicted to infect  
492 methanogens, and the three others to infect bacteria.



493

494

495

496

Figure 5. Heatmaps of the viral contigs of interest. The abundance values in RPKM were further normalized by centered-log ratio (CLR) transformation. The distance matrix was calculated with the Euclidean distance, and the clustering was performed with the method "average". A: 49 viral contigs of interest from metaviromes. B: 21 provirus contigs of interest from cellular metagenomes.



497 ***Viruses infecting methanogenic archaea in the formate incubation microcosms***

498 Contig V-322812 represents the partial genome (21.6 kb) of a virus of methanogen. Its  
499 relatively high abundance in the microcosms (RPKM values between 15 and 225) is  
500 consistent with the significant detection of methanogenic archaea in the heavy DNA  
501 fractions (fr3) (Fig 1B). Its taxonomic affiliation was unclear at a fine level, kaiju  
502 predicting *Candidatus Methanofastidiosum methylthiophilus* and WIsH, the family  
503 Methanobacteriaceae. Although a fine taxonomic affiliation of the host seemed difficult  
504 to establish, the prediction results still pointed to methanogenic archaea. Moreover,  
505 the contig encodes a MCM helicase, previously identified in several archaeal virus and  
506 plasmid families (Krupovič, Gribaldo, *et al.*, 2010b), and suggesting a circular genome  
507 topology. A subunit of an archaeal DNA polymerase II was also encoded in V-322812,  
508 further supporting the link with Archaea. Interestingly, this contig also encodes a  
509 protein shared with *Halorubrum* pleomorphic virus 1, a pleiomorphic virus infecting  
510 halophilic archaea. The only structural genes detected code for tail proteins, suggesting  
511 a possible head-tail morphotype. Considering the lack of encoded integrase, this virus  
512 could be purely virulent (lytic), however this remains hypothetical since the genome is  
513 not complete.

514

515

516 Contig C-41133 is the partial genome (31.8 kb) of a virus predicted to infect  
517 *Methanobacterium* archaea (kaiju, CAT). Its abundance was limited (RPKM values: 4.0-  
518 5.3) which is consistent with *Methanobacterium* proportion in the heavy DNA fractions  
519 (fr3). The presence of a gene coding for the long subunit of the terminase (TerL)  
520 suggests that C-41133 belongs to Caudoviricetes. The presence of an integrase gene  
521 is consistent with the detection of this partial genome in the cellular metagenomic  
522 dataset (provirus).

523 Contig C-330914 (82.9 kb) was detected as a viral contig in the cellular metagenomic  
524 data by only two tools out of three (VirSorter2 and ChecV, not VIBRANT). It likely  
525 originates from *Methanobrevibacter arboriphilus* (kaiju, CAT). It was present in all  
526 reactors (RPKM values: 0.03-24.7) and abundant only in reactor 13C\_rep2, in the  
527 intermediate DNA fraction (fr2). In this contig, many genes involved in the tail structure  
528 were detected in the right part of the contig (Fig 5). According to VIRFAM (Lopes *et al.*,  
529 2014), the corresponding viral particles have a myovirus-like morphology (contractile  
530 tail). However, many genes fell in the category "moron, auxiliary metabolic gene and  
531 host takeover" in its left part. This could be consistent with a proviral nature or could  
532 indicate that the left part of the contig is cellular, not belonging to the provirus



**Table3.** Abbreviations for protein annotation of the six contigs of interest

<b>Gene name</b>	<b>Function</b>	<b>Functional category</b>
<b>Ad</b>	head tail adaptor	Connector
<b>antiH</b>	anti Holin	Lysis
<b>BHP</b>	baseplate hub protein / baseplate protein	Tail
<b>Conn</b>	head-tail connector	Connector
<b>DNA-BP</b>	DNA binding protein	DNA, RNA and nucleotide metabolism
<b>DNA-CP</b>	DNA circularization protein	DNA, RNA and nucleotide metabolism
<b>DNApol</b>	DNA polymerase	DNA, RNA and nucleotide metabolism
<b>DT</b>	diffocin tube	Tail
<b>DTP</b>	distal tail protein	Tail
<b>Endo</b>	endonuclease	DNA, RNA and nucleotide metabolism
<b>Exc</b>	excisionase	Integration and excision
<b>Exo</b>	exonuclease	DNA, RNA and nucleotide metabolism
<b>HCP</b>	head completion protein	Head and packaging
<b>Helic</b>	DNA helicase	DNA, RNA and nucleotide metabolism
<b>Holin</b>	holin	Lysis
<b>HTA</b>	head tail adaptor	Connector
<b>Int</b>	serine integrase	Integration and excision
<b>MCM</b>	MCM protein	DNA, RNA and nucleotide metabolism
<b>mCP</b>	minor capsid protein	Head and packaging
<b>Methyl</b>	Modification methylase CviAll	DNA, RNA and nucleotide metabolism
<b>MT</b>	methyltransferase	DNA, RNA and nucleotide metabolism
<b>mTP</b>	minor tail protein	Tail
<b>Nuc</b>	nuclease	DNA, RNA and nucleotide metabolism
<b>PP</b>	portal protein	Head and packaging
<b>PAPSred</b>	Phosphoadenosine-phosphosulfate reductase	Moron, auxiliary metabolic gene and host takeover
<b>Prim</b>	DNA primase	DNA, RNA and nucleotide metabolism
<b>Prot</b>	prohead core protein protease	Head and packaging
<b>RapA</b>	RNA polymerase-associated protein RapA	Transcription regulation
<b>Rec</b>	recombination protein	DNA, RNA and nucleotide metabolism

<b>REN</b>	restriction endonuclease	DNA, RNA and nucleotide metabolism
<b>RepliOrg</b>	N-terminal phage replisome organiser	DNA, RNA and nucleotide metabolism
<b>RIP</b>	replication initiator protein	DNA, RNA and nucleotide metabolism
<b>RNApol</b>	RNA polymerase	DNA, RNA and nucleotide metabolism
<b>RNApσIH</b>	RNA polymerase sigma H factor	Transcription regulation
<b>Sheath</b>	tail sheath protein	Tail
<b>SP</b>	spike protein	Head and packaging
<b>Spanin</b>	spanin	Lysis
<b>TACP</b>	tail assembly chaperone protein	Tail
<b>TCP</b>	tail completion protein	Tail
<b>Ter</b>	terminase	Head and packaging
<b>TerL</b>	terminase large subunit	Head and packaging
<b>TerS</b>	terminase small subunit	Head and packaging
<b>TFP</b>	tail fiber protein	Tail
<b>THJP</b>	tail head junction protein	Connector
<b>TMP</b>	tape measure protein	Tail
<b>Transp</b>	transposase	Integration and excision
<b>TShIP</b>	tail sheath initiator protein	Tail
<b>TSP</b>	tail spike protein	Tail
<b>TTAP</b>	tail tip assembly protein	Tail
<b>TTP</b>	tail tip protein	Tail

536

537

538 ***Viruses infecting formatotrophic bacteria in the formate incubation microcosms***

539 Contig V-273599 is the complete circular genome (36.5 kb) of a virus (VIBRANT)  
540 predicted to infect Synergistaceae bacteria (WISH, SpacePHARER against a public and  
541 a home-made spacer databases). Based on its high abundance (RPKM values: 205-  
542 1 102), it could infects *Aminivibrio* or *Acetomicrobium* members, which are the  
543 dominant, active, Synergistaceae bacteria in the reactors. This is consistent with the  
544 predictions from SpacePharer (*Acetomicrobium*) and WISH (*Aminobacterium*).  
545 Interestingly, results of SpacePharer against the two spacer databases (public and  
546 home made) revealed both a hit with the same genus (*Acetomicrobium*). Since  
547 predictions from tools relying on spacer alignment can be considered as more robust  
548 compared to those based on k-mer signatures, we hypothesize an *Acetomicrobium*  
549 host. According to VIRFAM, V-273599 is a siphovirus (long, non-contractile tail). It  
550 belongs to the class *Caudoviricetes* (head-tail viruses), and all the functional modules  
551 required to complete an infection cycle could be identified: head and packaging,  
552 connector, tail, replication and lysis. No integrase gene was detected, however an  
553 excisionase was encoded, suggesting a link with temperate viruses.

554 Contig V-400577 is also the complete circular genome (135.7 kb) (VIBRANT) from a  
555 *Caudoviricetes* virus. It showed a moderate abundance (RPKM values: 7.7-77.9) and is  
556 predicted to infect Clostridiaceae members (Kaiju, WISH), possibly from the genus  
557 *Alkaliphilus*. The presence of integrase genes in the genome suggests a temperate  
558 lifestyle. Again, all the functional modules required for the infectious cycle of a  
559 temperate *Cauvoviricetes* virus were detected in the genome. It contained a gene  
560 encoding a phosphoadenosine-phosphosulfate reductase (PAPS-red), an AMG which  
561 was previously pointed out as abundant in anaerobic digester viromes (Willenbücher  
562 *et al.*, 2022)

563 Finally, C-470282 (73.8 kb) had in most cases a limited abundance (RPKM values: 0.8-  
564 62.1) and it could infect *Clostridium* species (Kaiju, CAT). Interestingly, Kaiju points to  
565 *Clostridium formicaceticum*. Although this taxonomic classification is not fully reliable,  
566 it is worth noting since this was the dominant active Clostridiaceae species in the

567 reactors. This is also supported by the abundance values, showing a systematic  
568 enrichment in the heavy fractions (fr3, RPKM: 10.6-62.1) compared to the intermediate  
569 ones (fr2, RPKM: 0.8-5.2). The gene content again indicates the belonging to the  
570 *Caudoviricetes* class and a temperate life style. VIRFAM predicts a myovirus-type.

### 571 **Concluding remarks**

572 Despite of a great number of experimental and bioinformatic approaches  
573 developed in recent years, establishing the link between viruses and their hosts  
574 within environmental ecosystems remains challenging in microbial ecology (Coclet and  
575 Roux, 2021). In our previous study, we proposed an original approach coupling SIP, an  
576 experimental method, with cellular metagenomic and metavirome analyses, in order to  
577 optimize the host prediction for viruses infecting methanogens in formate incubation  
578 microcosms. This approach showed its great utility, resulting in the discovery of several  
579 novel genomes of viruses infecting methanogens. In particular, two new families of  
580 spindle-shaped viruses were proposed, and this morphotype is specific to archaeal  
581 viruses and frequently observed in extreme environment. In fact, by focusing on specific  
582 active microorganisms through the SIP-DNA approach, we simplified the dataset to  
583 study and enabled to discriminate between viruses infecting microorganisms involved  
584 or not in formate bioconversion to biogas. Furthermore, the use of several  
585 bioinformatic tools with different strategies of host prediction (spacer alignment  
586 *versus* k-mer signature) and the use of both public and home-made databases  
587 improved greatly the robustness of the host prediction for the viral contigs. This  
588 analytic strategy is comparable with integrative tools such as VirHostMatcher-Net  
589 (Wang *et al.*, 2020) or PHISDetector (F. Zhang *et al.*, 2020) but our approach presents  
590 some advantages, especially in customization (each tool can be turned on/off) and in  
591 computational resource requirement.

592 In the present study, we could extend the approach to the study of both bacterial and  
593 archaeal viruses. We moreover improved our approach on two aspects. The co-  
594 assembly of the cellular metagenomes (instead of their individual assembly) reduced  
595 the redundancy of the MAG database. Besides, the search of proviruses in the cellular

596 metagenomes was implemented. It relies on the same bioinformatic tools and enabled  
597 to identify a significant number of additional (and distinct)viral genomes of interest.

598 To sum up, our original approach enabled to identify 89 complete or nearly complete  
599 genomes of (pro)viruses infecting bacteria and 38 genome of (pro)viruses infecting  
600 methanogens, involved in the bioconversion of formate to biogas. It led to the  
601 characterization of novel viruses infecting members of the genera *Clostridium*,  
602 *Methanobacterium*, *Methanobrevibacter*, and especially, for the first time a virus  
603 infecting the species *Acetomicrobium* (family Synergistaceae). Our results greatly  
604 expand the knowledge on the diversity of viruses within AD process. AMGs were  
605 present in many of the studied viral contigs: a larger-scale analysis of AMGs, similar to  
606 this recent study of anaerobic digester viromes (Willenbücher *et al.*, 2022), could be  
607 invaluable to further understand the effect of prokaryotic viruses in the AD process. In  
608 the future, it could also be interesting to evaluate the presence of the viral contigs  
609 identified in our two last study, in metaviromes from full-scale plants.

## 610 **Acknowledgments**

611 We are grateful to Chrystelle Bureau for her technical support on 16S rDNA  
612 metabarcoding, and to Angeline Guenne and Nadine Derlet for their technical  
613 assistance in analytical chemistry, in INRAE-PROSE. We also would like to acknowledge  
614 Christine Longin of the INRAE-MIMA2, platform for amazing TEM observations.

615 This work has benefited from the facilities and expertise of MIMA2 MET – GABI, INRAE,  
616 Agroparistech, 78352 Jouy-en-Josas, France.

617 This work was supported by Agence Nationale de la Recherche, France (ANR-17-CE05-  
618 0011, project VIRAME). MK was supported by Agence Nationale de la Recherche grant  
619 ANR-20-CE20-0009.

620 The authors declared to have no conflict of interest.



621 **References**

- 622 Anders, S., Pyl, P.T., and Huber, W. (2015) HTSeq—a Python framework to work with  
623 high-throughput sequencing data. *Bioinformatics* 31: 166–169.
- 624 Biswas, A., Staals, R.H.J., Morales, S.E., Fineran, P.C., and Brown, C.M. (2016)  
625 CRISPRDetect: A flexible algorithm to define CRISPR arrays. *BMC Genomics* 17:  
626 356.
- 627 Borrel, G., Colombet, J., Robin, A., Lehours, A.-C., Prangishvili, D., and Sime-Ngando, T.  
628 (2012) Unexpected and novel putative viruses in the sediments of a deep-dark  
629 permanently anoxic freshwater habitat. *ISME J* 6: 2119–2127.
- 630 Brysch, K., Schneider, C., Fuchs, G., and Widdel, F. (1987) Lithoautotrophic growth of  
631 sulfate-reducing bacteria, and description of *Desulfobacterium autotrophicum*  
632 gen. nov., sp. nov. *Arch Microbiol* 148: 264–274.
- 633 Buchfink, B., Xie, C., and Huson, D.H. (2015) Fast and sensitive protein alignment using  
634 DIAMOND. *Nat Methods* 12: 59–60.
- 635 Calusinska, M., Marynowska, M., Goux, X., Lentzen, E., and Delfosse, P. (2016) Analysis  
636 of dsDNA and RNA viromes in methanogenic digesters reveals novel viral  
637 genetic diversity. *Environ Microbiol* 18: 1162–1175.
- 638 Chapleur, O., Mazeas, L., Godon, J.-J., and Bouchez, T. (2016) Asymmetrical response of  
639 anaerobic digestion microbiota to temperature changes. *Appl Microbiol*  
640 *Biotechnol* 100: 1445–1457.
- 641 Chibani, C.M., Mahnert, A., Borrel, G., Almeida, A., Werner, A., Brugère, J.-F., et al. (2022)  
642 A catalogue of 1,167 genomes from the human gut archaeome. *Nat Microbiol*  
643 7: 48–61.
- 644 Chien, I.-C., Meschke, J.S., Gough, H.L., and Ferguson, J.F. (2013) Characterization of  
645 Persistent Virus-Like Particles in Two Acetate-Fed Methanogenic Reactors. *PLoS*  
646 *One* 8:.
- 647 Coclet, C. and Roux, S. (2021) Global overview and major challenges of host prediction  
648 methods for uncultivated phages. *Current Opinion in Virology* 49: 117–126.
- 649 Crowther, G.J., Kosály, G., and Lidstrom, M.E. (2008) Formate as the Main Branch Point  
650 for Methylophobic Metabolism in *Methylobacterium extorquens* AM1. *Journal*  
651 *of Bacteriology* 190: 5057–5062.
- 652 Edwards, R.A., McNair, K., Faust, K., Raes, J., and Dutilh, B.E. (2016) Computational  
653 approaches to predict bacteriophage–host relationships. *FEMS Microbiology*  
654 *Reviews* 40: 258–272.
- 655 Eryildiz, B., Lukitawesa, and Taherzadeh, M.J. (2020) Effect of pH, substrate loading,  
656 oxygen, and methanogens inhibitors on volatile fatty acid (VFA) production from

657 citrus waste by anaerobic digestion. *Bioresource Technology* 302: 122800.

658 Galiez, C., Siebert, M., Enault, F., Vincent, J., and Söding, J. (2017) WISH: who is the host?  
659 Predicting prokaryotic hosts from metagenomic phage contigs. *Bioinformatics*  
660 33: 3113–3114.

661 Guo, J., Bolduc, B., Zayed, A.A., Varsani, A., Dominguez-Huerta, G., Delmont, T.O., et al.  
662 (2021) VirSorter2: a multi-classifier, expert-guided approach to detect diverse  
663 DNA and RNA viruses. *Microbiome* 9: 37.

664 He, J., Wang, X., Yin, X., Li, Q., Li, X., Zhang, Y., and Deng, Y. (2018) Insights into  
665 biomethane production and microbial community succession during semi-  
666 continuous anaerobic digestion of waste cooking oil under different organic  
667 loading rates. *AMB Express* 8: 92.

668 Hernández, S. and Vives, M.J. (2020) Phages in Anaerobic Systems. *Viruses* 12: 1091.

669 Hyatt, D., Chen, G.-L., LoCasio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010)  
670 Prodigal: prokaryotic gene recognition and translation initiation site  
671 identification. *BMC Bioinformatics* 11: 119.

672 Inoue, K. and Iida, H. (1968) Bacteriophages of *Clostridium botulinum*. *J Virol* 2: 537–  
673 540.

674 Kanehisa, M., Sato, Y., and Morishima, K. (2016) BlastKOALA and GhostKOALA: KEGG  
675 Tools for Functional Characterization of Genome and Metagenome Sequences.  
676 *Journal of Molecular Biology* 428: 726–731.

677 Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., and Wang, Z. (2019) MetaBAT  
678 2: an adaptive binning algorithm for robust and efficient genome reconstruction  
679 from metagenome assemblies. *PeerJ* 7: e7359.

680 Kieft, K., Zhou, Z., and Anantharaman, K. (2020) VIBRANT: automated recovery,  
681 annotation and curation of microbial viruses, and evaluation of viral community  
682 function from genomic sequences. *Microbiome* 8: 90.

683 Kim, J.-G., Kim, S.-J., Cvirkaite-Krupovic, V., Yu, W.-J., Gwak, J.-H., López-Pérez, M., et al.  
684 (2019) Spindle-shaped viruses infect marine ammonia-oxidizing thaumarchaea.  
685 *PNAS* 116: 15645–15650.

686 Krupovič, M. and Bamford, D.H. (2008) Archaeal proviruses TKV4 and MVV extend the  
687 PRD1-adenovirus lineage to the phylum Euryarchaeota. *Virology* 375: 292–300.

688 Krupovič, M., Forterre, P., and Bamford, D.H. (2010) Comparative Analysis of the Mosaic  
689 Genomes of Tailed Archaeal Viruses and Proviruses Suggests Common Themes  
690 for Virion Architecture and Assembly with Tailed Viruses of Bacteria. *Journal of*  
691 *Molecular Biology* 397: 144–160.

692 Krupovič, M., Gribaldo, S., Bamford, D.H., and Forterre, P. (2010) The Evolutionary  
693 History of Archaeal MCM Helicases: A Case Study of Vertical Evolution

- 694 Combined with Hitchhiking of Mobile Genetic Elements. *Molecular Biology and*  
695 *Evolution* 27: 2716–2732.
- 696 Krupovic, M., Quemin, E.R.J., Bamford, D.H., Forterre, P., and Prangishvili, D. (2014)  
697 Unification of the Globally Distributed Spindle-Shaped Viruses of the Archaea.  
698 *Journal of Virology*.
- 699 Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat*  
700 *Methods* 9: 357–359.
- 701 Lemaire, O.N., Jespersen, M., and Wagner, T. (2020) CO<sub>2</sub>-Fixation Strategies in Energy  
702 Extremophiles: What Can We Learn From Acetogens? *Frontiers in Microbiology*  
703 11:.
- 704 Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015) MEGAHIT: an ultra-fast  
705 single-node solution for large and complex metagenomics assembly via succinct  
706 de Bruijn graph. *Bioinformatics* 31: 1674–1676.
- 707 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009) The  
708 Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–2079.
- 709 Liang, B., Wang, L.-Y., Mbadinga, S.M., Liu, J.-F., Yang, S.-Z., Gu, J.-D., and Mu, B.-Z.  
710 (2015) Anaerolineaceae and Methanosaeta turned to be the dominant  
711 microorganisms in alkanes-dependent methanogenic culture after long-term of  
712 incubation. *AMB Express* 5: 37.
- 713 Liu, Y., Demina, T.A., Roux, S., Aiewsakun, P., Kazlauskas, D., Simmonds, P., et al. (2021)  
714 Diversity, taxonomy, and evolution of archaeal viruses of the class  
715 Caudoviricetes. *PLOS Biology* 19: e3001442.
- 716 Lopes, A., Tavares, P., Petit, M.-A., Guérois, R., and Zinn-Justin, S. (2014) Automated  
717 classification of tailed bacteriophages according to their neck organization. *BMC*  
718 *Genomics* 15: 1027.
- 719 Madigou, C., Lê Cao, K.-A., Bureau, C., Mazéas, L., Déjean, S., and Chapleur, O. (2019)  
720 Ecological consequences of abrupt temperature changes in anaerobic digesters.  
721 *Chemical Engineering Journal* 361: 266–277.
- 722 Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., et al.  
723 (2020) Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and  
724 derived variants. *Nat Rev Microbiol* 18: 67–83.
- 725 von Meijenfeldt, F.A.B., Arkhipova, K., Cambuy, D.D., Coutinho, F.H., and Dutilh, B.E.  
726 (2019) Robust taxonomic classification of uncharted microbial sequences and  
727 bins with CAT and BAT. *Genome Biology* 20: 217.
- 728 Meile, L., Jenal, U., Studer, D., Jordan, M., and Leisinger, T. (1989) Characterization of  
729  $\psi$ M1, a virulent phage of *Methanobacterium thermoautotrophicum* Marburg.  
730 *Arch Microbiol* 152: 105–110.

- 731 Menzel, P., Ng, K.L., and Krogh, A. (2016) Fast and sensitive taxonomic classification for  
732 metagenomics with Kaiju. *Nature Communications* 7: 11257.
- 733 Molnár, J., Magyar, B., Schneider, G., Laczi, K., Valappil, S.K., Kovács, Á.L., et al. (2020)  
734 Identification of a novel archaea virus, detected in hydrocarbon polluted  
735 Hungarian and Canadian samples. *PLOS ONE* 15: e0231864.
- 736 Morales, C.A., Oakley, B.B., Garrish, J.K., Siragusa, G.R., Ard, M.B., and Seal, B.S. (2012)  
737 Complete genome sequence of the podoviral bacteriophage  $\Phi$ CP24R, which is  
738 virulent for *Clostridium perfringens*. *Arch Virol* 157: 769–772.
- 739 Nayfach, S., Camargo, A.P., Eloë-Fadrosch, E., Roux, S., and Kyrpides, N. (2021) CheckV:  
740 assessing the quality of metagenome-assembled viral genomes. *Nat Biotechnol*  
741 39: 578–585.
- 742 Neufeld, J.D., Vohra, J., Dumont, M.G., Lueders, T., Manefield, M., Friedrich, M.W., and  
743 Murrell, J.C. (2007) DNA stable-isotope probing. *Nat Protoc* 2: 860–866.
- 744 Ngo, V.-Q.-H., Enault, F., Midoux, C., Mariadassou, M., Chapleur, O., Mazéas, L., et al.  
745 (2022) Diversity of novel archaeal viruses infecting methanogens discovered  
746 through coupling of stable isotope probing and metagenomics. *Environ*  
747 *Microbiol.*
- 748 Nölling, J., Groffen, A., and de Vos, W.M. (1993)  $\phi$  F1 and  $\phi$ F3, two novel virulent,  
749 archaeal phages infecting different thermophilic strains of the genus  
750 *Methanobacterium*. *Microbiology* 139: 2511–2516.
- 751 Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015)  
752 CheckM: assessing the quality of microbial genomes recovered from isolates,  
753 single cells, and metagenomes. *Genome Res* 25: 1043–1055.
- 754 Parks, D.H., Rinke, C., Chuvochina, M., Chaumeil, P.-A., Woodcroft, B.J., Evans, P.N., et al.  
755 (2017) Recovery of nearly 8,000 metagenome-assembled genomes substantially  
756 expands the tree of life. *Nat Microbiol* 2: 1533–1542.
- 757 Pfister, P., Wasserfallen, A., Stettler, R., and Leisinger, T. (1998) Molecular analysis of  
758 *Methanobacterium* phage  $\Psi$ M2. *Molecular Microbiology* 30: 233–244.
- 759 Pickard, D.J.J. (2009) Preparation of bacteriophage lysates and pure DNA. *Methods Mol*  
760 *Biol* 502: 3–9.
- 761 Poirier, S., Desmond-Le Quéméner, E., Madigou, C., Bouchez, T., and Chapleur, O. (2016)  
762 Anaerobic digestion of biowaste under extreme ammonia concentration:  
763 Identification of key microbial phylotypes. *Bioresour Technol* 207: 92–101.
- 764 Puig-Castellví, F., Midoux, C., Guenne, A., Conteau, D., Franchi, O., Bureau, C., et al.  
765 (2022) Metataxonomics, metagenomics and metabolomics analysis of the  
766 influence of temperature modification in full-scale anaerobic digesters.  
767 *Bioresource Technology* 346: 126612.

- 768 Robledo-Mahón, T., Silva-Castro, G.A., Kuhar, U., Jamnikar-Ciglencečki, U., Barlič-  
769 Maganja, D., Aranda, E., and Calvo, C. (2019) Effect of Composting Under  
770 Semipermeable Film on the Sewage Sludge Virome. *Microb Ecol* 78: 895–903.
- 771 Si, B., Liu, Z., Zhang, Y., Li, J., Shen, R., Zhu, Z., and Xing, X. (2016) Towards biohythane  
772 production from biomass: Influence of operational stage on anaerobic  
773 fermentation and microbial community. *International Journal of Hydrogen*  
774 *Energy* 41: 4429–4438.
- 775 Tamaki, H., Zhang, R., Angly, F.E., Nakamura, S., Hong, P.-Y., Yasunaga, T., et al. (2012)  
776 Metagenomic analysis of DNA viruses in a wastewater treatment plant in tropical  
777 climate. *Environmental Microbiology* 14: 441–452.
- 778 Thauer, R.K., Kaster, A.-K., Seedorf, H., Buckel, W., and Hedderich, R. (2008)  
779 Methanogenic archaea: ecologically relevant differences in energy conservation.  
780 *Nat Rev Microbiol* 6: 579–591.
- 781 Thiroux, S., Dupont, S., Nesbø, C.L., Bienvenu, N., Krupovic, M., L’Haridon, S., et al. (2021)  
782 The first head-tailed virus, MFTV1, infecting hyperthermophilic methanogenic  
783 deep-sea archaea. *Environmental Microbiology* 23: 3614–3626.
- 784 Wang, K., Yin, J., Shen, D., and Li, N. (2014) Anaerobic digestion of food waste for  
785 volatile fatty acids (VFAs) production with different types of inoculum: Effect of  
786 pH. *Bioresource Technology* 161: 395–401.
- 787 Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007) Naïve Bayesian Classifier for  
788 Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Appl*  
789 *Environ Microbiol* 73: 5261–5267.
- 790 Wang, W., Ren, J., Tang, K., Dart, E., Ignacio-Espinoza, J.C., Fuhrman, J.A., et al. (2020) A  
791 network-based integrated framework for predicting virus–prokaryote  
792 interactions. *NAR Genomics and Bioinformatics* 2: lqaa044.
- 793 Wang, Y. and Qian, P.-Y. (2009) Conservative fragments in bacterial 16S rRNA genes  
794 and primer design for 16S ribosomal DNA amplicons in metagenomic studies.  
795 *PLoS One* 4: e7401.
- 796 Weidenbach, K., Nickel, L., Neve, H., Alkhnbashi, O.S., Künzel, S., Kupczok, A., et al.  
797 (2017) Methanosarcina Spherical Virus, a Novel Archaeal Lytic Virus Targeting  
798 Methanosarcina Strains. *Journal of Virology* 91: e00955-17.
- 799 Weidenbach, K., Wolf, S., Kupczok, A., Kern, T., Fischer, M.A., Reetz, J., et al. (2021)  
800 Characterization of Blf4, an Archaeal Lytic Virus Targeting a Member of the  
801 Methanomicrobiales. *Viruses* 13: 1934.
- 802 Willenbücher, K., Wibberg, D., Huang, L., Conrady, M., Ramm, P., Gätcke, J., et al. (2022)  
803 Phage Genome Diversity in a Biogas-Producing Microbiome Analyzed by  
804 Illumina and Nanopore GridION Sequencing. *Microorganisms* 10: 368.

805 Wolf, S., Fischer, M.A., Kupczok, A., Reetz, J., Kern, T., Schmitz, R.A., and Rother, M. (2019)  
806 Characterization of the lytic archaeal virus Drs3 infecting *Methanobacterium*  
807 *formicum*. *Arch Virol* 164: 667–674.

808 Wood, A.G., Whitman, W.B., and Konisky, J. (1989) Isolation and characterization of an  
809 archaeobacterial viruslike particle from *Methanococcus voltae* A3. *Journal of*  
810 *Bacteriology* 171: 93–98.

811 Zhai, N., Zhang, T., Yin, D., Yang, G., Wang, X., Ren, G., and Feng, Y. (2015) Effect of initial  
812 pH on anaerobic co-digestion of kitchen waste and cow manure. *Waste*  
813 *Management* 38: 126–131.

814 Zhang, F., Zhou, F., Gan, R., Ren, C., Jia, Y., Yu, L., and Huang, Z. (2020) PHISDetector: a  
815 tool to detect diverse in silico phage-host interaction signals for virome studies.  
816 661074.

817 Zhang, R., Mirdita, M., Levy Karin, E., Norroy, C., Galiez, C., and Söding, J. (2021)  
818 SpacePHARER: sensitive identification of phages from CRISPR spacers in  
819 prokaryotic hosts. *Bioinformatics*.

820 Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., et al. (2018) A  
821 Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred  
822 Server at its Core. *Journal of Molecular Biology* 430: 2237–2243.

823 Zolfo, M., Pinto, F., Asnicar, F., Manghi, P., Tett, A., Bushman, F.D., and Segata, N. (2019)  
824 Detecting contamination in viromes using ViromeQC. *Nat Biotechnol* 37: 1408–  
825 1412.

826

827

828

829

830 **Supplementary tables**

831

Table S1. Bioinformatic tool parameters and database versions

Tool	Version	Options	Employed database	Reference website
<b>fastp</b>	0.20.0	--length_required 50	NA	<a href="https://github.com/OpenGene/fastp">https://github.com/OpenGene/fastp</a>
<b>Megahit</b>	1.2.9	--continue --preset meta-large -f -- num-cpu-threads 20	NA	<a href="https://github.com/voutcn/megahit">https://github.com/voutcn/megahit</a>
<b>Prodigal</b>	2.6.3	-f gff -g 11 -p meta	NA	<a href="https://github.com/hyattpd/Prodigal">https://github.com/hyattpd/Prodigal</a>
<b>CAT</b>	5.2.3	default parameter	CAT database 07/01/2021	<a href="https://github.com/dutilh/CAT">https://github.com/dutilh/CAT</a>
<b>Kaiju</b>	1.7.3 1.8.1	default parameter	nr eukaryote 07/2021 AND nr virus 03/2021	<a href="https://github.com/bioinformatics-centre/kaiju">https://github.com/bioinformatics-centre/kaiju</a>
<b>Diamond</b>	0.9.26	--sensitive --max-target-seqs 1 -- outfmt 100 --salltitles --sallseqid -- verbose	nr eukaryote 07/2021	<a href="https://github.com/bbuchfink/diamond">https://github.com/bbuchfink/diamond</a>
<b>Bowtie2</b>	2.4.1	default parameter	NA	<a href="https://github.com/BenLangmead/bowtie2">https://github.com/BenLangmead/bowtie2</a>
<b>samtools</b>	1.9	default parameter	NA	<a href="https://github.com/samtools/samtools">https://github.com/samtools/samtools</a>
<b>CRISPRCasFinder</b>	4.2.20	-log -metagenome	NA	<a href="https://github.com/dcouvin/CRISPRCasFinder">https://github.com/dcouvin/CRISPRCasFinder</a>
<b>Metabat2</b>	2-2.12.1	--minContig 5000 --unbinned -- minClsSize 200000	NA	<a href="https://bitbucket.org/berkeleylab/metabat/">https://bitbucket.org/berkeleylab/metabat/</a>

<b>RefineM</b>	0.1.2	Functions: refineM taxon_filter & refineM filter_bins	Genome Taxonomy Database: gtdb_r95_protein_db.2020-07-30	<a href="https://github.com/dparks1134/RefineM">https://github.com/dparks1134/RefineM</a>
<b>CheckM</b>	1.0.18	lineage_wf -x fa	NA	<a href="https://github.com/Ecogenomics/CheckM">https://github.com/Ecogenomics/CheckM</a>
<b>WISH*</b>	1.1	default parameter	Homemade MAG database	<a href="https://github.com/soedinglab/WISH">https://github.com/soedinglab/WISH</a>
<b>CheckV</b>	0.7.0	checkv end_to_end	checkv-db-v1.0	<a href="https://bitbucket.org/berkeleylab/checkv/">https://bitbucket.org/berkeleylab/checkv/</a>
<b>VIBRANT</b>	1.2.1	-f nucl -virome	NA	<a href="https://github.com/AnantharamanLab/VIBRANT">https://github.com/AnantharamanLab/VIBRANT</a>
<b>VirSorter2</b>	2-2.2.3	default parameter	NA	<a href="https://github.com/jiarong/VirSorter2">https://github.com/jiarong/VirSorter2</a>
<b>ViromeQC</b>	1.0	default parameter	NA	<a href="https://github.com/SegataLab/viromeqc">https://github.com/SegataLab/viromeqc</a>
<b>SpacePharer</b>	2.8.12	easy-predict	spacers from local MAG databases and CRISPRCasdb (Cf.Mat&Met)	<a href="https://github.com/soedinglab/spacepharer">https://github.com/soedinglab/spacepharer</a>

832 \* For WISH, the models for the putative hosts were created using the 114 selected MAGs and the default options.

833 The null parameters were then obtained by using 11944 viral genomes downloaded from NCBI on the 5th and 6th of December,  
834 2020 and the R script provided by WISH authors:  
835 <https://raw.githubusercontent.com/soedinglab/WISH/master/computeNullParameters.R>

836 Finally, the prediction on the viral contigs were run as advised by WISH authors and using the file nullParameters.tsv generated  
837 at the previous step.



## 838 **Supplementary information**

### 839 *1.1. Isotopic analysis*

840 In order to identify the active methanogenesis pathway(s), the isotopic composition of  
841 the biogas was analyzed by coupling gas chromatograph (Trace GC Ultra, Thermo  
842 Fisher Scientific) with an Isotope-Ratio Mass Spectrometer (IRMS) (Delta V PLUS,  
843 Thermo Fisher Scientific). Delta expression  $\delta^{13}\text{C}$  (‰) of methane was used to identify  
844 the active methanogenesis pathway(s) (Whiticar *et al.*, 1986). This index was calculated  
845 using the following formula:

$$846 \quad \delta^{13}\text{C} = \frac{R_{\text{sample}} - R_{\text{standard}}}{R_{\text{standard}}} \times 1000 [\text{‰}]$$

847 where R was the ratio  $^{13}\text{C}/^{12}\text{C}$  and  $R_{\text{standard}}$  was equal to 0.0112372.

848 In the microcosms fed with unlabeled formate, the  $\delta^{13}\text{C}$  (‰) of methane was  $-59,3^0/00$   
849 for 12C\_rep1, indicating the dominance of the acetoclastic pathway in this sample.  
850 Concerning both other replicates, this value was  $-69,2^0/00$  and  $-73,5^0/00$  for 12C\_rep2  
851 and 12C\_rep3 respectively, indicating the dominance of the hydrogenotrophic  
852 pathway.

### 853 *1.2. Reproducibility and patterns of microbial communities across/among rep-* 854 *licate reactors*

855 Replicate reactors were introduced in the present study to be able to evaluate the  
856 reproducibility with respect to microbial community composition, and to verify the lack  
857 of effect of the  $^{13}\text{C}$ -labeled substrate compared to the unlabeled one.

858 To examine these aspects, 7 distinct datasets were considered: the 16S rDNA  
859 metabarcoding data, which inform on microbial community composition but may have  
860 strong PCR biases; the cellular metagenomics data (with and without removal of the  
861 low abundant contigs); the metavirome data (again with and without removal of the  
862 low abundant contigs). As the full metavirome must contain cellular contaminants, we  
863 also included a subset including only the contigs detected as viral by one of VirSorter2,

864 Vibrant or CheckV, and the subset of 49 contigs of interest identified in the metavirome  
865 dataset.

866 Regarding biogas production, different kinetic profiles were obtained according to the  
867 reactors: reactor 12C\_rep1 had the fastest dynamics, 13C\_rep1 and 13C\_rep3 had the  
868 slowest one, and the three remaining reactors, 12C\_rep2, 12C\_rep3, and 13C\_rep2, had  
869 similar, intermediate biogas production rates.

870 Regarding the influence of the <sup>13</sup>C-labeled compound, PERMANOVA (~  
871 labeling\_status) was applied to different relevant datasets (Table S1). The influence of  
872 the labeling was never significant, confirming the lack of major effect of employing <sup>13</sup>C-  
873 formate compared to unlabeled formate in our study, on the composition of microbial  
874 communities.

875 [Table S2. Results of PERMANOVA analyses conducted with the adonis function \(vegan package\) with](#)  
876 [the model ~labeling\\_status, where labeling\\_status has for value <sup>12</sup>C or <sup>13</sup>C according to the nature of](#)  
877 [the substrate added in the reactor.](#)

<b>Dataset</b>	<b>Number of ASVs/contigs</b>	<b>R<sup>2</sup></b>	<b>Pr(&gt;F)</b>	<b>Number of permutations</b>
<b>16S rDNA amplicons</b>	227	0.09169	1	719
<b>Metavirome</b>	56 613	0.22677	0.4	719
<b>Metavirome (RPKM &gt; 1)</b>	16 687	0.20651	0.4	719
<b>Metavirome (viral)</b>	15 294	0.22013	0.4	719
<b>Metavirome (49 contigs)</b>	49	0.27781	0.3	719

878

## Supplementary figures

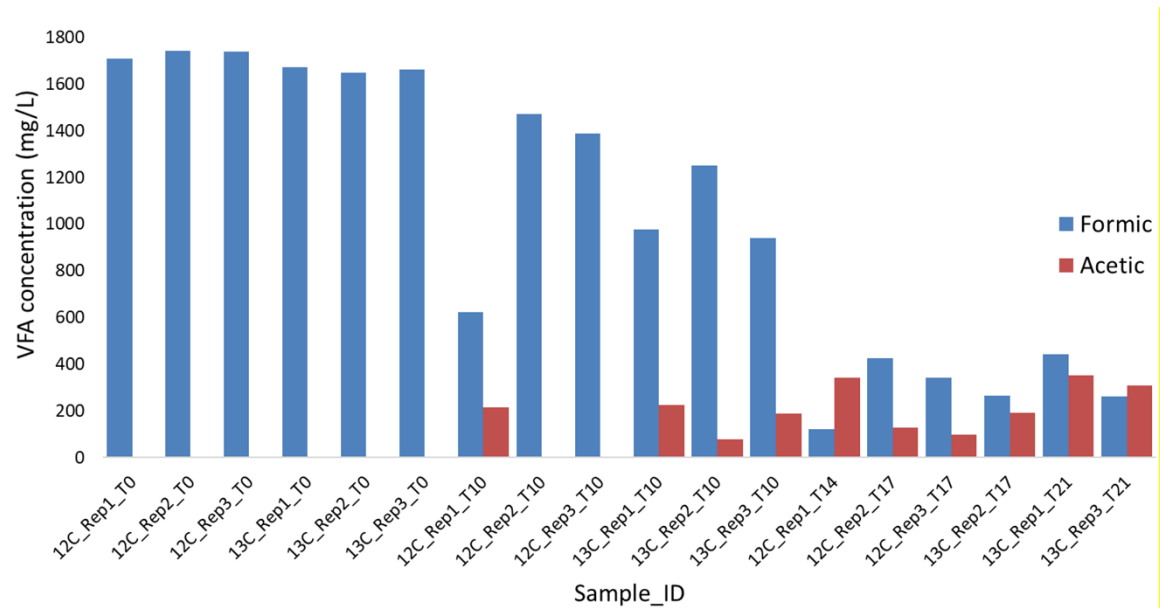


Figure S1. Formate and acetate concentration during the incubation

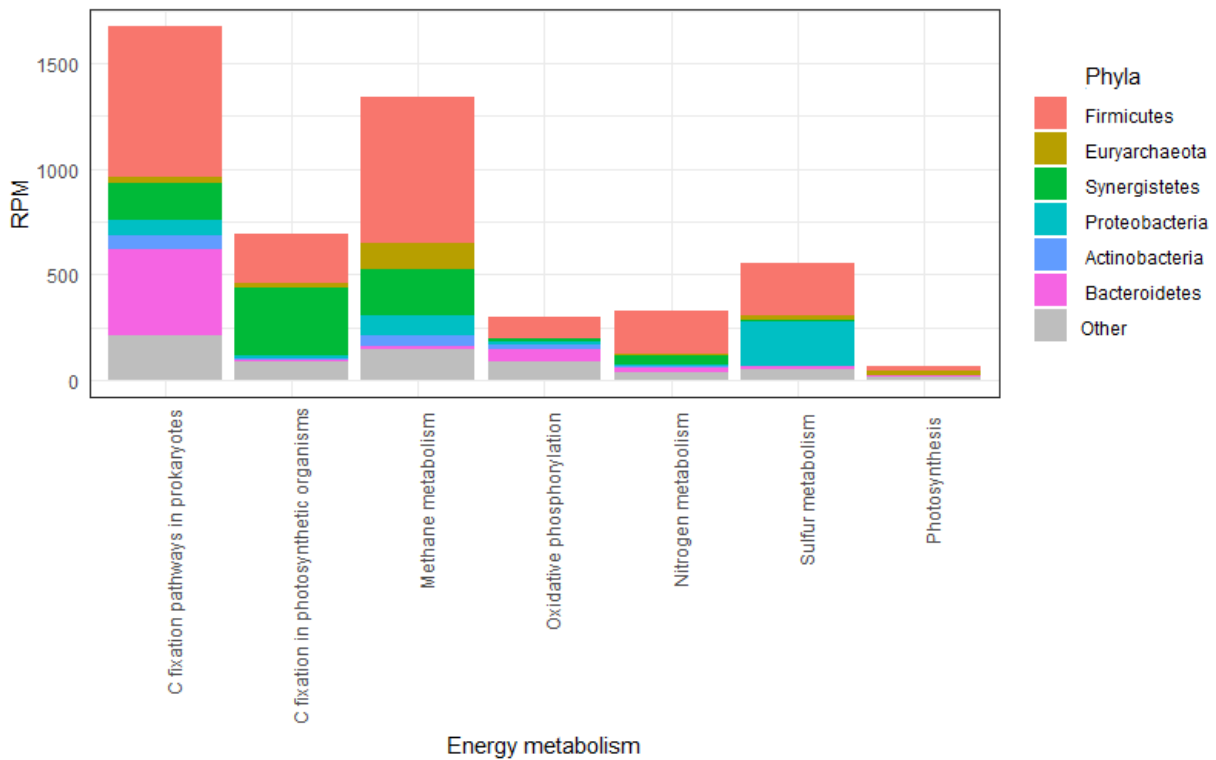


Figure S2. Taxonomic assignment (Phylum level) of the microorganisms involved in energy metabolism, based on the functional analysis of the cellular shotgun metagenome

# Chapitre 3 : Preuve de concept de l'utilisation du SIP pour les ADN viraux

## 1. Introduction

Les deux séries d'expériences présentées dans les deux précédents chapitres ont donc mis en évidence l'intérêt du SIP pour identifier les virus de microorganismes impliqués dans la métabolisation d'un substrat marqué au sein d'un écosystème. Il semble possible d'aller plus loin dans cette démarche, en appliquant la séparation sur gradient de densité non-seulement aux ADN cellulaires, mais également aux acides nucléiques viraux. En effet, les virus ayant infecté des cellules marquées au  $^{13}\text{C}$  devraient également être marqués au  $^{13}\text{C}$ , en particulier leurs acides nucléiques.

Afin d'établir la preuve expérimentale de ce concept, un modèle biologique de souche pure d'*Escherichia coli* et de son bactériophage T4, a été utilisé. Différents ratios de  $^{13}\text{C}$  et  $^{12}\text{C}$ -glucose ont été employés dans le milieu minimal utilisé pour la croissance des cellules. Dans ce chapitre, je présente ces premières expériences de preuve de concept où est analysée la cohérence entre la proportion de  $^{13}\text{C}$ -glucose employée, la densité des ADN cellulaires obtenus en l'absence de T4, et la densité des ADN de T4. Ces premiers résultats suggèrent que le SIP appliqué aux acides nucléiques viraux permettrait d'étudier spécifiquement des populations virales actives, associées aux microorganismes assimilant un substrat particulier

## 2. Plan expérimental

Dans un premier temps, des cultures de bactéries *E. coli* en présence de bactériophages T4 ont été effectuées, en milieu minimal contenant soit du  $^{13}\text{C}$ -glucose, soit du glucose non-marqué comme unique source de carbone (voir Matériels et Méthodes, section 8).

Cette expérience avait pour but de confirmer la bonne incorporation du  $^{13}\text{C}$  dans les cellules d'*E. coli* et de démontrer le même phénomène pour les particules de bactériophage T4 obtenues à partir du milieu contenant du glucose marqué. Elle a également permis d'évaluer la sensibilité de la méthode, en utilisant des quantités variables d'ADN lors de leur fractionnement selon leur densité, sur gradient isopycnique.

Dans un second temps, afin d'évaluer la cohérence entre la proportion de substrat marqué utilisée, et le taux d'enrichissement en  $^{13}\text{C}$  de l'ADN cellulaire et de l'ADN viral, des cultures en milieu minimum, en employant des mélanges de glucose non marqué et de  $^{13}\text{C}$ -glucose en proportions variées, ont été réalisées. Les pourcentages indiqués dans ce chapitre correspondent au pourcentage de  $^{13}\text{C}$ -glucose par rapport à la totalité du glucose ajouté. La concentration totale initiale de substrat reste toujours identique pour toutes les expériences (1% m/V). Une gamme de taux de marquage de 0% à 35% (0%, 5%, 10%, 20%, 35%) a été appliquée pour les cultures cellulaires d'*E. coli* en absence de bactériophage T4 et une gamme de marquage de 0% à 100% par pas de 20% a été appliquée pour les cultures cellulaires d'*E. coli* en présence du bactériophage T4.

Pour simuler le contexte de SIP appliqué en microbiologie environnementale, qui résulte d'un mélange d'ADN marqué (issu des microorganismes actifs) et non marqué, une analyse de SIP-ADN a également été effectuée avec un mélange de  $^{12}\text{C}$ -ADN et  $^{13}\text{C}$ -ADN d'*E. coli* à un ratio de 90 :10.

### 3. Résultats et discussions

#### 3.1. Mise en évidence du marquage au $^{13}\text{C}$

##### a. Marquage des cellules

Le taux d'enrichissement en  $^{13}\text{C}$  des cellules d'*E. coli* a été mesuré par IRMS. Pour les échantillons non-enrichis, le taux de  $^{13}\text{C}$  était inférieur au seuil de détection, ce qui est cohérent avec le faible taux de l'isotope  $^{13}\text{C}$  dans la nature ( $\approx 1,1\%$ ) (Source : Scientific Instrument Services <https://www.sisweb.com/referenc/source/exactmas.htm> ). Pour les

échantillons provenant de la culture sur  $^{13}\text{C}$ -glucose, les taux d'enrichissement variaient entre 89,4% et 94,9% avec une moyenne de 93,49%. Ces valeurs sont inférieures à celles attendues (de l'ordre de 100%). Différentes raisons ont été supposées pour expliquer ce phénomène. Le  $^{12}\text{C}$  pourrait provenir des traces du milieu LB utilisé pour les précultures d'*E. coli*, et avoir été introduit dans les cultures avec substrat marqué, lors de l'ensemencement avec 5 $\mu\text{L}$  de solution de préculture. Une autre explication pourrait être une précision limitée de mesure lors de l'analyse d'IRMS. En effet, une dilution importante de l'échantillon a été réalisée automatiquement au sein de l'IRMS avant analyse en raison de la quantité élevée de cellules présentes dans chaque capsule en étain, utilisée lors de l'analyse.

Ces résultats suffisent cependant à démontrer une bonne incorporation du  $^{13}\text{C}$  dans les cellules d'*E. coli*, en cas de croissance sur  $^{13}\text{C}$ -glucose.

#### b. Marquage des virus

Suite à la préparation des particules de bactériophage T4, l'ADN viral a été extrait et analysé par ultracentrifugation sur gradient isopycnique. Pour chacune des deux conditions ( $^{13}\text{C}$ -glucose ou glucose non marqué), des triplicats ont été réalisés. Les profils des concentrations d'ADN viral en fonction de leur densité ont mis en évidence deux groupes de pics distincts. L'un d'eux correspondait aux ADNs du bactériophage T4 issus des incubations en présence de  $^{12}\text{C}$ -glucose (ADNs à des densités plus faibles). Le second groupe correspondait à ceux issus des incubations en présence de  $^{13}\text{C}$ -glucose (marquage à 100%) (ADNs à des densités plus élevées) (Figure 41). Pour mieux visualiser la position des pics de concentration d'ADN, les valeurs de concentration ont été normalisées en divisant, pour chaque échantillon, les concentrations d'ADN par la valeur maximale obtenue pour cet échantillon (Figure 42). Une bonne reproductibilité a pu être notée pour les deux types d'échantillons. Ainsi, les pics de concentration en ADN des échantillons issus des cultures sur  $^{13}\text{C}$ -glucose étaient atteints à des densités de 1,745 ; 1,743 et 1,744 (moyenne de 1,744). Pour les ADNs obtenus en l'absence de substrat marqué, elles étaient de 1,704 ; 1,702 et 1,706 (moyenne de 1,704). D'après un t-test de Welch, cette différence

de densité massique entre les deux groupes de pics était significative ( $p = 8,60e-5$ ). Cela montre que le  $^{13}\text{C}$  a bien été incorporé dans les particules du bactériophage T4, et que les ADN ont bien pu être séparés selon leur densité. En outre, la différence moyenne de densité maximale, entre ces deux groupes de pics, était donc de 0,040 (1,744 – 1,704), proche de celle attendue d'après la littérature, c'est-à-dire 0,035 (Birnie and Rickwood, 2014).

Différentes quantités d'ADN viral ont été testées pour évaluer la sensibilité de l'étape de fractionnement selon la densité (Figure 41). Ainsi, les concentrations maximales en ADN obtenues dans les fractions, en l'absence de marquage, étaient de 960, 345 et 58 ng/mL. De façon similaire, pour les échantillons marqués, les concentrations maximales valaient 1 316 ng/mL, 250 ng/mL et 122 ng/mL. Cette capacité de détection de l'ADN viral à de faibles concentrations, à l'issue du fractionnement, est encourageante pour l'application de cette approche à de l'ADN viral environnemental.

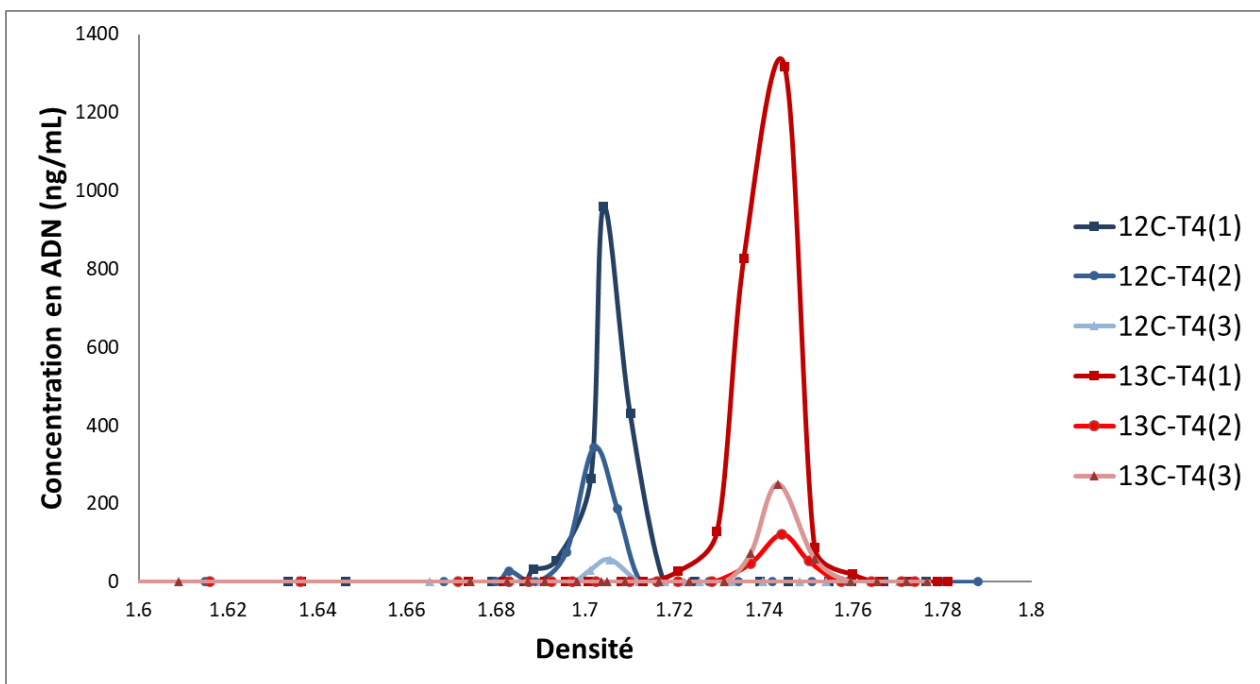


Figure 41. Profils de densité des ADNs du bactériophage T4 obtenus à partir des incubations avec du  $^{12}\text{C}$ -glucose ou du  $^{13}\text{C}$ -glucose.

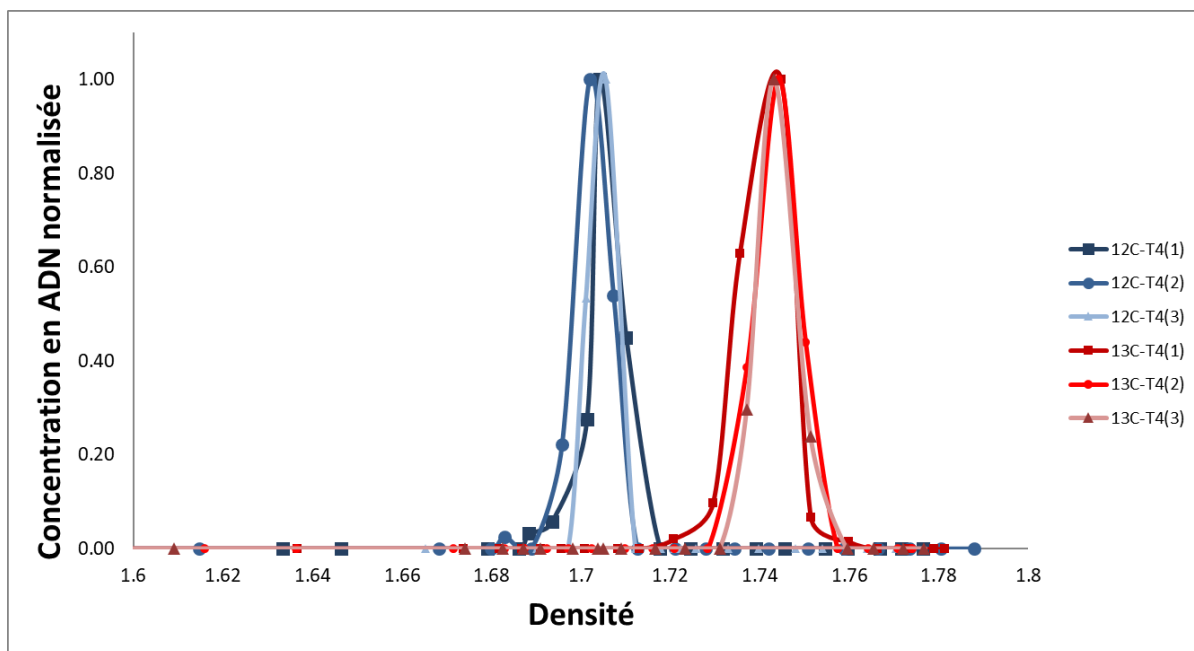


Figure 42. Profils de densité des ADNs du bactériophage T4, normalisés, obtenus à partir des incubations avec du  $^{12}\text{C}$ -glucose ou du  $^{13}\text{C}$ -glucose.

### 3.2. Cohérence entre la proportion de substrat marqué utilisée, et le taux d'enrichissement en $^{13}\text{C}$ des ADN cellulaires et viraux

#### a. Marquage des cellules

Une analyse IRMS a été réalisée sur des cellules d'*E. coli* collectées à partir de cultures réalisées selon une gamme de proportions en glucose, allant de 0% à 35% (voir Plan d'expériences). De façon similaire à ce qui avait été observé lors de la première expérience (section 3.1 a), un très faible signal a été obtenu pour le  $^{13}\text{C}$  pour les cellules issues d'une croissance sur  $^{12}\text{C}$ -glucose. Concernant la gamme de marquage, les taux de marquage des cellules étaient, dans la plupart des cas, légèrement inférieurs à ceux attendus, sauf pour l'échantillon correspondant à 5% de  $^{13}\text{C}$ -glucose, pour lequel le taux de marquage moyen des cellules atteignait 5,06% (Figure 43). Ceci pourrait être expliqué par les hypothèses et les biais présentés dans la section précédente.

Toutefois, on remarque que les taux d'enrichissement mesurés varient de façon linéaire en fonction des taux d'enrichissement attendus, avec une pente légèrement inférieure à 1. En conclusion, le taux de marquage en  $^{13}\text{C}$  des cellules d'*E. coli* est cohérent avec le taux de



$^{13}\text{C}$ -glucose utilisé pour la culture.

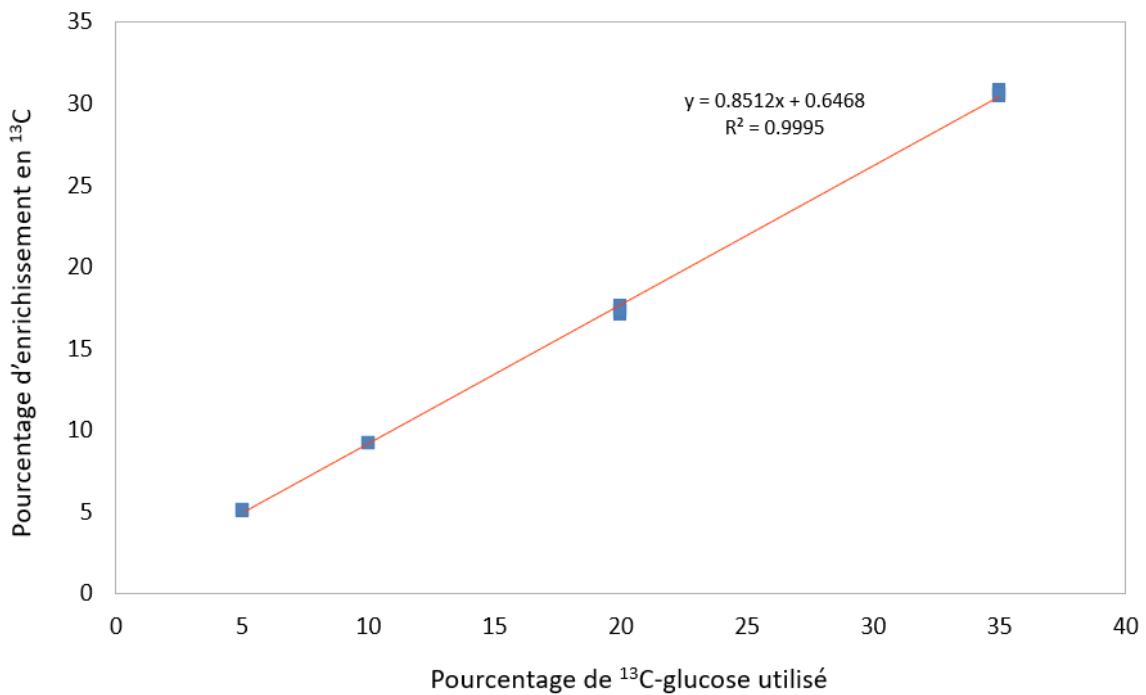


Figure 43. Taux d'enrichissement en  $^{13}\text{C}$  dans les cellules en fonction du pourcentage de  $^{13}\text{C}$ -glucose utilisé

#### b. Marquage des virus

Des analyses de SIP des ADN du bactériophage T4 collectés à partir des cultures avec différentes proportions de  $^{13}\text{C}$ -glucose (0%, 20 %, 40%, 60%, 80%, 100%) ont été réalisées. Une différence significative de densité entre les positions des pics à différentes proportions de glucose marqué a été constatée (Figure 44).

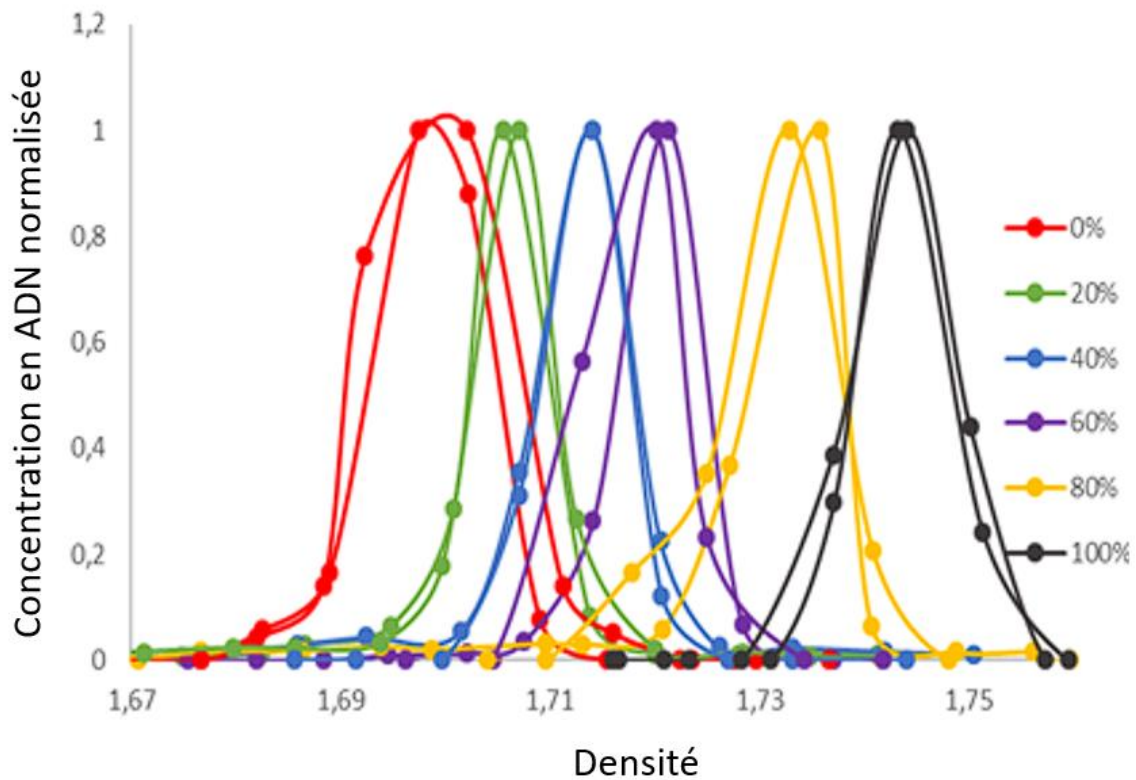


Figure 44. Profils de densité des ADNs du bactériophage T4, normalisés, obtenus à partir des incubations avec différentes proportions de  $^{13}\text{C}$ -glucose

Un calcul de densité théorique des pics a été réalisé pour évaluer le lien entre le taux de  $^{13}\text{C}$  glucose utilisé et le marquage en  $^{13}\text{C}$  dans l'ADN du phage T4. Ce calcul, basé sur l'hypothèse que le taux  $^{13}\text{C}$ -ADN du phage T4 est identique avec celui de  $^{13}\text{C}$ -glucose, suit l'équation suivante (Birnie and Rickwood, 2014) :

$$\rho = 0.098 * (\%GC) + 0.035 * (\%^{13}\text{C}) + 1.660$$

Où :  $\rho$  la est la densité du pic

(%GC) est le pourcentage en GC de l'ADN étudié (35,30% pour le bactériophage T4)

(% $^{13}\text{C}$ ) le taux de marquage en  $^{13}\text{C}$

Une variation linéaire est observée entre la position théorique du pic et sa position observée (Figure 45), démontrant une bonne cohérence entre la proportion de  $^{13}\text{C}$  glucose utilisée et le taux de marquage en  $^{13}\text{C}$  de l'ADN du bactériophage T4. Cela permet de

confirmer l'hypothèse du marquage indirect de l'ADN viral grâce à l'utilisation d'un substrat qui favorise la croissance de son hôte.

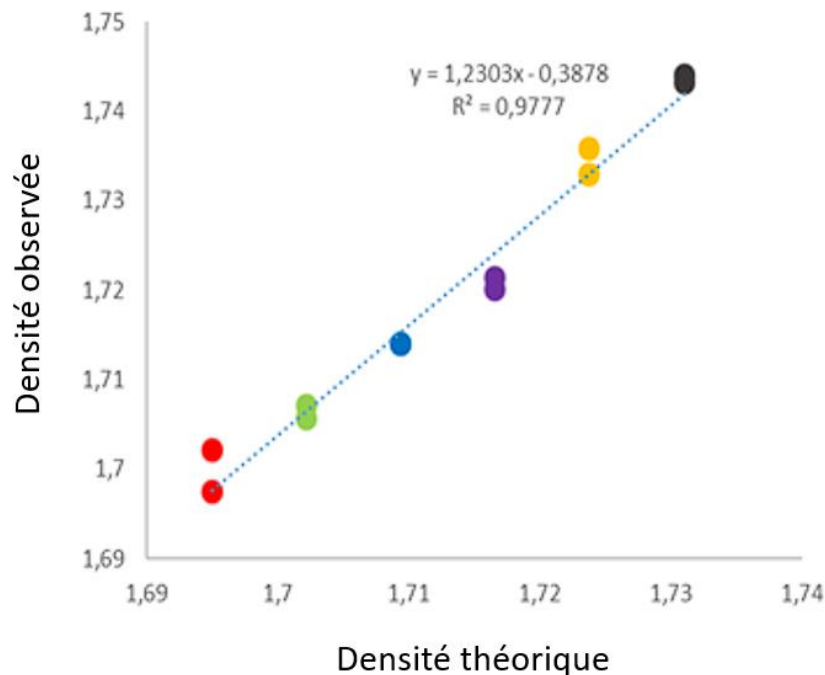


Figure 45. Densité observée pour les ADNs du bactériophage T4, en fonction de la densité théorique calculée à partir du contenu en GC du génome viral et du pourcentage de  $^{13}\text{C}$ -glucose utilisé

### 3.3. Perspectives d'utilisation du SIP appliqué à l'ADN viral dans le contexte de la microbiologie environnementale

Dans un échantillon issu d'un environnement complexe, tel que les procédés de DA, les communautés microbiennes sont très diverses. L'utilisation d'un substrat spécifique ne permet d'activer qu'un nombre limité de groupes microbiens. Lors d'application du SIP appliquées à ce type d'environnement, l'ADN total comprend donc un mélange d'ADN marqué et non marqué. Afin d'évaluer la séparation de l'ADN marqué et non marqué au sein d'un même échantillon, un mélange de  $^{12}\text{C}$ -ADN et  $^{13}\text{C}$ -ADN du bactériophage T4 a été séparé sur gradient isopycnique (Figure 46). Deux pics distincts ont été observés. La densité est à 1,7355 pour le pic de  $^{13}\text{C}$ -ADN et à 1,693 pour le pic de  $^{12}\text{C}$ -ADN. En raison de la faible quantité de  $^{13}\text{C}$ -ADN disponible, aucun réplica n'a été effectué. En revanche, le résultat obtenu a montré une sensibilité suffisante pour distinguer les 2 pics, malgré une faible quantité d'ADN lourd. Bien qu'en réalité, les différences de taux de marquages

seraient sans doute moins nettes, ces résultats demeurent prometteurs pour l'application du SIP aux ADN viraux.

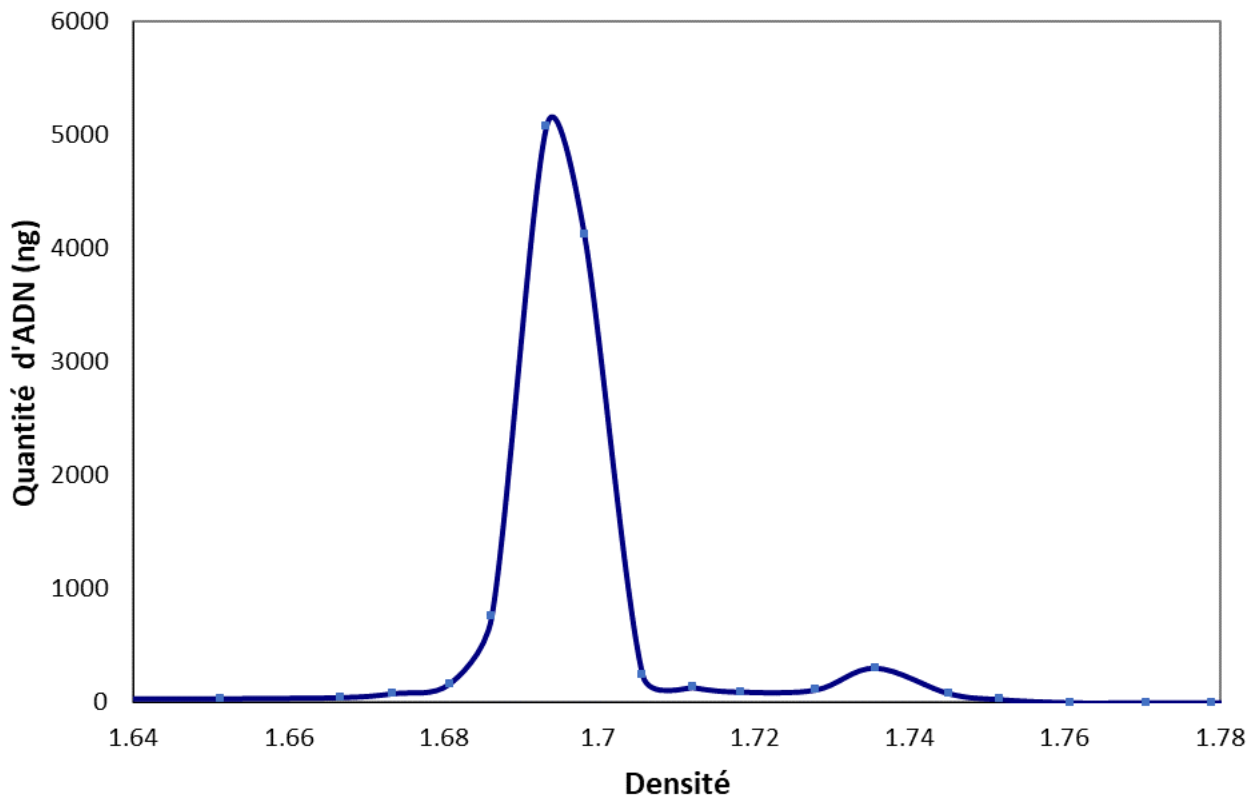


Figure 46. Profil de densité, normalisé, d'un mélange d'ADNs non marqués et marqués à 100% au  $^{13}\text{C}$ , du bactériophage T4.

Il est important de noter que cette expérience a été réalisée avec une très grande quantité d'ADN, de l'ordre de quelques  $\mu\text{g}$ . Il est possible d'utiliser de quantités moindres, de l'ordre de 500 ng. Ceci reste cependant un facteur limitant pour appliquer cette approche aux études environnementales, car étant donné la taille typique des génomes viraux, de telles quantités d'acides nucléiques peuvent être difficiles à atteindre.

Une autre limite est liée à la complexité et la diversité virale dans un échantillon environnemental. Comme la séparation des ADN dépend non seulement de leur taux de marquage en  $^{13}\text{C}$  mais aussi de leur contenu en GC, il est difficile de séparer parfaitement les  $^{13}\text{C}$ -ADN et  $^{12}\text{C}$ -ADN (Voir chapitre 1 et 2). Dans le cas des virus, cette séparation pourrait être encore plus difficile en raison d'une grande amplitude possible dans leur contenu en GC (de  $\approx 12\%$  à  $\approx 70\%$ ) (Figure 47).

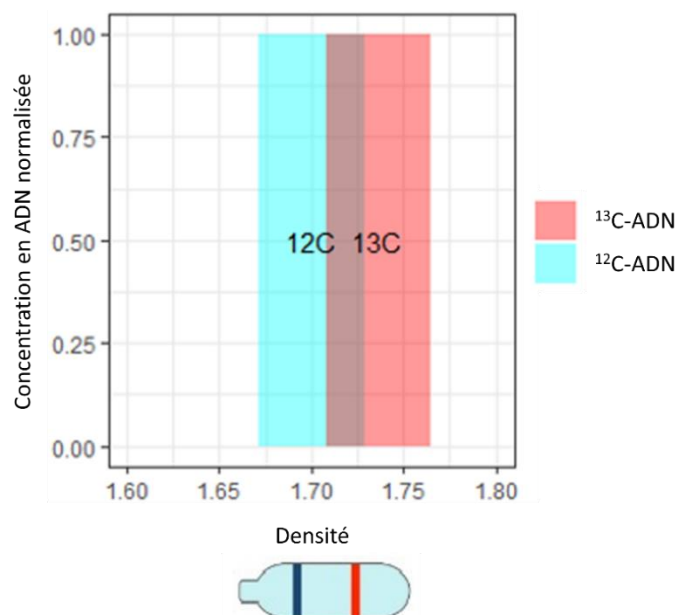


Figure 47. Intervalles de densité attendus pour les ADNs viraux, dans le cas de l'application du SIP à des échantillons environnementaux (seuil de pourcentage en GC de 20 à 70%)

En outre, le coût des molécules marquées au  $^{13}\text{C}$  est souvent très élevé, ce qui pourrait aussi être considéré comme un facteur limitant pour ce genre d'approche. Il est donc nécessaire de continuer à effectuer des tests pour déterminer la quantité d'ADN viral minimale permettant la distinction efficace des virus d'hôtes impliqués ou non-impliqués dans la dégradation d'un substrat donné.

#### 4. Conclusions et perspectives

Durant ce court chapitre, nous avons établi la preuve de concept de l'application du SIP aux ADN viraux. Malgré la présence de certaines limites, cette approche semble prometteuse pour identifier les virus infectant des hôtes qui assimilent des substrats spécifiques au sein de communautés microbiennes complexes. Dans le cadre de ma recherche, l'utilisation des substrats marqués tels que le formate, l'acétate, ou le méthanol, permettrait de mieux cibler les virus actifs au cours du métabolisme de ces molécules afin de faciliter la prédiction de leur hôte *via* des approches bioinformatiques. Ainsi, une série expérience d'incubation en microcosmes avec l'utilisation de méthanol non marqué ou de  $^{13}\text{C}$ -methanol comme source unique de carbone a été planifiée dans le prolongement de cette thèse.

---

# Discussion et perspectives

## 1. Virus au sein des procédés de DA

Les travaux réalisés au cours de cette thèse étaient une des premières expériences d'analyse de métaviromes et celle de métagénomés cellulaires issus de la même incubation en conditions contrôlées de laboratoires. En effet, les expériences en triplicata, ainsi que la contrôle de la quantité d'inoculum et de substrat utilisée ont permis de mieux comprendre la variabilité des communautés microbiennes au cours des procédés de DA, ainsi que la complexité des interactions entre les populations constituant ces communautés (procaryotes-procaryotes et virus-procaryotes).

Ces travaux ont apporté de nouvelles connaissances sur la diversité virale au sein des procédé de DA. Les observations au MET ont relevé une grande diversité morphologique de particules virales présentes dans tous nos microcosmes. Ainsi, outre les particules virales de type tête-queue qui sont dominantes, d'autres morphotypes ont été également détecté, en particulier des particules virales fusiformes (Voir les figures 2 dans les chapitres 1 et 2). Ces dernières sont spécifiques aux archées et elles ont souvent été observées dans des environnements extrêmes (sources hydrothermales, hypersalines), ce qui constitue un argument suggérant la présence de nouveaux virus d'archées méthanogènes au sein de procédés de DA en conditions mésophiles. En outre, de nouvelles particules atypiques et de structures indéfinies fréquemment présentes dans les digesteurs anaérobies ont été décrites (Figure 48). Bien que la nature de ces particules reste mystérieuse, certaines pourraient être d'origine virale et représenter des nouvelles morphologies, encore jamais rapportée dans la littérature.

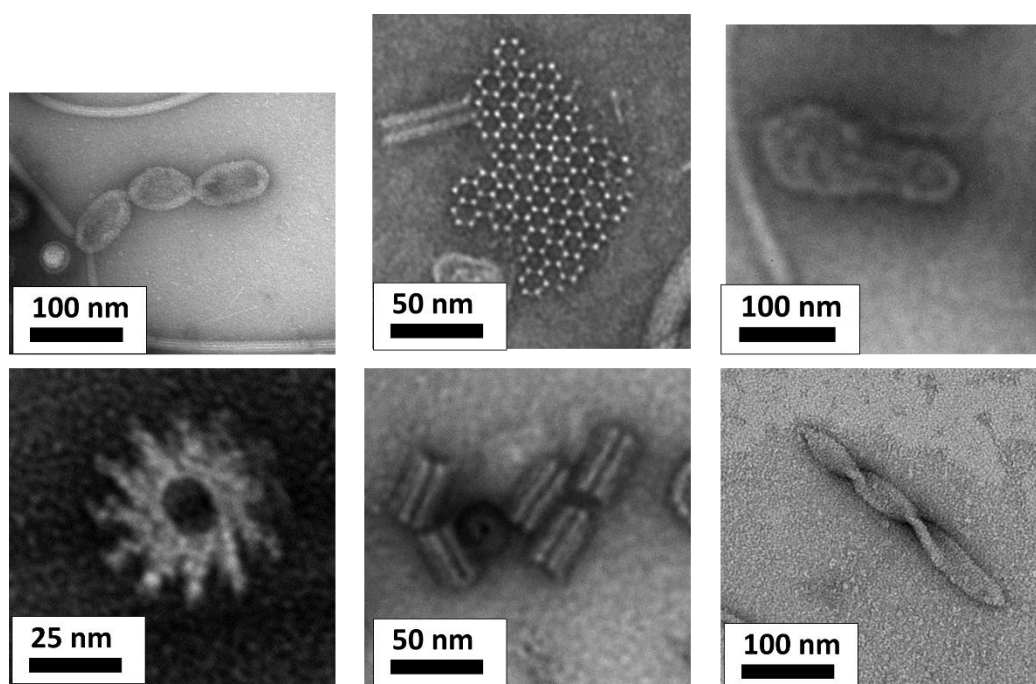


Figure 48. Particules et structures non-identifiées présentes dans les digesteurs anaérobies de laboratoire

En complément de ces observations, différents nouveaux génomes viraux ont été identifiés *via* des analyses bioinformatiques. En effet, de nouvelles familles de virus infectant les archées méthanogènes ont été caractérisées lors du chapitre 1. En particulier, j'ai décrit pour la première fois deux nouveaux virus fusiformes infectant des membres des ordres Methanobacteriales et Methanosarcinales. Cette découverte était en accord avec les observations en MET qui ont révélé la présence des particules virales fusiformes à différentes tailles dans mes microcosmes. En outre, plusieurs nouveaux génomes complets des virus infectant des bactéries impliquées dans les métabolismes de formate ont été également rapportés (Voir chapitre 2). En particulier, pour la première fois à notre connaissance, un génome complet de virus infectant un membre de la famille Synergistaceae a été identifié. Cette famille de bactéries acétogènes est fréquemment identifiées dans les digesteurs anaérobies, et elle est impliquée dans plusieurs voies de fixation du carbone (Poirier, Bize, *et al.*, 2016; Veeravalli *et al.*, 2017; Lv *et al.*, 2019; Nakasaki *et al.*, 2020). Dans la mesure où les acétogènes sont un groupe fonctionnel clé pour la régulation de la méthanisation, en interactions avec la fermentation en amont et la méthanogenèse en aval, il est intéressant de connaître leurs virus pour mieux comprendre

le fonctionnement de ce groupe clé. Mes travaux ont permis d'enrichir les connaissances sur la diversité virale dans le processus de méthanisation.

Permettant non seulement d'identifier de nouveaux virus, ces travaux ont également fourni des éléments pour mieux comprendre les associations hôte-virus au sein des bioprocédés de DA. En effet, les analyses bioinformatiques de contenu génomique ont montré dans la plupart des cas une cohérence entre l'abondance relative d'un virus et celle de son hôte prédict. Cette observation suggère que l'infection virale et la lyse des cellules hôte de ces virus dépendent de la densité cellulaire de l'hôte. En outre, la détection de gènes auxiliaires du métabolisme (AMG – « Auxiliary Metabolic Genes ») au sein de contigs viraux apporte la preuve d'interactions hôte-virus dans le passé. En effet, ces gènes sont connus pour leur vaste présence chez les virus et aussi pour leur d'origine cellulaire. L'acquisition d'AMGs par un virus, suite à l'infection virale d'un nouvel hôte, peut non seulement améliorer la réplication de l'ADN viral mais aussi affecter potentiellement différents cycles biogéochimiques en modulant des flux métaboliques microbiens clés, tels que le cycle du soufre ou du carbone, ou en affectant d'autres fonctions biologiques telles que la synthèse d'acides nucléiques ou de métabolites, ou la tolérance au stress. (Crummett *et al.*, 2016; Mara *et al.*, 2020). Dans le cadre de ma recherche, l'analyse des AMGs des contigs d'intérêt, menée lors du chapitre 2, a montré la présence de plusieurs AMGs chez des virus infectant les membres du genre *Clostridium*, présents dans nos microcosmes. Bien qu'aucun de ces AMGs ne soit impliqué directement dans le métabolisme du formate, certains codent pour des enzymes impliquées dans le métabolisme des acides aminés. Par ailleurs, des cytosines méthyltransferases ont également été identifiées, telles que K17398-DNA (cytosine-5)-methyltransferase 3A [EC:2.1.1.37], ou K00558-DNA (cytosine-5)-methyltransferase 1 [EC:2.1.1.37]). La présence de ces gènes pourrait refléter l'acquisition par les virus d'une stratégie pour protéger leur acide nucléique contre les systèmes de restriction modification de leur hôte (Kumar *et al.*, 1994; Walsh and Xu, 2006). Ceci constitue un bel exemple de course à l'armement entre les virus et leurs hôtes au sein des procédés de DA.



## 2. Développement méthodologique pour la métagénomique virale

Malgré le développement de différentes approches expérimentales et bioinformatiques depuis ces dernières années (voir « Synthèse bibliographique », section 4.3), la prédiction de l'hôte des virus reste toujours un défi en écologie microbienne. Deux grandes limites souvent rencontrées dans ces approches sont : la nécessité de connaître la séquence du génome viral pour certaines approches *single-cell* (ex : phageFISH, epicPCR) et la fiabilité de la prédiction pour les approches bioinformatiques. Dans ce manuscrit, je propose une approche expérimentale originale couplant SIP et analyses métagénomiques avec différents outils bioinformatiques. En sélectionnant les communautés microbiennes actives marquées, impliquées dans la méthanisation du formate, le SIP a permis de simplifier les communautés microbiennes et d'améliorer la construction de MAGs qui sont ensuite utilisés pour créer des bases de données locales pour la prédiction d'hôtes. En fait, pour une espèce considérée, les génomes présents dans un échantillon donné possèdent en général des différences avec ceux des bases de données publiques, en raison de différents phénomènes évolutifs ou parce qu'il ne s'agit pas exactement des mêmes souches. Ces différences, même très légères, peuvent affecter la précision des approches bioinformatiques de prédiction d'hôte, notamment les approches basées sur la composition des oligonucléotides (signature k-mer). Ces bases de données locales permettent de surmonter ce biais.

Afin d'assurer la reproductibilité de mes analyses, ainsi que de faciliter l'accès à ma stratégie d'analyses bioinformatiques, un pipeline bioinformatique a été développé ([https://for-gemia.inra.fr/cedric.midoux/workflow\\_metagenomics](https://for-gemia.inra.fr/cedric.midoux/workflow_metagenomics)) (voir « Matériels et Méthodes », section 6.2). Il s'agit d'une analyse complète à partir des données brutes, applicable pour la métagénomique cellulaire et virale, avec différentes options. Concernant la prédiction de l'hôte, différents outils bioinformatiques sont intégrés dans notre pipeline. Il s'agit d'un couplage entre deux types d'approches : –alignement de *spacers* et –signatures génomiques (k-mers). De plus, la comparaison entre les résultats obtenus à partir des bases de données publiques avec ceux obtenus à partir des bases de données locales et/ou entre

les résultats de chaque outil bioinformatique permet d'affiner la qualité de prédiction. L'annotation taxonomique des contigs viraux contre la base de données NCBI nr apporte également des informations intéressantes concernant l'hôte de ces virus (voir chapitre 1 et 2).

Mon approche est similaire à d'autres approches intégratives telles que celle de PHISDetector (F. Zhang *et al.*, 2020) ou de Virhostmatcher-Net (Wang *et al.*, 2020) s'appuyant également sur le couplage des méthodes d'alignement et des méthodes de similarité en k-mer. En comparaison à ces approches, la nôtre présente plusieurs avantages : elle est modulable en fonction du jeu de données avec un choix possible des outils à inclure dans l'analyse, elle est organisée sous la forme d'un pipe-line performant mais moins gourmand en termes de ressources informatiques requises. Cependant, notre pipeline présente toujours certains points à améliorer : certains outils nécessitent une intervention manuelle, notamment dans la construction des bases de données locales.

### 3. Perspectives

#### 3.1. Mise en évidence des nouveaux virus découverts grâce aux analyses bioinformatiques

Bien que plusieurs nouveaux virus aient été découverts durant ces travaux de thèse, il est toujours nécessaire d'aller plus loin dans la caractérisation de ces virus et de confirmer la solidité de notre approche en mettant en évidence les interactions hôte-virus par des approches expérimentales complémentaires. Grâce à la métagénomique, j'ai obtenu des séquences (complètes ou partielles) du génome viral, ce qui rend possible l'application de différentes approches d'hybridation.

La première approche à laquelle j'ai pensé est le PhageFISH (Allers *et al.*, 2013), dont le principe est basé sur la détection de gènes viraux par hybridation *in situ* fluorescente (FISH). Cette approche comprend 2 étapes principales : (i) fixation et identification des cellules hôtes avec des amorces spécifiques ciblant les ARNr et (ii) détection des gènes viraux en utilisant les amorces virales marquées avec de la digoxigénine (Dig). Le détail du

protocole est présenté dans (Barrero-Canosa and Moraru, 2019). Un biais de cette approche est la construction des amorces virales qui nécessite une certaine connaissance de la séquence du génome viral cible. Dans le cadre de nos études, cette connaissance a été obtenue grâce aux données de métagénomique virale. Pour préparer cette approche de marquage, les cellules issues de la deuxième série d'expériences (Chapitre 2) ont été fixées au paraformaldehyde et conservées à -20°C, ce qui laisse la possibilité d'appliquer cette méthode dans le futur.

D'autres approches pourraient également être appliquées. La cytométrie en flux, couplée avec une approche *single cell* (Tomaru and Nagasaki, 2007; Frossard *et al.*, 2016) permet de trier les cellules contenant le signal d'hybridation du gène viral cible. De plus, une variante de cette approche, la virométrie en flux, a bénéficié des progrès technologiques actuels (Gaudin and Barteneva, 2015; Zamora and Aguilar, 2018) et permet de trier directement des particules virales de façon non destructive.

Toutes ces approches permettent de non seulement identifier expérimentalement les interactions hôte-virus, mais également de favoriser ensuite la reconstruction du génome viral complet.

### 3.2. SIP virus : une approche prometteuse en écologie virale

Les travaux présentés dans ce manuscrit ont confirmé l'intérêt de l'approche SIP-ADN pour améliorer la prédiction d'hôte des virus. Pendant cette thèse, j'ai également établi la preuve de concept de l'application du SIP aux ADN viraux. Par rapport au SIP appliqué aux ADN cellulaires, elle semble plus avantageuse. En effet, le SIP-virus permet de simplifier le virome à analyser, ce qui entraîne la diminution du nombre de contigs à étudier, réduisant ainsi la durée d'analyse. Outre sa plus grande spécificité, cette sélection directe d'ADN viraux favorise l'amplification des génomes viraux d'intérêt, ce qui pourrait favoriser une meilleure reconstruction de génomes viraux complets par la suite.

De plus, le couplage de ces deux analyses de SIP permettrait une double simplification des communautés procaryotiques et virales, et entrainerait une bien meilleure fiabilité de la prédiction de l'hôte des virus par des approches bioinformatiques.

### 3.3. Approfondissement des recherches sur les virus d'archées méthanogènes : expériences avec du <sup>13</sup>C-méthanol

Les tests d'incubation en microcosmes présentés dans le chapitre 1 ont montré que les communautés microbiennes au sein des microcosmes variaient en fonction du substrat utilisé. En particulier, cet impact a été nettement observé dans les communautés d'archée méthanogènes : Methanobacteriales était l'ordre dominant dans les microcosmes alimentés avec le formate tandis que Methanosarcinales dominait dans les microcosmes alimentés avec de l'acétate ou du méthanol (voir chapitre 1). Les archées méthanogènes sont des acteurs clés de la méthanogenèse, l'étape ultime du processus de DA. Les connaissances sur les virus qui les infectent permettront de mieux comprendre le fonctionnement de ce processus. Pour cela, une série d'expériences d'incubation en microcosmes avec du <sup>13</sup>C-méthanol comme source unique de carbone a été planifiée pour la suite de cette thèse. Cette série d'expérience sera également l'occasion de mettre en œuvre l'approche SIP-virus.

### 3.4. Diversité virale dans les installations industrielles de DA

Afin de mieux comprendre la diversité virale au sein des bioprocédés de DA à l'échelle industrielle, six différents digesteurs anaérobies à pleine échelle ont été échantillonnés. Ces digesteurs possèdent des caractéristiques très variées (Tableau 11), permettant *a priori* de refléter une plus grande diversité microbienne par rapport à mes expériences en microcosmes.

Tableau 11. Caractéristiques des méthaniseurs échantillonnés

Nom de digesteur	Température	Nature de l'alimentation
M1	Mésophile	Territoriale (ordures ménagères, graisses)
P1	Psychrophile	Déchets de la ferme (lisiers de vache)
M2	Mésophile	Déchets de la ferme (fumiers d'herbivores, déchets verts)
M3	Mésophile	Boues de station d'épuration urbaine
T1	Thermophile	Boues de station d'épuration urbaine
M4	Mésophile	Territoriale (fumiers d'herbivores, déchets verts, biodéchets)

Les résultats préliminaires ont montré une variabilité des communautés procaryotes et des paramètres physico-chimiques, fortement en lien avec le type des digesteurs et la nature des substrats. Cela suggère donc une variabilité des communautés virales. Ainsi, les observations au MET ont montré une grande diversité de particules virales. Les métaviromes ont été également collectés puis séquencés et sont en attente d'analyses bioinformatiques.

En parallèle, appliquer notre pipeline d'analyse bioinformatique aux métaviromes issus des digesteurs anaérobies déjà publiés dans la littérature pourrait apporter de nouvelles connaissances sur la diversité virale et permettrait également de comparer toutes ces données avec les miennes pour identifier les familles virales présentes fréquemment dans ces procédés de DA.

## 4. Conclusions

En proposant d'une approche originale couplant le SIP-ADN et différents outils bioinformatiques, mes travaux ont fourni des stratégies pour l'identification des liens hôte-virus au sein des communautés microbiennes complexes, ainsi que l'accès à l'analyse de ces liens par des approches *in silico* à partir de données métagénomiques, sous la forme d'un pipeline bioinformatique.

En résumé, les travaux effectués durant ma thèse ont permis de significativement développer les connaissances sur la diversité virale au sein des bioprocédés de DA, notamment pour les virus d'archées méthanogènes et des bactéries impliquées dans la méthanisation du formate, ainsi que d'apporter les éléments pour répondre à des questions plus fondamentales concernant les interactions entre les virus et leur hôte, ainsi que leur prévalence dans ces écosystèmes. La question d'une éventuelle application biotechnologique de ces résultats demeure à ce jour une question ouverte. Ces connaissances sur les virus pourront-elles un jour être utilisées pour modifier la composition des microbiotes présents dans les bioréacteurs ? L'utilisation des virus ou de molécules d'origines virales spécifiques permettrait de proposer des leviers technologiques pour maintenir le bon fonctionnement de la méthanisation ou encore à plus long terme, pour contrôler la production des molécules de haute valeur économique ajoutée que le méthane, dans un contexte de bioraffinerie de déchets organiques. Ainsi, ce concept de biocontrôle avec l'utilisation des virus virulents appliqué au digesteurs anaérobies a déjà été évoqué dans la littérature (Kotay *et al.*, 2011; Zhang *et al.*, 2017). Cependant, pour des communautés microbiennes aussi complexes que celles présentes au sein des procédés de DA, des connaissances bien approfondies de la diversité virale sont requises. Mes travaux s'insèrent dans ce contexte, visant à contribuer aux bases scientifiques nécessaires pour pouvoir imaginer des applications technologiques des virus de procaryotes dans le domaine des biotechnologies environnementales.

# Bibliographie

- Abrescia, N.G.A., Bamford, D.H., Grimes, J.M., and Stuart, D.I. (2012) Structure unifies the viral universe. *Annual Review of Biochemistry* **81**: 795–822.
- Ahlgren, N.A., Ren, J., Lu, Y.Y., Fuhrman, J.A., and Sun, F. (2017) Alignment-free oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. *Nucleic Acids Research* **45**: 39–53.
- Ahring, B.K. (2003) Perspectives for anaerobic digestion. In *Biomethanation I. Advances in Biochemical Engineering/Biotechnology*. Ahring, B.K., Angelidaki, I., de Macario, E.C., Gavala, H.N., Hofman-Bang, J., Macario, A.J.L., et al. (eds). Berlin, Heidelberg: Springer, pp. 1–30.
- Alauzet, C. (2009) Taxonomie des bactéries anaérobies: De la reclassification à la découverte de nouveaux pathogènes.
- Alkhnabshi, O.S., Meier, T., Mitrofanov, A., Backofen, R., and Voß, B. (2019) CRISPR-Cas bioinformatics. *Methods*.
- Allers, E., Moraru, C., Duhaime, M.B., Beneze, E., Solonenko, N., Barrero-Canosa, J., et al. (2013) Single-cell and population level viral infection dynamics revealed by phageFISH, a method to visualize intracellular and free viruses. *Environmental Microbiology* **15**: 2306–2318.
- Amani, T., Nosrati, M., and Sreekrishnan, T.R. (2010) Anaerobic digestion from the viewpoint of microbiological, chemical, and operational aspects — a review. *Environ Rev* **18**: 255–278.
- Ameen, A. and Raza, S. (2017) Metaproteomics approaches and techniques: A review. *International Journal of Advances in Scientific Research* **3**: 49–51.
- Amgarten, D., Iha, B.K.V., Piroupo, C.M., Silva, A.M. da, and Setubal, J.C. (2020) vHULK, a new tool for bacteriophage host prediction based on annotated genomic features and deep neural networks.
- Anders, S., Pyl, P.T., and Huber, W. (2015) HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**: 166–169.
- Bairoch, A. and Apweiler, R. (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Research* **28**: 45–48.
- Baker, B.J., De Anda, V., Seitz, K.W., Dombrowski, N., Santoro, A.E., and Lloyd, K.G. (2020) Diversity, ecology and evolution of Archaea. *Nat Microbiol* **5**: 887–900.
- Baltimore, D. (1971) Expression of animal virus genomes. *BACTERIOL REV* **35**: 7.
- Baquero, D.P., Liu, Y., Wang, F., Egelman, E.H., Prangishvili, D., and Krupovic, M. (2020) Chapter Four - Structure and assembly of archaeal viruses. In *Advances in Virus Research. Virus Assembly and Exit Pathways*. Kielian, M., Mettenleiter, T.C., and

- Roossinck, M.J. (eds). Academic Press, pp. 127–164.
- Barns, S.M., Delwiche, C.F., Palmer, J.D., and Pace, N.R. (1996) Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proceedings of the National Academy of Sciences* **93**: 9188–9193.
- Barrero-Canosa, J. and Moraru, C. (2019) PhageFISH for Monitoring Phage Infections at Single Cell Level. *Methods Mol Biol* **1898**: 1–26.
- Bath, C. and Dyall-Smith, M.L. (1998) His1, an Archaeal Virus of the Fuselloviridae Family That Infects *Haloarcula hispanica*. *Journal of Virology*.
- Bell, T.H., Yergeau, E., Martineau, C., Juck, D., Whyte, L.G., and Greer, C.W. (2011) Identification of Nitrogen-Incorporating Bacteria in Petroleum-Contaminated Arctic Soils by Using [15N]DNA-Based Stable Isotope Probing and Pyrosequencing. *Applied and Environmental Microbiology* **77**: 4163–4171.
- Berghuis, B.A., Yu, F.B., Schulz, F., Blainey, P.C., Woyke, T., and Quake, S.R. (2019) Hydrogenotrophic methanogenesis in archaeal phylum Verstraetearchaeota reveals the shared ancestry of all methanogens. *PNAS* **116**: 5037–5044.
- Birnie, G.D. and Rickwood, D. (2014) Centrifugal Separations in Molecular and Cell Biology, Butterworth-Heinemann.
- Biswas, A., Staals, R.H.J., Morales, S.E., Fineran, P.C., and Brown, C.M. (2016) CRISPRDetect: A flexible algorithm to define CRISPR arrays. *BMC Genomics* **17**: 356.
- Boeckaerts, D., Stock, M., Criel, B., Gerstmans, H., De Baets, B., and Briers, Y. (2021) Predicting bacteriophage hosts based on sequences of annotated receptor-binding proteins. *Sci Rep* **11**: 1467.
- Borja, R. and Rincón, B. (2017) Biogas Production. In *Reference Module in Life Sciences*. Elsevier.
- Borrel, G., Adam, P.S., McKay, L.J., Chen, L.-X., Sierra-García, I.N., Sieber, C.M.K., et al. (2019) Wide diversity of methane and short-chain alkane metabolisms in uncultured archaea. *Nat Microbiol* **4**: 603–613.
- Borrel, G., Colombet, J., Robin, A., Lehours, A.-C., Prangishvili, D., and Sime-Ngando, T. (2012) Unexpected and novel putative viruses in the sediments of a deep-dark permanently anoxic freshwater habitat. *ISME J* **6**: 2119–2127.
- Bouchez, T., Bliex, A.L., Dequiedt, S.S., Domaizon, I., Dufresne, A., Ferreira, S., et al. (2016) La microbiologie moléculaire au service du diagnostic environnemental. Synthèse de Recherche. Contrat n°1406C0050. Projet de recherche coordonné par: "Observatoire des sols vivants", ADEME Agence de l'Environnement et de la Maitrise de l'Energie.
- Brito, I.L. (2021) Examining horizontal gene transfer in microbial communities. *Nat Rev Microbiol* **19**: 442–453.
- Bryant, M.P. (1979) Microbial Methane Production—Theoretical Aspects. *J Anim Sci* **48**:



193–201.

- Brysch, K., Schneider, C., Fuchs, G., and Widdel, F. (1987) Lithoautotrophic growth of sulfate-reducing bacteria, and description of *Desulfobacterium autotrophicum* gen. nov., sp. nov. *Arch Microbiol* **148**: 264–274.
- Buchfink, B., Xie, C., and Huson, D.H. (2015) Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12**: 59–60.
- Calusinska, M., Marynowska, M., Goux, X., Lentzen, E., and Delfosse, P. (2016) Analysis of dsDNA and RNA viromes in methanogenic digesters reveals novel viral genetic diversity. *Environ Microbiol* **18**: 1162–1175.
- Chapleur, O. (2012) Ingénierie écologique des communautés microbiennes de méthanisation des déchets ligno-cellulosiques.
- Chapleur, O., Mazeas, L., Godon, J.-J., and Bouchez, T. (2016) Asymmetrical response of anaerobic digestion microbiota to temperature changes. *Appl Microbiol Biotechnol* **100**: 1445–1457.
- Chater, K.F. and Wilde, L.C.Y. (1980) *Streptomyces albus* G Mutants Defective in the SalGI Restriction-Modification System. *Microbiology* **116**: 323–334.
- Chen, Y., Cheng, J.J., and Creamer, K.S. (2008) Inhibition of anaerobic digestion process: A review. *Bioresour Technol* **99**: 4044–4064.
- Chibani, C.M., Mahnert, A., Borrel, G., Almeida, A., Werner, A., Brugère, J.-F., et al. (2022) A catalogue of 1,167 genomes from the human gut archaeome. *Nat Microbiol* **7**: 48–61.
- Chien, I.-C., Meschke, J.S., Gough, H.L., and Ferguson, J.F. (2013) Characterization of Persistent Virus-Like Particles in Two Acetate-Fed Methanogenic Reactors. *PLoS One* **8**:
- Coclet, C. and Roux, S. (2021) Global overview and major challenges of host prediction methods for uncultivated phages. *Current Opinion in Virology* **49**: 117–126.
- Conrad, R. (2020) Importance of hydrogenotrophic, acetoclastic and methylotrophic methanogenesis for methane production in terrestrial, aquatic and other anoxic environments: A mini review. *Pedosphere* **30**: 25–39.
- Cotton, C.A., Claassens, N.J., Benito-Vaquerizo, S., and Bar-Even, A. (2020) Renewable methanol and formate as microbial feedstocks. *Current Opinion in Biotechnology* **62**: 168–180.
- Coutinho, F.H., Zaragoza-Solas, A., López-Pérez, M., Barylski, J., Zielezinski, A., Dutilh, B.E., et al. (2021) RaFAH: Host prediction for viruses of Bacteria and Archaea based on protein content. *Patterns* **2**: 100274.
- Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B., et al. (2018) CRISPRCasFinder, an update of CRISPRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids*

*Research* **46**: W246–W251.

- Croteau, F.R., Rousseau, G.M., and Moineau, S. (2018) Le système CRISPR-Cas - Au-delà de l'édition génomique. *Med Sci (Paris)* **34**: 813–819.
- Crowther, G.J., Kosály, G., and Lidstrom, M.E. (2008) Formate as the Main Branch Point for Methylophilic Metabolism in *Methylobacterium extorquens* AM1. *Journal of Bacteriology* **190**: 5057–5062.
- Crummett, L.T., Puxty, R.J., Weihe, C., Marston, M.F., and Martiny, J.B.H. (2016) The genomic content and context of auxiliary metabolic genes in marine cyanomyoviruses. *Virology* **499**: 219–229.
- Csardi, G. and Nepusz, T. (2005) The Igraph Software Package for Complex Network Research. *InterJournal Complex Systems*: 1695.
- Danovaro, R., Dell'Anno, A., Trucco, A., Serresi, M., and Vanucci, S. (2001) Determination of Virus Abundance in Marine Sediments. *Applied and Environmental Microbiology* **67**: 1384–1387.
- Deng, L., Gregory, A., Yilmaz, S., Poulos, B.T., Hugenholtz, P., and Sullivan, M.B. Contrasting Life Strategies of Viruses that Infect Photo- and Heterotrophic Bacteria, as Revealed by Viral Tagging. *mBio* **3**: e00373-12.
- Deng, L., Ignacio-Espinoza, J.C., Gregory, A.C., Poulos, B.T., Weitz, J.S., Hugenholtz, P., and Sullivan, M.B. (2014) Viral tagging reveals discrete populations in *Synechococcus* viral genome sequence space. *Nature* **513**: 242–245.
- Eason, R. and Campbell, A.M. (1978) 8 - Analytical Ultracentrifugation. In *Centrifugal Separations in Molecular and Cell Biology*. Birnie, G.D. and Rickwood, D. (eds). Butterworth-Heinemann, pp. 251–287.
- Eastman, J.A. and Ferguson, J.F. (1981) Solubilization of Particulate Organic Carbon during the Acid Phase of Anaerobic Digestion. *Journal (Water Pollution Control Federation)* **53**: 352–366.
- Edwards, R.A., McNair, K., Faust, K., Raes, J., and Dutilh, B.E. (2016) Computational approaches to predict bacteriophage–host relationships. *FEMS Microbiology Reviews* **40**: 258–272.
- Eichorst, S.A. and Kuske, C.R. (2012) Identification of Cellulose-Responsive Bacterial and Fungal Communities in Geographically and Edaphically Different Soils by Using Stable Isotope Probing. *Applied and Environmental Microbiology* **78**: 2316–2327.
- Enright, A.J., Van Dongen, S., and Ouzounis, C.A. (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* **30**: 1575–1584.
- Enzmann, F., Mayer, F., Rother, M., and Holtmann, D. (2018) Methanogens: biochemical background and biotechnological applications. *AMB Express* **8**.
- Eryildiz, B., Lukitawesa, and Taherzadeh, M.J. (2020) Effect of pH, substrate loading, oxygen, and methanogens inhibitors on volatile fatty acid (VFA) production from citrus

- waste by anaerobic digestion. *Bioresource Technology* **302**: 122800.
- Escudié, F., Auer, L., Bernard, M., Mariadassou, M., Cauquil, L., Vidal, K., et al. (2018) FROGS: Find, Rapidly, OTUs with Galaxy Solution. *Bioinformatics* **34**: 1287–1294.
- Evans, P.N., Boyd, J.A., Leu, A.O., Woodcroft, B.J., Parks, D.H., Hugenholtz, P., and Tyson, G.W. (2019) An evolving view of methane metabolism in the Archaea. *Nature Reviews Microbiology*.
- Evans, P.N., Parks, D.H., Chadwick, G.L., Robbins, S.J., Orphan, V.J., Golding, S.D., and Tyson, G.W. (2015) Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. *Science* **350**: 434–438.
- Eydal, H.S.C., Jägevall, S., Hermansson, M., and Pedersen, K. (2009) Bacteriophage lytic to *Desulfovibrio aespoensis* isolated from deep groundwater. *ISME J* **3**: 1139–1147.
- Fdz-Polanco, F., Fdz-Polanco, M., Fernández, N., Urueña, M.A., García, P.A., and Villaverde, S. (2001) Combining the biological nitrogen and sulfur cycles in anaerobic conditions. *Water Science and Technology* **44**: 77–84.
- Fokine, A. and Rossmann, M.G. (2014) Molecular architecture of tailed double-stranded DNA phages. *Bacteriophage* **4**: e28281.
- Folimonova, S.Y. (2012) Superinfection Exclusion Is an Active Virus-Controlled Function That Requires a Specific Viral Protein. *Journal of Virology* **86**: 5554–5561.
- Forterre, P. (2011) Manipulation of cellular syntheses and the nature of viruses: The virocell concept. *Comptes Rendus Chimie* **14**: 392–399.
- Fox, G.E., Pechman, K.R., and Woese, C.R.Y. 1977 (1977) Comparative Cataloging of 16S Ribosomal Ribonucleic Acid: Molecular Approach to Prokaryotic Systematics. *International Journal of Systematic and Evolutionary Microbiology* **27**: 44–57.
- Frossard, A., Hammes, F., and Gessner, M.O. (2016) Flow Cytometric Assessment of Bacterial Abundance in Soils, Sediments and Sludge. *Front Microbiol* **7**..
- Galiez, C., Siebert, M., Enault, F., Vincent, J., and Söding, J. (2017) WIsH: who is the host? Predicting prokaryotic hosts from metagenomic phage contigs. *Bioinformatics* **33**: 3113–3114.
- Gaudin, R. and Barteneva, N.S. (2015) Sorting of small infectious virus particles by flow virometry reveals distinct infectivity profiles. *Nat Commun* **6**: 6022.
- Gazitúa, M.C., Vik, D.R., Roux, S., Gregory, A.C., Bolduc, B., Widner, B., et al. (2021) Potential virus-mediated nitrogen cycling in oxygen-depleted oceanic waters. *ISME J* **15**: 981–998.
- Godon, J.-J., Bize, A., Ngo, H., Cauquil, L., Almeida, M., Petit, M.-A., and Zemb, O. (2021) Bacterial Consumption of T4 Phages. *Microorganisms* **9**: 1852.
- Graham, E.B., Paez-Espino, D., Brislawn, C., Hofmockel, K.S., Wu, R., Kyrpides, N.C., et al. (2019) Untapped viral diversity in global soil metagenomes. 583997.

- Guo, J., Bolduc, B., Zayed, A.A., Varsani, A., Dominguez-Huerta, G., Delmont, T.O., et al. (2021a) VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* **9**: 37.
- Haichar, F. el Z., Roncato, M.-A., and Achouak, W. (2012) Stable isotope probing of bacterial community structure and gene expression in the rhizosphere of *Arabidopsis thaliana*. *FEMS Microbiology Ecology* **81**: 291–302.
- Handley, J., Adams, V., and Akagi, J.M. (1973) Morphology of Bacteriophage-Like Particles from *Desulfovibrio vulgaris*. *Journal of Bacteriology* **115**: 1205–1207.
- Hargreaves, K.R. and Clokie, M.R.J. (2015) A Taxonomic Review of *Clostridium difficile* Phages and Proposal of a Novel Genus, "Phimmp04likevirus." *Viruses* **7**: 2534–2541.
- Harirchi, S., Wainaina, S., Sar, T., Nojoudi, S.A., Parchami, Milad, Parchami, Mohsen, et al. (2022) Microbiological insights into anaerobic digestion for biogas, hydrogen or volatile fatty acids (VFAs): a review. *Bioengineered* **13**: 6521–6557.
- He, J., Wang, X., Yin, X., Li, Q., Li, X., Zhang, Y., and Deng, Y. (2018) Insights into biomethane production and microbial community succession during semi-continuous anaerobic digestion of waste cooking oil under different organic loading rates. *AMB Express* **8**: 92.
- Hejna, J.A. and Moses, R.E. (2009) DNA Replication. In *Encyclopedia of Microbiology (Third Edition)*. Schaechter, M. (ed). Oxford: Academic Press, pp. 113–122.
- Hernández, S. and Vives, M.J. (2020) Phages in Anaerobic Systems. *Viruses* **12**: 1091.
- Herrmann, S., Kleinstuber, S., Chatzinotas, A., Kuppardt, S., Lueders, T., Richnow, H.-H., and Vogt, C. (2010) Functional characterization of an anaerobic benzene-degrading enrichment culture by DNA stable isotope probing. *Environmental Microbiology* **12**: 401–411.
- Hille, F., Richter, H., Wong, S.P., Bratovič, M., Ressel, S., and Charpentier, E. (2018) The Biology of CRISPR-Cas: Backward and Forward. *Cell* **172**: 1239–1259.
- Hinnebusch, J. and Tilly, K. (1993) Linear plasmids and chromosomes in bacteria. *Mol Microbiol* **10**: 917–922.
- Höck, J. and Meister, G. (2008) The Argonaute protein family. *Genome Biol* **9**: 210.
- Huber, H., Hohn, M.J., Stetter, K.O., and Rachel, R. (2003) The phylum Nanoarchaeota: Present knowledge and future perspectives of a unique form of life. *Research in Microbiology* **154**: 165–171.
- Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**: 119.
- Inoue, K. and Iida, H. (1968) Bacteriophages of *Clostridium botulinum*. *J Virol* **2**: 537–540.
- Iranzo, J., Koonin, E.V., Prangishvili, D., and Krupovic, M. (2016) Bipartite Network Analysis

- of the Archaeal Virosphere: Evolutionary Connections between Viruses and Capsidless Mobile Elements. *J Virol* **90**: 11043–11055.
- Iranzo, J., Krupovic, M., and Koonin, E.V. (2016) The Double-Stranded DNA Virosphere as a Modular Hierarchical Network of Gene Sharing. *mBio* **7**: e00978-16.
- Jackson, E.F. and Jackson, C.R. (2008) Viruses in wetland ecosystems. *Freshwater Biology* **53**: 1214–1227.
- Jahn, M.T., Lachnit, T., Markert, S.M., Stigloher, C., Pita, L., Ribes, M., et al. (2021) Lifestyle of sponge symbiont phages by host prediction and correlative microscopy. *ISME J* **15**: 2001–2011.
- Jang, H.B., Chittick, L., Li, Y.-F., Zablocki, O., Sanderson, C.M., Carrillo, A., et al. (2022) Viral tag and grow: a scalable approach to capture and characterize infectious virus–host pairs. *ISME COMMUN* **2**: 1–11.
- Jiang, W., Wu, N., Wang, X., Chi, Y., Zhang, Y., Qiu, X., et al. (2015) Dysbiosis gut microbiota associated with inflammation and impaired mucosal immune function in intestine of humans with non-alcoholic fatty liver disease. *Sci Rep* **5**: 8096.
- Jones, M.D., Crandell, D.W., Singleton, D.R., and Aitken, M.D. (2011) Stable-isotope probing of the polycyclic aromatic hydrocarbon-degrading bacterial guild in a contaminated soil. *Environmental Microbiology* **13**: 2623–2632.
- Kamimura, K. and Araki, M. (1989) Isolation and Characterization of a Bacteriophage Lytic for *Desulfovibrio salexigens*, a Salt-Requiring, Sulfate-Reducing Bacterium. *Applied and Environmental Microbiology* **55**: 645–648.
- Kanehisa, M., Sato, Y., and Morishima, K. (2016) BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology* **428**: 726–731.
- Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., and Wang, Z. (2019a) MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**: e7359.
- Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., and Wang, Z. (2019b) MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**: e7359.
- Kazda, M., Langer, S., and Bengelsdorf, F.R. (2014) Fungi open new possibilities for anaerobic fermentation of organic residues. *Energ Sustain Soc* **4**: 6.
- Kieft, K., Zhou, Z., and Anantharaman, K. (2020a) VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* **8**: 90.
- Kim, J.-G., Kim, S.-J., Cvirkaite-Krupovic, V., Yu, W.-J., Gwak, J.-H., López-Pérez, M., et al. (2019) Spindle-shaped viruses infect marine ammonia-oxidizing thaumarchaea. *PNAS* **116**: 15645–15650.

- Kim, M., Ahn, Y.-H., and Speece, R.E. (2002) Comparative process stability and efficiency of anaerobic digestion; mesophilic vs. thermophilic. *Water Res* **36**: 4369–4385.
- Kim, Y.M., Jang, H.M., Lee, K., Chantrasakdakul, P., Kim, D., and Park, K.Y. (2015) Changes in bacterial and archaeal communities in anaerobic digesters treating different organic wastes. *Chemosphere* **141**: 134–137.
- Kirschke, S., Bousquet, P., Ciais, P., Saunois, M., Canadell, J.G., Dlugokencky, E.J., et al. (2013) Three decades of global methane sources and sinks. *Nature Geosci* **6**: 813–823.
- Klumpp, K. and Crépin, T. (2014) Capsid proteins of enveloped viruses as antiviral drug targets. *Current Opinion in Virology* **5**: 63–71.
- Koonin, E.V., Krupovic, M., and Agol, V.I. (2021) The Baltimore Classification of Viruses 50 Years Later: How Does It Stand in the Light of Virus Evolution? *Microbiol Mol Biol Rev* **85**: e0005321.
- Koonin, E.V. and Yutin, N. (2019) Chapter Five - Evolution of the Large Nucleocytoplasmic DNA Viruses of Eukaryotes and Convergent Origins of Viral Gigantism. In *Advances in Virus Research*. Kielian, M., Mettenleiter, T.C., and Roossinck, M.J. (eds). Academic Press, pp. 167–202.
- Koonin, E.V. and Yutin, N. (2020) The crAss-like Phage Group: How Metagenomics Reshaped the Human Virome. *Trends Microbiol* **28**: 349–359.
- Köster, J. and Rahmann, S. (2012) Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* **28**: 2520–2522.
- Kotay, S.M., Datta, T., Choi, J., and Goel, R. (2011) Biocontrol of biomass bulking caused by *Haliscomenobacter hydrossis* using a newly isolated lytic bacteriophage. *Water Research* **45**: 694–704.
- Krupovič, M. and Bamford, D.H. (2008) Archaeal proviruses TKV4 and MVV extend the PRD1-adenovirus lineage to the phylum Euryarchaeota. *Virology* **375**: 292–300.
- Krupovic, M., Cvirkaite-Krupovic, V., Iranzo, J., Prangishvili, D., and Koonin, E.V. (2018) Viruses of archaea: Structural, functional, environmental and evolutionary genomics. *Virus Res* **244**: 181–193.
- Krupovic, M., Dolja, V.V., and Koonin, E.V. (2020) The LUCA and its complex virome. *Nat Rev Microbiol* **18**: 661–670.
- Krupovič, M., Forterre, P., and Bamford, D.H. (2010a) Comparative Analysis of the Mosaic Genomes of Tailed Archaeal Viruses and Proviruses Suggests Common Themes for Virion Architecture and Assembly with Tailed Viruses of Bacteria. *Journal of Molecular Biology* **397**: 144–160.
- Krupovič, M., Gribaldo, S., Bamford, D.H., and Forterre, P. (2010a) The Evolutionary History of Archaeal MCM Helicases: A Case Study of Vertical Evolution Combined with Hitchhiking of Mobile Genetic Elements. *Molecular Biology and Evolution* **27**: 2716–2732.

- Krupovic, M., Quemin, E.R.J., Bamford, D.H., Forterre, P., and Prangishvili, D. (2014) Unification of the Globally Distributed Spindle-Shaped Viruses of the Archaea. *Journal of Virology*.
- Kumar, S., Cheng, X., Klimasauskas, S., Mi, S., Posfai, J., Roberts, R.J., and Wilson, G.G. (1994) The DNA (cytosine-5) methyltransferases. *Nucleic Acids Res* **22**: 1–10.
- Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–359.
- Lara, E., Vaqué, D., Sà, E.L., Boras, J.A., Gomes, A., Borrull, E., et al. Unveiling the role and life strategies of viruses from the surface to the dark ocean. *Science Advances* **3**: e1602565.
- L. Bräuer, S., Basiliko, N., M. P. Siljanen, H., and H. Zinder, S. (2020) Methanogenic archaea in peatlands. *FEMS Microbiology Letters* **367**: fnaa172.
- Lei, Y., Wei, L., Liu, T., Xiao, Y., Dang, Y., Sun, D., and Holmes, D.E. (2018) Magnetite enhances anaerobic digestion and methanogenesis of fresh leachate from a municipal solid waste incineration plant. *Chemical Engineering Journal* **348**: 992–999.
- Leite, D.M.C., Brochet, X., Resch, G., Que, Y.-A., Neves, A., and Peña-Reyes, C. (2018) Computational prediction of inter-species relationships through omics data analysis and machine learning. *BMC Bioinformatics* **19**: 420.
- Lemaire, O.N., Jespersen, M., and Wagner, T. (2020) CO<sub>2</sub>-Fixation Strategies in Energy Extremophiles: What Can We Learn From Acetogens? *Frontiers in Microbiology* **11**.
- Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**: 1674–1676.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009a) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.
- Li, M., Wang, Y., Li, F., Zhao, Y., Liu, M., Zhang, S., et al. (2021) A Deep Learning-Based Method for Identification of Bacteriophage-Host Interaction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **18**: 1801–1810.
- Li, T., Mazéas, L., Sghir, A., Leblon, G., and Bouchez, T. (2009) Insights into networks of functional microbes catalysing methanization of cellulose under mesophilic conditions. *Environ Microbiol* **11**: 889–904.
- Li, Y., Gordon, E., Shean, R.C., Idle, A., Deng, X., Greninger, A.L., and Delwart, E. (2021) CrAssphage and its bacterial host in cat feces. *Sci Rep* **11**: 815.
- Liang, B., Wang, L.-Y., Mbadinga, S.M., Liu, J.-F., Yang, S.-Z., Gu, J.-D., and Mu, B.-Z. (2015) Anaerolineaceae and Methanosaeta turned to be the dominant microorganisms in alkanes-dependent methanogenic culture after long-term of incubation. *AMB Express* **5**: 37.
- Lin, Q., De Vrieze, J., Li, J., and Li, X. (2016) Temperature affects microbial abundance,

- activity and interactions in anaerobic digestion. *Bioresource Technology* **209**: 228–236.
- Liou, J.S.-C. and Madsen, E.L. (2008) Microbial Ecological Processes: Aerobic/Anaerobic. In *Encyclopedia of Ecology*. Jørgensen, S.E. and Fath, B.D. (eds). Oxford: Academic Press, pp. 2348–2357.
- Liu, D., Ma, Y., Jiang, X., and He, T. (2019) Predicting virus-host association by Kernelized logistic matrix factorization and similarity network fusion. *BMC Bioinformatics* **20**: 594.
- Liu, Y., Demina, T.A., Roux, S., Aiewsakun, P., Kazlauskas, D., Simmonds, P., et al. (2021) Diversity, taxonomy, and evolution of archaeal viruses of the class Caudoviricetes. *PLOS Biology* **19**: e3001442.
- Lopatina, A., Tal, N., and Sorek, R. (2020) Abortive Infection: Bacterial Suicide as an Antiviral Immune Strategy. *Annual Review of Virology* **7**: 371–384.
- Lopes, A., Tavares, P., Petit, M.-A., Guérois, R., and Zinn-Justin, S. (2014) Automated classification of tailed bacteriophages according to their neck organization. *BMC Genomics* **15**: 1027.
- Łoś, M. and Węgrzyn, G. (2012) Pseudolysogeny. *Adv Virus Res* **82**: 339–349.
- Lu, C., Zhang, Z., Cai, Z., Zhu, Z., Qiu, Y., Wu, A., et al. (2021) Prokaryotic virus host predictor: a Gaussian model for host prediction of prokaryotic viruses in metagenomics. *BMC Biology* **19**: 5.
- Luo, Y., Pfister, P., Leisinger, T., and Wasserfallen, A. (2001) The Genome of Archaeal Prophage  $\Psi$ M100 Encodes the Lytic Enzyme Responsible for Autolysis of *Methanothermobacter wolfeii*. *Journal of Bacteriology*.
- Lv, Z., Chen, Z., Chen, X., Liang, J., Jiang, J., and Loake, G.J. (2019) Effects of various feedstocks on isotope fractionation of biogas and microbial community structure during anaerobic digestion. *Waste Management* **84**: 211–219.
- Lyu, Z., Shao, N., Akinyemi, T., and Whitman, W.B. (2018) Methanogenesis. *Current Biology* **28**: R727–R732.
- Madigou, C., Lê Cao, K.-A., Bureau, C., Mazéas, L., Déjean, S., and Chapleur, O. (2019) Ecological consequences of abrupt temperature changes in anaerobic digesters. *Chemical Engineering Journal* **361**: 266–277.
- Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J.J., Charpentier, E., Horvath, P., et al. (2011) Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* **9**: 467–477.
- Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., et al. (2020) Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat Rev Microbiol* **18**: 67–83.
- Manrique, P., Bolduc, B., Walk, S.T., van der Oost, J., de Vos, W.M., and Young, M.J. (2016)



- Healthy human gut phageome. *Proceedings of the National Academy of Sciences* **113**: 10400–10405.
- Mao, C., Feng, Y., Wang, X., and Ren, G. (2015) Review on research achievements of biogas from anaerobic digestion. *Renewable and Sustainable Energy Reviews* **45**: 540–555.
- Mara, P., Vik, D., Pachiadaki, M.G., Suter, E.A., Poulos, B., Taylor, G.T., et al. (2020) Viral elements and their potential influence on microbial processes along the permanently stratified Cariaco Basin redoxcline. *ISME J* **14**: 3079–3092.
- Marbouty, M., Cournac, A., Flot, J.-F., Marie-Nelly, H., Mozziconacci, J., and Koszul, R. (2014) Metagenomic chromosome conformation capture (meta3C) unveils the diversity of chromosome organization in microorganisms. *Elife* **3**: e03318.
- Marbouty, M. and Koszul, R. (2015) Metagenome Analysis Exploiting High-Throughput Chromosome Conformation Capture (3C) Data. *Trends in Genetics* **31**: 673–682.
- Maron, P.-A., Ranjard, L., Mougel, C., and Lemanceau, P. (2007) Metaproteomics: A New Approach for Studying Functional Microbial Ecology. *Microb Ecol* **53**: 486–493.
- Martineau, C., Whyte, L.G., and Greer, C.W. (2010) Stable Isotope Probing Analysis of the Diversity and Activity of Methanotrophic Bacteria in Soils from the Canadian High Arctic. *Applied and Environmental Microbiology* **76**: 5773–5784.
- Maxam, A.M. and Gilbert, W. (1977) A new method for sequencing DNA. *Proceedings of the National Academy of Sciences* **74**: 560–564.
- Mayumi, D., Mochimaru, H., Tamaki, H., Yamamoto, K., Yoshioka, H., Suzuki, Y., et al. (2016) Methane production from coal by a single methanogen. *Science* **354**: 222–225.
- McDonald, W.H. and Yates III, J.R. (2002) Shotgun proteomics and biomarker discovery. *Disease Markers* **18**: 99–105.
- von Meijenfeldt, F.A.B., Arkhipova, K., Cambuy, D.D., Coutinho, F.H., and Dutilh, B.E. (2019a) Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. *Genome Biology* **20**: 217.
- Meile, L., Jenal, U., Studer, D., Jordan, M., and Leisinger, T. (1989) Characterization of  $\psi$ M1, a virulent phage of *Methanobacterium thermoautotrophicum* Marburg. *Arch Microbiol* **152**: 105–110.
- Menzel, P., Ng, K.L., and Krogh, A. (2016) Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications* **7**: 11257.
- Meyer, J.R., Dobias, D.T., Weitz, J.S., Barrick, J.E., Quick, R.T., and Lenski, R.E. (2012) Repeatability and Contingency in the Evolution of a Key Innovation in Phage Lambda. *Science* **335**: 428–432.
- Mirzaei, M.K. and Maurice, C.F. (2017) Ménage à trois in the human gut: interactions between host, bacteria and phages. *Nat Rev Microbiol* **15**: 397–408.
- Moletta, R. (2015) La méthanisation, 3<sup>è</sup>. Lavoisier.

- Molnár, J., Magyar, B., Schneider, G., Laczi, K., Valappil, S.K., Kovács, Á.L., et al. (2020) Identification of a novel archaea virus, detected in hydrocarbon polluted Hungarian and Canadian samples. *PLOS ONE* **15**: e0231864.
- Mora, C., Tittensor, D.P., Adl, S., Simpson, A.G.B., and Worm, B. (2011) How Many Species Are There on Earth and in the Ocean? *PLOS Biology* **9**: e1001127.
- Morales, C.A., Oakley, B.B., Garrish, J.K., Siragusa, G.R., Ard, M.B., and Seal, B.S. (2012) Complete genome sequence of the podoviral bacteriophage ΦCP24R, which is virulent for *Clostridium perfringens*. *Arch Virol* **157**: 769–772.
- Morella, N.M., Yang, S.C., Hernandez, C.A., and Koskella, B. (2018) Rapid quantification of bacteriophages and their bacterial hosts in vitro and in vivo using droplet digital PCR. *Journal of Virological Methods* **259**: 18–24.
- Mosbæk, F., Kjeldal, H., Mulat, D.G., Albertsen, M., Ward, A.J., Feilberg, A., and Nielsen, J.L. (2016) Identification of syntrophic acetate-oxidizing bacteria in anaerobic digesters by combined protein-based stable isotope probing and metagenomics. *ISME J* **10**: 2405–2418.
- Mosier-Boss, P.A., Lieberman, S.H., Andrews, J.M., Rohwer, F.L., Wegley, L.E., and Breitbart, M. (2003) Use of Fluorescently Labeled Phage in the Detection and Identification of Bacterial Species. *Appl Spectrosc, AS* **57**: 1138–1144.
- Murin, C.D., Wilson, I.A., and Ward, A.B. (2019) Antibody responses to viral infections: a structural perspective across three different enveloped viruses. *Nat Microbiol* **4**: 734–747.
- Mushegian, A.R. (2020) Are There 1031 Virus Particles on Earth, or More, or Fewer? *Journal of Bacteriology* **202**: e00052-20.
- Nakasaki, K., Nguyen, K.K., Ballesteros, F.C., Maekawa, T., and Koyama, M. (2020) Characterizing the microbial community involved in anaerobic digestion of lipid-rich wastewater to produce methane gas. *Anaerobe* **61**: 102082.
- Nale, J.Y., Spencer, J., Hargreaves, K.R., Buckley, A.M., Trzepiński, P., Douce, G.R., and Clokie, M.R.J. (2016) Bacteriophage Combinations Significantly Reduce *Clostridium difficile* Growth In Vitro and Proliferation In Vivo. *Antimicrob Agents Chemother* **60**: 968–981.
- Nasko, D.J., Ferrell, B.D., Moore, R.M., Bhavsar, J.D., Polson, S.W., and Wommack, K.E. (2019) CRISPR Spacers Indicate Preferential Matching of Specific Virioplankton Genes. *mBio*.
- Nayfach, S., Camargo, A.P., Eloë-Fadrosh, E., Roux, S., and Kyrpides, N. (2021) CheckV: assessing the quality of metagenome-assembled viral genomes. *Nat Biotechnol* **39**: 578–585.
- Neufeld, J.D., Schäfer, H., Cox, M.J., Boden, R., McDonald, I.R., and Murrell, J.C. (2007) Stable-isotope probing implicates *Methylophaga* spp and novel

- Gammaproteobacteria in marine methanol and methylamine metabolism. *ISME J* **1**: 480–491.
- Neufeld, J.D., Vohra, J., Dumont, M.G., Lueders, T., Manefield, M., Friedrich, M.W., and Murrell, J.C. (2007) DNA stable-isotope probing. *Nat Protoc* **2**: 860–866.
- Ngo, V.-Q.-H., Enault, F., Midoux, C., Mariadassou, M., Chapleur, O., Mazéas, L., et al. (2022) Diversity of novel archaeal viruses infecting methanogens discovered through coupling of stable isotope probing and metagenomics. *Environ Microbiol.*
- Nölling, J., Groffen, A., and de Vos, W.M. (1993)  $\phi$  F1 and  $\phi$  F3, two novel virulent, archaeal phages infecting different thermophilic strains of the genus *Methanobacterium*. *Microbiology* **139**: 2511–2516.
- Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P. (2016) metaSPAdes: a new versatile de novo metagenomics assembler. *arXiv:160403071 [q-bio]*.
- Olo Ndela, E., Enault, F., and Toussaint, A. (2021) Transposable Prophages in *Leptospira*: An Ancient, Now Diverse, Group Predominant in Causative Agents of Weil's Disease. *International Journal of Molecular Sciences* **22**: 13434.
- Ortmann, A.C. and Suttle, C.A. (2005) High abundances of viruses in a deep-sea hydrothermal vent system indicates viral mediated microbial mortality. *Deep Sea Research Part I: Oceanographic Research Papers* **52**: 1515–1527.
- Ostenfeld, L.J., Munk, P., Aarestrup, F.M., and Otani, S. (2022) Detection of specific uncultured bacteriophages by fluorescence in situ hybridisation in pig microbiome. 2022.02.18.481118.
- Owen, S.V., Wenner, N., Canals, R., Makumi, A., Hammarlöf, D.L., Gordon, M.A., et al. (2017) Characterization of the Prophage Repertoire of African *Salmonella* Typhimurium ST313 Reveals High Levels of Spontaneous Induction of Novel Phage BTP1. *Frontiers in Microbiology* **8**.
- Pal, C., Maciá, M.D., Oliver, A., Schachar, I., and Buckling, A. (2007) Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature* **450**: 1079–1081.
- Palm, P., Schleper, C., Grampp, B., Yeats, S., McWilliam, P., Reiter, W.-D., and Zillig, W. (1991) Complete nucleotide sequence of the virus SSV1 of the archaeobacterium *Sulfolobus shibatae*. *Virology* **185**: 242–250.
- Pappas, N., Roux, S., Hölzer, M., Lamkiewicz, K., Mock, F., Marz, M., and Dutilh, B.E. (2021) Virus Bioinformatics. *Encyclopedia of Virology* 124–132.
- Park, G.W., Ng, T.F.F., Freeland, A.L., Marconi, V.C., Boom, J.A., Staat, M.A., et al. (2020) CrAssphage as a Novel Tool to Detect Human Fecal Contamination on Environmental Surfaces and Hands. *Emerg Infect Dis* **26**: 1731–1739.
- Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**: 1043–1055.

- Parks, D.H., Rinke, C., Chuvochina, M., Chaumeil, P.-A., Woodcroft, B.J., Evans, P.N., et al. (2017) Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* **2**: 1533–1542.
- Pavlostathis, S.G. and Giraldo-Gomez, E. (1991) Kinetics of Anaerobic Treatment. *Water Sci Technol* **24**: 35–59.
- Pfister, P., Wasserfallen, A., Stettler, R., and Leisinger, T. (1998) Molecular analysis of Methanobacterium phage  $\Psi$ M2. *Molecular Microbiology* **30**: 233–244.
- Philosof, A., Yutin, N., Flores-Uribe, J., Sharon, I., Koonin, E.V., and Béjà, O. (2017) Novel Abundant Oceanic Viruses of Uncultured Marine Group II Euryarchaeota. *Curr Biol* **27**: 1362–1368.
- Pickard, D.J.J. (2009) Preparation of bacteriophage lysates and pure DNA. *Methods Mol Biol* **502**: 3–9.
- Pina, M., Bize, A., Forterre, P., and Prangishvili, D. (2011) The archeoviruses. *FEMS Microbiol Rev* **35**: 1035–1054.
- Poirier, S., Bize, A., Bureau, C., Bouchez, T., and Chapleur, O. (2016) Community shifts within anaerobic digestion microbiota facing phenol inhibition: Towards early warning microbial indicators? *Water Research* **100**: 296–305.
- Poirier, S., Déjean, S., and Chapleur, O. (2018) Support media can steer methanogenesis in the presence of phenol through biotic and abiotic effects. *Water Research* **140**: 24–33.
- Poirier, S., Desmond-Le Quéméner, E., Madigou, C., Bouchez, T., and Chapleur, O. (2016) Anaerobic digestion of biowaste under extreme ammonia concentration: Identification of key microbial phylotypes. *Bioresour Technol* **207**: 92–101.
- Poirier, S., Madigou, C., Bouchez, T., and Chapleur, O. (2017) Improving anaerobic digestion with support media: Mitigation of ammonia inhibition and effect on microbial communities. *Bioresource Technology* **235**: 229–239.
- Pons, J.C., Paez-Espino, D., Riera, G., Ivanova, N., Kyrpides, N.C., and Llabrés, M. (2021) VPF-Class: taxonomic assignment and host prediction of uncultivated viruses based on viral protein families. *Bioinformatics* **37**: 1805–1813.
- Prangishvili, D., Bamford, D.H., Forterre, P., Iranzo, J., Koonin, E.V., and Krupovic, M. (2017) The enigmatic archaeal virosphere. *Nature Reviews Microbiology* **15**: 724–739.
- Puig-Castellví, F., Midoux, C., Guenne, A., Conteau, D., Franchi, O., Bureau, C., et al. (2022) Metataxonomics, metagenomics and metabolomics analysis of the influence of temperature modification in full-scale anaerobic digesters. *Bioresource Technology* **346**: 126612.
- Radajewski, S., Ineson, P., Parekh, N.R., and Murrell, J.C. (2000) Stable-isotope probing as a tool in microbial ecology. *Nature* **403**: 646–649.
- Raoult, D. and Forterre, P. (2008) Redefining viruses: lessons from Mimivirus. *Nat Rev*

*Microbiol* **6**: 315–319.

- Rapp, B.J. and Wall, J.D. (1987) Genetic transfer in *Desulfovibrio desulfuricans*. *Proceedings of the National Academy of Sciences* **84**: 9128–9130.
- Redmond, M.C., Valentine, D.L., and Sessions, A.L. (2010) Identification of Novel Methane-, Ethane-, and Propane-Oxidizing Bacteria at Marine Hydrocarbon Seeps by Stable Isotope Probing. *Applied and Environmental Microbiology* **76**: 6412–6422.
- Reeburgh, W.S. (2007) Oceanic Methane Biogeochemistry. *Chem Rev* **107**: 486–513.
- Robinson, C.M. and Pfeiffer, J.K. (2014) Viruses and the Microbiota. *Annual Review of Virology* **1**: 55–69.
- Robledo-Mahón, T., Silva-Castro, G.A., Kuhar, U., Jamnikar-Ciglenc̃ki, U., Barlić-Maganja, D., Aranda, E., and Calvo, C. (2019) Effect of Composting Under Semipermeable Film on the Sewage Sludge Virome. *Microb Ecol* **78**: 895–903.
- Rodríguez-Ramos, J.A., Borton, M.A., McGivern, B.B., Smith, G.J., Solden, L.M., Shaffer, M., et al. (2022) Genome-resolved metaproteomics decodes the microbial and viral contributions to coupled carbon and nitrogen cycling in river sediments. 2022.03.11.484050.
- Roseman, D. and Richardson, R.L. (1969) Isolation of bacteriophage for *Clostridium tetani*. *J Virol* **3**: 350.
- Roux, S., Brum, J.R., Dutilh, B.E., Sunagawa, S., Duhaime, M.B., Loy, A., et al. (2016) Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* **537**: 689–693.
- Roux, S., Hawley, A.K., Torres Beltran, M., Scofield, M., Schwientek, P., Stepanauskas, R., et al. (2014) Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by single-cell- and meta-genomics. *eLife* **3**: e03125.
- Ryu, W.-S. (2017) Discovery and Classification. *Molecular Virology of Human Pathogenic Viruses* 3–20.
- Sakowski, E.G., Arora-Williams, K., Tian, F., Zayed, A.A., Zablocki, O., Sullivan, M.B., and Preheim, S.P. (2021) Interaction dynamics and virus–host range for estuarine actinophages captured by epicPCR. *Nat Microbiol* **6**: 630–642.
- Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences* **74**: 5463–5467.
- Seal, B.S. (2013) Characterization of bacteriophages virulent for *Clostridium perfringens* and identification of phage lytic enzymes as alternatives to antibiotics for potential control of the bacterium. *Poult Sci* **92**: 526–533.
- Si, B., Liu, Z., Zhang, Y., Li, J., Shen, R., Zhu, Z., and Xing, X. (2016) Towards biohythane production from biomass: Influence of operational stage on anaerobic fermentation and microbial community. *International Journal of Hydrogen Energy* **41**: 4429–4438.

- Sieiro, C., Areal-Hermida, L., Pichardo-Gallardo, Á., Almuiña-González, R., de Miguel, T., Sánchez, S., et al. (2020) A Hundred Years of Bacteriophages: Can Phages Replace Antibiotics in Agriculture and Aquaculture? *Antibiotics* **9**: 493.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**: 539.
- Sikora, A., Detman, A., Chojnacka, A., and Błaszczuk, M.K. (2017) Anaerobic Digestion: I. A Common Process Ensuring Energy Flow and the Circulation of Matter in Ecosystems. II. A Tool for the Production of Gaseous Biofuels. *Fermentation Processes*.
- Smallridge, R. (2008) A virus gets a virus. *Nat Rev Microbiol* **6**: 714–714.
- Snyder, J.C., Bolduc, B., and Young, M.J. (2015) 40 Years of archaeal virology: Expanding viral diversity. *Virology* **479–480**: 369–378.
- Söding, J. (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics* **21**: 951–960.
- Spencer, S.J., Tamminen, M.V., Preheim, S.P., Guo, M.T., Briggs, A.W., Brito, I.L., et al. (2016) Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers. *The ISME Journal* **10**: 427–436.
- Steinegger, M., Meier, M., Mirdita, M., Vöhringer, H., Haunsberger, S.J., and Söding, J. (2019) HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* **20**: 1–15.
- Steinegger, M. and Söding, J. (2017) MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* **35**: 1026–1028.
- Stephenson, M. and Stickland, L.H. (1933) Hydrogenase: The bacterial formation of methane by the reduction of one-carbon compounds by molecular hydrogen. *Biochemical Journal* **27**: 1517–1527.
- Stern, A. and Sorek, R. (2011) The phage-host arms race: Shaping the evolution of microbes. *BioEssays* **33**: 43–51.
- Sun, H., Yang, Z., Shi, G., Arhin, S.G., Papadakis, V.G., Goula, M.A., et al. (2021) Methane production from acetate, formate and H<sub>2</sub>/CO<sub>2</sub> under high ammonia level: Modified ADM1 simulation and microbial characterization. *Science of The Total Environment* **783**: 147581.
- Suttle, C.A. (2016) Environmental microbiology: Viral diversity on the global stage. *Nat Microbiol* **1**: 1–2.
- Suttle, C.A. (2007) Marine viruses — major players in the global ecosystem. *Nature Reviews Microbiology* **5**: 801–812.
- Suttle, C.A. (2005) Viruses in the sea. *Nature* **437**: 356–361.

- Tadmor, A.D., Ottesen, E.A., Leadbetter, J.R., and Phillips, R. (2011) Probing Individual Environmental Bacteria for Viruses by Using Microfluidic Digital PCR. *Science* **333**: 58–62.
- Tamaki, H., Zhang, R., Angly, F.E., Nakamura, S., Hong, P.-Y., Yasunaga, T., et al. (2012) Metagenomic analysis of DNA viruses in a wastewater treatment plant in tropical climate. *Environmental Microbiology* **14**: 441–452.
- Terzian, P., Olo Ndela, E., Galiez, C., Lossouarn, J., Pérez Bucio, R.E., Mom, R., et al. (2021) PHROG: families of prokaryotic virus proteins clustered using remote homology. *NAR Genomics and Bioinformatics* **3**: lqab067.
- Thauer, R.K., Kaster, A.-K., Seedorf, H., Buckel, W., and Hedderich, R. (2008) Methanogenic archaea: ecologically relevant differences in energy conservation. *Nat Rev Microbiol* **6**: 579–591.
- Thiroux, S., Dupont, S., Nesbø, C.L., Bienvenu, N., Krupovic, M., L'Haridon, S., et al. (2021) The first head-tailed virus, MFTV1, infecting hyperthermophilic methanogenic deep-sea archaea. *Environmental Microbiology* **23**: 3614–3626.
- Thompson, J.N. (2012) The Role of Coevolution. *Science* **335**: 410–411.
- Tomaru, Y. and Nagasaki, K. (2007) Flow cytometric detection and enumeration of DNA and RNA viruses infecting marine eukaryotic microalgae. *J Oceanogr* **63**: 215–221.
- Tominaga, K., Morimoto, D., Nishimura, Y., Ogata, H., and Yoshida, T. (2020) In silico Prediction of Virus-Host Interactions for Marine Bacteroidetes With the Use of Metagenome-Assembled Genomes. *Front Microbiol* **11**:
- Tran, P.Q. and Anantharaman, K. (2021) Biogeochemistry Goes Viral: towards a Multifaceted Approach To Study Viruses and Biogeochemical Cycling. *mSystems* **6**: e01138-21.
- Tuyishime, P. and Sinumvayo, J.P. (2020) Novel outlook in engineering synthetic methylotrophs and formatotrophs: a course for advancing C1-based chemicals production. *World J Microbiol Biotechnol* **36**: 118.
- Veeravalli, S.S., Lalman, J.A., Chaganti, S.R., and Heath, D.D. (2017) Continuous hydrogen production using upflow anaerobic sludge blanket reactors: effect of organic loading rate on microbial dynamics and H<sub>2</sub> metabolism. *Journal of Chemical Technology & Biotechnology* **92**: 544–551.
- Venkiteshwaran, K., Bocher, B., Maki, J., and Zitomer, D. (2015) Relating Anaerobic Digestion Microbial Community and Process Function: Supplementary Issue: Water Microbiology. *Microbiol Insights* **8s2**: MBI.S33593.
- Vijayavel, K., Fujioka, R., Ebdon, J., and Taylor, H. (2010) Isolation and characterization of Bacteroides host strain HB-73 used to detect sewage specific phages in Hawaii. *Water Res* **44**: 3714–3724.
- Villarroel, J., Kleinheinz, K.A., Jurtz, V.I., Zschach, H., Lund, O., Nielsen, M., and Larsen, M.V.

- (2016) HostPhinder: A Phage Host Prediction Tool. *Viruses* **8**: 116.
- Walker, C.B., Stolyar, S.S., Pinel, N., Yen, H.-C.B., He, Z., Zhou, J., et al. (2006) Recovery of temperate *Desulfovibrio vulgaris* bacteriophage using a novel host strain. *Environmental Microbiology* **8**: 1950–1959.
- Walsh, C.P. and Xu, G.L. (2006) Cytosine Methylation and DNA Repair. In *DNA Methylation: Basic Mechanisms*. Current Topics in Microbiology and Immunology. Doerfler, W. and Böhm, P. (eds). Berlin, Heidelberg: Springer, pp. 283–315.
- Wang, J., Zhao, Z., and Zhang, Y. (2021) Enhancing anaerobic digestion of kitchen wastes with biochar: Link between different properties and critical mechanisms of promoting interspecies electron transfer. *Renewable Energy* **167**: 791–799.
- Wang, K., Yin, J., Shen, D., and Li, N. (2014) Anaerobic digestion of food waste for volatile fatty acids (VFAs) production with different types of inoculum: Effect of pH. *Bioresource Technology* **161**: 395–401.
- Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007) Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Appl Environ Microbiol* **73**: 5261–5267.
- Wang, W., Ren, J., Tang, K., Dart, E., Ignacio-Espinoza, J.C., Fuhrman, J.A., et al. (2020) A network-based integrated framework for predicting virus–prokaryote interactions. *NAR Genomics and Bioinformatics* **2**: lqaa044.
- Wang, X., Sharp, C.E., Jones, G.M., Grasby, S.E., Brady, A.L., and Dunfield, P.F. (2015) Stable-Isotope Probing Identifies Uncultured Planctomycetes as Primary Degradors of a Complex Heteropolysaccharide in Soil. *Applied and Environmental Microbiology* **81**: 4607–4615.
- Wang, Y. and Qian, P.-Y. (2009) Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. *PLoS One* **4**: e7401.
- Webster, G., Rinna, J., Roussel, E.G., Fry, J.C., Weightman, A.J., and Parkes, R.J. (2010) Prokaryotic functional diversity in different biogeochemical depth zones in tidal sediments of the Severn Estuary, UK, revealed by stable-isotope probing. *FEMS Microbiology Ecology* **72**: 179–197.
- Weidenbach, K., Nickel, L., Neve, H., Alkhnbashi, O.S., Künzel, S., Kupczok, A., et al. (2017) Methanosarcina Spherical Virus, a Novel Archaeal Lytic Virus Targeting Methanosarcina Strains. *Journal of Virology* **91**: e00955-17.
- Weidenbach, K., Wolf, S., Kupczok, A., Kern, T., Fischer, M.A., Reetz, J., et al. (2021) Characterization of Blf4, an Archaeal Lytic Virus Targeting a Member of the Methanomicrobiales. *Viruses* **13**: 1934.
- Whiticar, M.J., Faber, E., and Schoell, M. (1986) Biogenic methane formation in marine and freshwater environments: CO<sub>2</sub> reduction vs. acetate fermentation—Isotope



- evidence. *Geochimica et Cosmochimica Acta* **50**: 693–709.
- Willenbücher, K., Wibberg, D., Huang, L., Conrady, M., Ramm, P., Gätcke, J., et al. (2022) Phage Genome Diversity in a Biogas-Producing Microbiome Analyzed by Illumina and Nanopore GridION Sequencing. *Microorganisms* **10**: 368.
- Woese, C.R. (1987) Bacterial evolution. *Microbiol Rev* **51**: 221–271.
- Woese, C.R. and Fox, G.E. (1977) Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences* **74**: 5088–5090.
- Woese, C.R., Fox, G.E., Zablen, L., Uchida, T., Bonen, L., Pechman, K., et al. (1975) Conservation of primary structure in 16S ribosomal RNA. *Nature* **254**: 83–86.
- Woese, C.R., Kandler, O., and Wheelis, M.L. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences* **87**: 4576–4579.
- Wolf, S., Fischer, M.A., Kupczok, A., Reetz, J., Kern, T., Schmitz, R.A., and Rother, M. (2019) Characterization of the lytic archaeal virus Drs3 infecting *Methanobacterium formicicum*. *Arch Virol* **164**: 667–674.
- Wood, A.G., Whitman, W.B., and Konisky, J. (1989) Isolation and characterization of an archaeobacterial viruslike particle from *Methanococcus voltae* A3. *Journal of Bacteriology* **171**: 93–98.
- Wu, Q. and Liu, W.-T. (2009) Determination of virus abundance, diversity and distribution in a municipal wastewater treatment plant. *Water Research* **43**: 1101–1109.
- Yan, Z., Song, Z., Li, D., Yuan, Y., Liu, X., and Zheng, T. (2015) The effects of initial substrate concentration, C/N ratio, and temperature on solid-state anaerobic digestion from composting rice straw. *Bioresource Technology* **177**: 266–273.
- Yin, X., Wu, W., Maeke, M., Richter-Heitmann, T., Kulkarni, A.C., Oni, O.E., et al. (2019) CO<sub>2</sub> conversion to methane and biomass in obligate methylotrophic methanogens in marine sediments. *The ISME Journal* **1**.
- Yishai, O., Lindner, S.N., Gonzalez de la Cruz, J., Tenenboim, H., and Bar-Even, A. (2016) The formate bio-economy. *Current Opinion in Chemical Biology* **35**: 1–9.
- Zamora, J.L.R. and Aguilar, H.C. (2018) Flow virometry as a tool to study viruses. *Methods* **134–135**: 87–97.
- Zander, A., Willkomm, S., Ofer, S., van Wolferen, M., Egert, L., Buchmeier, S., et al. (2017) Guide-independent DNA cleavage by archaeal Argonaute from *Methanocaldococcus jannaschii*. *Nat Microbiol* **2**: 1–10.
- Zhai, N., Zhang, T., Yin, D., Yang, G., Wang, X., Ren, G., and Feng, Y. (2015) Effect of initial pH on anaerobic co-digestion of kitchen waste and cow manure. *Waste Management* **38**: 126–131.
- Zhang, F., Zhou, F., Gan, R., Ren, C., Jia, Y., Yu, L., and Huang, Z. (2020) PHISDetector: a tool

- to detect diverse in silico phage-host interaction signals for virome studies. 661074.
- Zhang, J., Gao, Q., Zhang, Q., Wang, T., Yue, H., Wu, L., et al. (2017) Bacteriophage–prokaryote dynamics and interaction within anaerobic digestion processes across time and space. *Microbiome* **5**: 57.
- Zhang, J., Tian, H., Wang, X., and Tong, Y.W. (2020) Effects of activated carbon on mesophilic and thermophilic anaerobic digestion of food waste: Process performance and life cycle assessment. *Chemical Engineering Journal* **399**: 125757.
- Zhang, R., Mirdita, M., Levy Karin, E., Norroy, C., Galiez, C., and Söding, J. (2021) SpacePHARER: sensitive identification of phages from CRISPR spacers in prokaryotic hosts. *Bioinformatics*.
- Zhang, S., Wang, Y., Song, J., Sheng, C., Shang, Z., Wang, R., et al. (2021) Investigation on the Interactive Effects between Temperature and Chemical Composition of Organic Wastes on Anaerobic Co-Digestion Performance. *Processes* **9**: 1682.
- Zhou, G.-M. and Fang, H.H.P. (1998) Competition between methanogenesis and sulfidogenesis in anaerobic wastewater treatment. *Water Science and Technology* **38**: 317–324.
- Zhu, Y. (2018) Mass spectrometry based proteomics: data analysis and applications.
- Ziemiński, K. and Frąc, M. (2012) Methane fermentation process as anaerobic digestion of biomass: Transformations, stages and microorganisms. *African Journal of Biotechnology* **11**: 4127–4139–4139.
- Zimmerman, A.E., Howard-Varona, C., Needham, D.M., John, S.G., Worden, A.Z., Sullivan, M.B., et al. (2020) Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. *Nat Rev Microbiol* **18**: 21–34.
- Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., et al. (2018) A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *Journal of Molecular Biology* **430**: 2237–2243.
- Zinder, S.H. (1993) Physiological Ecology of Methanogens. In *Methanogenesis: Ecology, Physiology, Biochemistry & Genetics*. Ferry, J.G. (ed). Boston, MA: Springer US, pp. 128–206.
- Zolfo, M., Pinto, F., Asnicar, F., Manghi, P., Tett, A., Bushman, F.D., and Segata, N. (2019) Detecting contamination in viromes using ViromeQC. *Nat Biotechnol* **37**: 1408–1412.
- Zuckermandl, E. and Pauling, L. (1965) Molecules as documents of evolutionary history. *Journal of Theoretical Biology* **8**: 357–366.

# Valorisations scientifiques

## 1. Publications

Pendant ma thèse, j'ai participé aux travaux ayant permis la découverte du phénomène de prédation des particules virales (bactériophages T4) par les bactéries (*Aeromonas*) présentes dans un échantillon de boue de station d'épuration à condition aérobie. Ce phénomène a été mise en évidence par l'approche de marquage isotopique (Stable isotope probing).

Godon, J.-J., Bize, A., Ngo, H., Cauquil, L., Almeida, M., Petit, M.-A., Zemb, O., 2021. Bacterial Consumption of T4 Phages. *Microorganisms* 9, 1852. (DOI: [10.3390/microorganisms9091852](https://doi.org/10.3390/microorganisms9091852))

Pour les travaux de ma thèse, un article a été accepté en juillet 2022 dans *Environmental microbiology*

[Diversity of novel archaeal viruses infecting methanogens discovered through coupling of stable isotope probing and metagenomics \(submitted\)](#)

## 2. Communications

Présentation d'un poster à la conférence « Hands-on phage biotechnology 4th » (Braga-Portugal) en 2019 :

["Stable isotope probing \(SIP\) for phage DNA"](#)

Présentation de 2 posters à la conférence « Phages.fr in Grenoble » en 2019 :

[" A molecular ecology approach using stable isotope probing and metagenomics to study viruses of methanogens' diversity"](#)

["Detection of an archaeal-specific viral family, previously thought to infect only hyperthermophilic archaea, in human gut metaviromes"](#)

Présentation d'un poster à la conférence « AFEM 2019 » (Association francophone

d'écologie microbienne) à Bussang:

« Exploration de la diversité des virus d'archées méthanogènes dans les procédés de digestion anaérobie, par des approches isotopiques et métagénomiques »

2 présentations orales aux séminaires de l'équipe Phage, INRAE Micalis (Jouy-en-Josas)

Présentation d'un poster à la conférence virtuelle « JOBIM 2021 visio » (Journées Ouvertes en Biologie, Informatique et Mathématiques):

"A bioinformatic pipeline to elucidate the links between viruses and their hosts in microbial communities, applied to viruses in anaerobic digestion processes"

Présentation orale à la conférence SFM Microbe 2021, Nantes :

" Characterization of the diversity of viruses infecting methanogens in anaerobic digestion processes by shotgun sequencing metagenomics"

### 3. Enseignement

Pendant ma thèse, j'ai donné 8h de cours sur le traitement biologique des déchets non dangereux aux étudiants de la Licence Professionnelle Biologie analytique et expérimentale à l'Ecole Supérieure des Techniques de Biologie Appliquée (ESTBA) à Paris en 2020.

### 4. Formations

Tout au long de ma thèse, j'ai suivi différentes formations répondant aux besoins pour le développement de ma carrière :

- Formation aux techniques appliquées aux virus : Hands-on Phage International courses (Braga, Portugal 2019)
- Formation à l'éthique de recherche (MOOC 2021)
- Formations sur le traitement et d'analyse des méta-données. (INRAE Migale, Jouy en Josas, 2019 – 2021)
- Formation au projet professionnel (ED Abies, Paris, 2022)