

THESE DE DOCTORAT DE

L'UNIVERSITE DE NANTES
COMUE UNIVERSITE BRETAGNE LOIRE

ECOLE DOCTORALE N° 598
Sciences de la Mer et du littoral
Spécialité : Microbiologie

Par

Sofia STRUBBIA

« Norovirus et huître : infectiosité et approche génomique »

Thèse présentée et soutenue à Nantes, le 04/12/2019

Unité de recherche : Laboratoire Santé Environnement et Microbiologie, SG2M, IFREMER

Rapporteurs avant soutenance :

Albert Bosch Professeur, Université de Barcelone, Espagne
Maël Bessaud Chercheur, Institut Pasteur, Paris

Composition du Jury :

Présidente : Nathalie Ruvoën Professeur, Oniris, Nantes
Examineur : Marion Koopmans Professeur, Erasmus University, Rotterdam, Pays-Bas

Dir. de thèse : Soizick Le Guyader Directrice de recherche, Ifremer, Nantes
Co-dir. de thèse : Jacques Le Pendu Directeur de recherche, Inserm, Nantes

REMERCIEMENTS

Je remercie avant tout **Soizick Le Guyader**, pour m'avoir encadrée depuis ma première arrivée au laboratoire en octobre 2014. Depuis, vous n'avez jamais manqué de dispenser votre aide et vos conseils guidés par une vraie passion pour votre travail, j'ai énormément appris au cours de ces 5 ans. Merci également pour m'avoir donné l'opportunité de me former à travers la participation à de nombreux congrès et formations en France comme à l'étranger. Je tiens en particulier à vous remercier pour vos attentions, votre compréhension et soutien à certains moments difficiles où j'ai pu compter sur vous. Merci !

Je tiens à remercier **Jacques Le Pendu**, co-directeur de thèse, pour les conseils et pour toutes les explications autour des mucines et des glycanes. Merci aussi pour avoir participé à mes différents comités de jury de thèse avec le Pr. **Berthe Marie Imbert-Marcille**, que je remercie également. Je remercie **Sylvie Lapegue** pour son soutien et ces conseils.

Merci à **Adrien Breiman**, ta disponibilité et ton aide ont été précieuses.

Je remercie les rapporteurs pour avoir accepté de lire, évaluer et discuter ce travail malgré leur emploi du temps très chargé : **Albert Bosch** et **Maël Bessaud**.

Merci également aux membres du jury, **Nathalie Ruvoën** qui continue à suivre et encourager les étudiants du master 2 manIMAL, expérience enrichissante que je porte encore dans mon cœur ; et merci à **Marion Koopmans** pour être ici aujourd'hui et pour avoir accepté que je réalise une partie de mes travaux de thèse dans son laboratoire l'année dernière. Pendant trois mois j'ai pu utiliser le matériel et les infrastructures du laboratoire ViroScience (EMC) et bénéficier de l'expertise et des conseils des collègues néerlandais. J'en profite pour remercier **Miranda, Bas, David, Nele, Swheta, Jasmine, Emmanuelle, My Phan** et **Matt**. Si mon séjour à Rotterdam a été professionnellement et humainement riche, c'est grâce à vous !

Le plus grand et le plus chaleureux de mes mercis va à toute **l'équipe LSEM Nantes et Brest**. Merci pour votre accueil, vos sourires, votre aide constante, votre esprit de cohésion et de partage. En anglais, « franglais » jusqu'au verlan, j'ai passé avec vous 5 ans merveilleux et vous quitter me fait mal au cœur.

Un merci particulier aux *runners* de l'équipe : **Antoine, Sylvain, Jean-Côme, Cédric**, pour avoir partagé les bords de l'Erdre et du Cens. Un énorme merci à **Julien**, ange gardien de cette thèse, ton aide et ta disponibilité n'ont pas de limites et ce travail n'aurait pas été le même sans toi ! Merci à **Marion**, pour ta rigueur scientifique, ta disponibilité et tes conseils qui ont été d'une grande aide lors des manips et aussi pendant la rédaction de ce manuscrit. Merci à **Jessica** pour ta patience et ta disponibilité dans l'organisation des différentes missions et aussi pour m'avoir aidé dans la mise en forme de ce manuscrit. Merci à **Françoise** pour ton côté calme et rassurant et pour m'avoir initié à la couture ! Merci à **Joanna** pour ton optimisme et pour voir toujours la vie *en rose*. Merci à **Candice** pour ta gentillesse et ta spontanéité et pour avoir accepté de relire et « remettre en français » plusieurs parties de ce

manuscrit.

Merci à **Sophie, Gaëlle, Pascal, Emilie, Adeline, Solène** ... Merci à vous tous ! Ça va être dur de retrouver une autre équipe comme celle-ci !!!

J'exprime aussi une grande reconnaissance à **Alban**, ton séjour au LSEM n'a pas été long mais indispensable pour ce travail. Et si la bioinfo me fait un peu moins peur, c'est aussi grâce à toi ! Merci pour ton partage. Merci à **David** qui m'a initié à la culture cellulaire, je garde un bon souvenir de nos essais de *rolling bottles*, de ton côté pédagogique et festif.

Un remerciement tout particulier pour **Cécile**, collègue mais surtout une grande amie. Je sais que je pourrai toujours compter sur toi, et toi sur moi. Nos balades matinales style *Walt-Disney* vont me manquer énormément... mais la fin de cette thèse ne suffira pas pour te débarrasser de moi ;))

Dans les bureaux d'Ifremer une autre belle amitié se cache, merci **Sophie** pour tes *Ferrero Rochers* qui me faisaient courir vers ton bureau et pour être toujours disponible à m'aider avec le côté administratif français, si mystérieux pour moi !

Merci à **Pierre et Roselyne** de la crêperie *Fleur de sel*, pour leur soutien, leurs encouragements et pour avoir fait de moi une « chercheuse-crêpière ». C'est avec une grande satisfaction que j'ai travaillé pour et avec vous.

Merci à tous les **loulous** et tous les **kikous** qui en dehors du portail d'Ifremer ont énormément participé à ce travail. Eh bien oui, les soirées avec vous, les rigolades, les sorties et les bons moments partagés ont contribué au bien-être de la thésarde donc de ces recherches! J'ai beaucoup de chance, j'ai trouvé en vous des amis exceptionnels ! En particulier un grand merci à **Nadège**, pour m'avoir toujours soutenue et pour avoir cru en ce travail.

Merci à ma famille, qui m'a toujours soutenue dans mes choix, même quand cela a impliqué de s'éloigner et d'être parfois nostalgique. En Italie, comme en Nouvelle Zélande et en Argentine, vous êtes tous dans mon cœur et ce travail vous appartient aussi. Sébastien, Enea et Anna, une pensée spéciale à vous mes petits amours, vous me manquez tellement !

Enfin merci au dernier beau cadeau que cette thèse m'a offert. Parler de gastroentérite, échantillons de selles et assainissement des eaux usées peut aussi amener vers de très belles rencontres et vers des paysages inattendus ! J'ai hâte d'en découvrir tant d'autres à tes côtés.

*Je dédie cette thèse à ma mère,
qui m'a transmis l'enthousiasme et la détermination à travers ses plus beaux sourires.*

AVANT-PROPOS

Ces travaux de thèse ont été réalisés sous la direction de Soizick Le Guyader dans le Laboratoire Santé Environnement et Microbiologie (LSEM) de l'IFREMER (centre de Nantes) et ont été co-encadrés par Jacques Le Pendu dans l'équipe CRCINA, U1232, INSERM. Ils ont bénéficié d'un financement par le projet européen H2020 COMPARE et par l'IFREMER (Direction Scientifique). Par ailleurs, grâce à un appel à mobilité internationale, financé par la Direction Scientifique de l'IFREMER, j'ai eu l'opportunité de réaliser une partie des ces travaux dans le laboratoire ViroScience (Erasmus Medical Center, Rotterdam), dirigé par M. Koopmans pour une durée de trois mois.

Ces travaux ont fait l'objet de publications, communication orales et affichées présentées ci-dessous :

PUBLICATIONS

- **Strubbia, S.**, Schaeffer, J., Besnard, A., Wacrenier, C., Le Mennec, C., Desdouits, M., Le Guyader, F.S., (2019). Metagenomic to detect norovirus and other Human enteric viruses in oysters: impact on hexamer selection and targeted capture-based enrichment. (Manuscript soumis à International Journal of Food Microbiology).
- **Strubbia, S.**, Schaeffer, J., Oude Munnink, B., Besnard A., Phan, M. V. T., Nieuwenhuijse, D., de Graaf, M., Schapendonk, C., Wacrenier, C., Cotten, M., Koopmans, M., Le Guyader, F.S. (2019). Metavirome sequencing to evaluate norovirus diversity in sewage and related bioaccumulated oysters. *Frontiers in Microbiology*, 10. <https://doi.org/10.3389/fmicb.2019.02394>
- **Strubbia, S.**, Phan, M. V. T., Schaeffer, J., Koopmans, M., Cotten, M., Le Guyader, F. S. (2019). Characterization of Norovirus and Other Human Enteric Viruses in Sewage and Stool Samples Through Next-Generation Sequencing. *Food and Environmental Virology*. <https://doi.org/10.1007/s12560-019-09402-3>
- Rincé, A., Balière, C., Hervio-Heath, D., Cozien, J., Lozach, S., Parnaudeau, S., Le Guyader F., Le Hello S., Giard J.C., Sauvageot N., Benachour A., **Strubbia S.**, Gourmelon, M. (2018). Occurrence of bacterial pathogens and human noroviruses in shellfish- Harvesting areas and their catchments in France. *Frontiers in Microbiology*, 9. <https://doi.org/10.3389/fmicb.2018.02443>

COMMUNICATIONS DANS DES CONGRES

- **Strubbia S.**, Schaeffer J., Besnard A., Phan M.V.T., Cotten M., Oude Munnik B.B., Nieuwenhuijse D.F., Koopmans M., Le Guyader S.F., 2019 « Metagenomic approach to study norovirus diversity in sewage and oyster samples ». 7^{ème} congrès international Calicivirus, 17 octobre, Sydney (Australie). Sélectionnée pour une bourse de voyage étudiant. *Communication orale*
- **Strubbia S.**, Schaeffer J., Besnard A., Desdouits M., Le Guyader S., 2019 « Capture-based deep sequencing in sewage and in three oyster types using In-house designed hexamer set ». 7^{ème} congrès international Calicivirus, 17 octobre, Sydney (Australie). *Communication affichée*
- Le Guyader S.F., Desdouits M., **Strubbia S.**, Le Mennec C., Polo D., Zeng S., Ettayebi K., Estes M.K., 2019 « Using enteroids to follow infectious human norovirus in seawater ». 7^{ème} congrès international Calicivirus, 17 octobre, Sydney (Australie). *Communication affichée*
- Schaeffer J., Desdouits M., Besnard A., **Strubbia S.**, Le Guyader S.F., 2019 « Diversité des virus entériques dans des eaux usées collectées durant la dernière décennie ». 15^{ème} congrès national de la Société Française de Microbiologie, 30 septembre – 2 octobre 2019, Paris (France). *Communication affichée*
- **Strubbia S.**, Schaeffer J., Phan M.V.T., Cotten M., Oude Munnik B.B., Nieuwenhuijse D.F., Koopmans M., Le Guyader S. 2019 « Metagenomic approach to study norovirus diversity in sewage and oyster samples ». European Congress of Virology, 28 avril – 1 mai, Rotterdam (Pays Bas). *Communication affichée*
- Schaeffer J., Besnard A., **Strubbia S.**, Le Guyader S.F., 2019 « Enteric virus diversity in sewage collected during last decade » European Congress of Virology, 28 avril – 1 mai, Rotterdam (Pays Bas). *Communication orale*
- **Strubbia S.**, Schaeffer J., Besnard A., Le Guyader S.L., 2018 « Etude de la diversité des norovirus dans les eaux usées par la métagénomique ». Journées francophones de virologie, 27 – 29 mars. Lyon (France). *Communication affichée*

COMMUNICATIONS DANS DES GROUPES DE TRAVAIL

- **Strubbia S.**, Schaeffer J., Phan M.V.T., Cotten M., Oude Munnik B.B., Nieuwenhuijse D.F., Koopmans M., Le Guyader S., 2018 «Double approach to estimate norovirus diversity in sewage using Next generation sequencing». Young COMPARE meeting, 27 février, Copenhague (Danemark). *Communication orale*
- **Strubbia S.**, Schaeffer J., Besnard A., Le Guyader S., 2018 « Norovirus et huîtres, analyse metagenomique d'échantillons complexes ». Journées bio-informatiques, 12-13 novembre, Nantes (France). *Communication orale*
- **Strubbia S.**, Schaeffer J., Le Guyader S., 2018 « Sample preparation : how to eliminate bacteria ? » 2018 COMPARE general meeting 28 février, Copenhague (Danemark). *Communication orale*
- Wylezich C., Zitzow P., Kauer L., Zoli M., Belka A., Pohlmann A., **Strubbia S.**, Le Guyader S., Bee M., Höper D., 2017. « Metagenomics for food safety » COMPARE general meeting, 28 février-3 mars, Rotterdam. *Communication affichée*
- **Strubbia S.**, Schaeffer J., Desdouits M., Phan M.V.T., Cotten M., Koopmans M., Le Guyader S., 2018. « Metagenomics and metabarcoding approaches to evaluate norovirus diversity in wastewater samples ». COMPARE general meeting, 28 février-2 mars 2018, Copenhague (Danemark). *Communication affichée*

ACTIVITÉS COMPLEMENTAIRES

- Visite au laboratoire Friedrich-Loeffler-Institut (FLI) à Greifswald (Allemagne) du 4 au 10 décembre 2016 sous la direction de Dr. Dirk Hoepfer afin de tester des protocoles d'élution et d'extraction en vue du séquençage metagénomique
- Experience d'enseignement à l'UFR Sciences et Techniques de Nantes : 64 heures en 2016/2017 aux étudiants en première et deuxième année de Licence Sciences de la Vie en travaux pratiques (microbiologie et biologie cellulaire) et en travaux dirigés (génétique)
- Mobilité internationale dans le laboratoire ViroScience (Erasmus Medical Center), à Rotterdam (Pays Bas), du juin à août 2018 sous la direction de M. Koopmans afin de développer les approches utilisées pour le séquençage haut-débit des échantillons faiblement contaminés, analyser et représenter graphiquement les données bio-informatiques.

- Selectionnée pour participer à une formation organisée par l'Université de Stirling (Ecosse) sous le projet AQUAEXCEL2020 du 26 au 30 aout 2019. Training course : Introductory Bioinformatic Course to Sequencing Data Processing.

TABLE DES MATIERES

REMERCIEMENTS	2
AVANT-PROPOS	5
TABLE DES MATIERES	9
TABLE DES FIGURES	12
TABLE DES TABLEAUX	15
ABBREVIATIONS	16
INTRODUCTION	17
1. NOROVIRUS	19
1.1 HISTORIQUE	19
1.2 LES CALICIVIRUS	19
1.3 STRUCTURE DU GENOME VIRAL	21
1.4 STRUCTURE DE LA CAPSIDE VIRALE	22
1.5 TAXONOMIE ET CLASSIFICATION	23
1.6 MUTATION ET EVOLUTION	24
1.7 RECOMBINAISONS GÉNÉTIQUES	26
2. NOROVIRUS CHEZ L'HOMME	27
2.1 SENSIBILITE GENETIQUE AUX NOVs	27
2.2 FACTEURS INFLUENCANT L'EPIDEMIE	28
2.3 PATHOGENÈSE	28
2.4 VOIE DE TRANSMISSION	29
2.5 CYCLE DE REPLICATION	30
2.6 IMMUNOLOGIE	31
2.7 EPIDEMIOLOGIE	32
3. LES VIRUS ENTERIQUES HUMAINS ET L'ENVIRONNEMENT	35
3.1 VIRUS ENTERIQUES	35
3.2 NOVs DANS L'ENVIRONNEMENT	36
3.3 PRINCIPES ET FONCTIONNEMENT DES SYSTEMES D'EPURATION	37
3.4 EFFICACITE DES TRAITEMENTS D'EPURATION	38
4. NOROVIRUS ET HUITRE	40
4.1 LES HOMMES ET LES HUITRES, UNE HISTOIRE ANCIENNE	40
4.2 ANATOMIE DE L'HUÎTRE	41
4.3 LES HUÎTRES ET LES EPIDEMIES DE GASTROENTERITE	42
4.4 DES LIGANDS SPECIFIQUES POUR LES NOVs DANS LES HUITRES	43
4.5 EPIDEMIES DE NOVs ET HUITRES	44
5. METHODES DE DETECTION	48
5.1 ELUTION VIRALE	48
5.2 ELUTION CIBLEE DES NOVs	52
5.3 EXTRACTION DES ACIDES NUCLÉIQUES ET QUANTIFICATION	53
6. SEQUENCAGE HAUT DEBIT	55

6.1 LE SEQUENCAGE, DE GENERATION EN GENERATION	55
6.2 METAGENOMIQUE ET METABARCODING	57
6.3 HTS DANS LE CONTEXTE DE LA METAGENOMIQUE VIRALE	58
6.4 PREPARATION DES ECHANTILLONS	58
6.5 PREPARATION DES LIBRAIRIES	60
6.6 CAPTURE ET ENRECHISSIMENT	64
6.7 PIPELINE BIOINFORMATIQUE POUR L'ANALYSE DES DONNEES DE SEQUENCAGE	64
6.8 BIAS ET ERREURS DE SEQUENCAGE	70
6.9 CONTROLES ET RISQUES DE CONTAMINATIONS	71
6.10 NOVs : UN CHALLENGE METAGENOMIQUE	72
7. APPROCHES COMPLEMENTAIRES	74
7.1 PHAGE DISPLAY	74
7.2 PMA	75
7.3 LES APTAMERES	76
7.4 TECHNOLOGIE MICROFLUIDIQUE	77
7.5 LES MUCINES GASTRIQUES PORCINES	78
7.6 VIRUS TULANE COMME MODELE DE NOVs	79
8. REPLICATION DES NOVs SUR CULTURE CELLULAIRE	81
8.1 PHYSIOLOGIE ET ANATOMIE DE L'INTESTIN GRÊLE CHEZ L'HOMME	81
8.2 MISE EN CULTURE DE CELLULES SOUCHES ET LEUR DIFFERENTIATION	82
8.3 PREPARATION DE L'INOCULUM ET INFECTION	85
8.4 REPLICATION DES SOUCHES DE NOVs	86
8.5 APPLICATION ET DEVELOPPEMENTS FUTURS	87
OBJECTIFS DES TRAVAUX DE THESE	88
TRAVAUX ORIGINAUX	89
CHAPITRE 1 : ETUDE DE LA SPECIFICITÉ DES PGM	89
INTRODUCTION	89
MATERIEL ET METHODES	90
RESULTATS PRELIMINAIRES	92
DISCUSSION	95
CHAPITRE 2 : ETUDE DE LA PERSISTANCE DES NOVs DANS L'EAU DE MER	97
INTRODUCTION	97
MATERIEL ET METHODES	97
RESULTATS PRELIMINAIRES	98
DISCUSSION	102
ARTICLE I	103
1.1 INTRODUCTION A L'ETUDE	103
1.2 OBJECTIF DE L'ETUDE	104
1.3 ARTICLE I	105
1.4 DISCUSSION	115
1.5 CONCLUSIONS ET PERSPECTIVES	116
ARTICLE II	117
2.1 INTRODUCTION A L'ETUDE	117
2.2 OBJECTIF DE L'ETUDE	118
2.3 ARTICLE II	119

2.4 DISCUSSION	132
2.5 CONCLUSIONS ET PERSPECTIVES	133
ARTICLE III	134
3.1 INTRODUCTION A L'ETUDE	134
3.2 OBJECTIF DE L'ETUDE	135
3.3 ARTICLE III	136
3.4 DISCUSSION	143
3.5 CONCLUSIONS ET PERSPECTIVES	164
CONCLUSIONS GENERALES ET PERSPECTIVES	165
REFERENCES	170
LIENS INTERNET	188
ANNEXE I	189
ANNEXE II	191
ANNEXE III	192
ANNEXE IV	193
ANNEXE V	194
ANNEXE VI	195
COMMUNICATIONS AFFICHÉES	197

TABLE DES FIGURES

Figure 1 Agrégation des particules du diamètre de 27nm observées par immunomicroscopie électronique.....	19
Figure 2 Représentation du génome de NoVs.....	21
Figure 3 Structure de la capside virale.....	22
Figure 4 Classification des NoVs en sept génogroupes selon la séquence en acides aminés de la protéine structurale VP1.	24
Figure 5 Evolution temporelle de apparition d'apparition des variantes GII.4 et d'autres génotypes récents.....	25
Figure 6 Modèle d'évolution de NoVs.....	26
Figure 7 Variabilité du profil de glycosilation dans la population et son impact sur la sensibilité aux NoV humains.....	27
Figure 8 Alerte RASFF.....	30
Figure 9 Modèle d'entrée du NoVs.....	31
Figure 10 Variabilité temporelle de NoV GI (bleu), GII (rouge) et GIV (vert) à partir de janvier 2005 jusqu'à janvier 2019.....	33
Figure 11 Variabilité de la région ORF1 des souches de NoVs impliqués dans des épidémies entre janvier 2005 et janvier 2019.....	33
Figure 12 Variabilité de la region ORF2 des souches de NoVs impliqués dans des épidémies entre janvier 2005 et janvier 2019.....	34
Figure 13 Distribution saisonnière et concentration de NoVs GI (noir) et GII (rouge) dans les échantillons d'eaux usées.....	36
Figure 14 Schéma des différents procédés d'épuration.....	38
Figure 15 Anatomie de l'huître creuse et de l'huître plate.....	41
Figure 16 Représentation du modèle probabiliste de type SEIR	43
Figure 17 Localisation des NoVs GI.1 au niveau du tissu digestif de l'huître.....	44
Figure 18 Détection des NoVs dans les zones de production française.....	47
Figure 19 Représentation schématisée des étapes de concentration primaire.....	48
Figure 20 Résumé des étapes clés pour l'éluion des particules virales de NoVs à partir d'échantillons d'eaux et d'huîtres.....	52
Figure 21 Schéma de fonctionnement de la digital PCR.....	54
Figure 22 Dispositif MinION d'Oxford Nanopore.....	57
Figure 23 Schématisation du protocole proposé par Conceição-Neto et collègues.....	59

Figure 24 Schéma de l'étape de tailing et ligation.....	63
Figure 25 Amplification à pont des bibliothèques avant séquençage per Illumina.....	63
Figure 26 A) Schéma des étapes principales pour l'analyse des données de séquençage. B) Panoramique des différents pipelines utilisés et identification des étapes incluses.....	65
Figure 27 Schématisation de l'assemblage des reads.	67
Figure 28 Graphe de De Bruijn pour l'assemblage.....	68
Figure 29 Impact de la présence des séquences trop riches ou trop pauvres en nucléotides G et C sur le séquençage haut débit.....	70
Figure 30 Détection des VLP de NoVs GI.1 à l'aide de nanoparticules de phage marquées dans un essai à flux latéral.....	75
Figure 31 Approche de qRT-PCR en utilisant le PMA pour différencier les NoVs infectieux et non infectieux selon la perméabilité de la capsid virale.....	76
Figure 32 Diagrammes de flux montrant une approche de détection de norovirus basée sur l'utilisation d'aptamères.....	77
Figure 33 Diagrammes de flux montrant une approche de détection de norovirus en utilisant la technologie microfluidique.....	78
Figure 34 Smartphone et dispositif microfluidique sur papier pour la détection des NoVs....	78
Figure 35 Persistance du virus Tulane (TV) dans les huîtres.....	80
Figure 36 Organisation de l'épithélium intestinal.....	82
Figure 37 Modèle de culture des entéroïdes.....	83
Figure 38 Schématisation des phases de développement des entéroïdes.....	84
Figure 39 Diagramme schématisé du passage des entéroïdes de 3D à 2D pour la formation de monocouches.	85
Figure 40 Structure et composition des mucines porcines et bovines.....	90
Figure 41 Choix des conditions à adopter pour favoriser le « coating » des mucines.....	92
Figure 42 Test ELISA qui montre les ligands glycosylés exprimés par les mucines de porc (bleu) et de bœuf (vert) ainsi que l'efficacité de la désialylation (mucines de porc désialylés en orange, de bœuf en gris).....	93
Figure 43 Courbe effet dose-dépendante de la fixation des VLP GII.4 sur les PGM.....	94
Figure 44 Courbe effet dose-dépendante des liaisons entre PGM et VLP GI.1.....	94
Figure 45 Diagramme en violon de la récupération d'ARN de NoVs GII.3, GII.4 et TV suite au protocole de purification des échantillons d'eau de mer.....	99
Figure 46 Concentration de l'ARN viral de TV et des souches GII.3 et GII.4 de NoVs	

quantifiées par rRT-PCR.....	100
Figure 47 Ratio de répliation des souches NoVs GII.3 et GII.4 au cours du temps.....	101
Figure 48 Décroissance du titre infectieux de TV indiquée par TCDI50/mL.....	101

TABLE DES TABLEAUX

Tableau 1 Classification des calicivirus incluant les genres et espèces types.....	20
Tableau 2 Nomenclature des protéines structurales et non structurales qui composent le génome des NoVs, et leurs fonctions.....	22
Tableau 3 Principaux virus entériques humains rejetés dans l'environnement.....	36
Tableau 4 Exemple d'études sur la prévalence des NoVs dans les coquillages.....	45
Tableau 5 Comparaison des techniques de séquençage HTS de deuxième génération.....	56
Tableau 6 Description des étapes prévues pour différents kits utilisés pour la préparation des bibliothèques Illumina.....	60
Tableau 7 Estimation de la qualité des séquences obtenues.....	66
Tableau 8 Exemple d'une manip faite en utilisant une souche GI.1.....	95

ABBREVIATIONS

ADN	Acide DésoxyriboNucléique
ADNds	ADN double brin
AN	Acides Nucléiques
ARN	Acide RiboNucléique
ARNsb	ARN simple brin
BSM	Bovine Submaxillary Mucins
Ct	Cycle threshold
DO	Densité Optique
ELISA	Enzyme Linked Immunosorbent Assay
EFSA	European Food Safety Authority
HBGA	Histo-Blood Group Antigens
HTS	High Throughput Sequencing
ISO	Organisation Internationale de Normalisation
MOI	Multiplicity Of Infection
NCBI	National Center for Biotechnology Information
NoVs	Norovirus
Nt	Nucléotide
ORF	Open Reading Frame
Pb	Pair de base
PEG	Polyéthylène Glycol
PGM	Porcin Gastric Mucins
R	Ratio
RT-PCR	ReverseTranscription et Polymerase Chain Reaction
TCID₅₀	Median Tissue Culture Infective Dose
TD	Tissu Digestif
VLP	Virus Like Particles

INTRODUCTION

La biodiversité virale est extrêmement abondante. On estime que le corps humain contient 100 fois plus de virus (10^{15}) que de cellules (10^{13}). Plus largement, 10^{31} virus sont estimés être contenus dans la biosphère (Wobus *et al.*, 2012 ; Zarate *et al.*, 2017). Dans les océans, plus de 200.000 populations virales ont été identifiées lors de l'expédition scientifique *Tara Oceans* qui entre 2009 et 2013 a prélevé 35.000 échantillons de plancton marin autour de la planète (Gregory *et al.*, 2019).

Outre cette immense abondance, les virus sont également caractérisés par une variabilité très importante, en termes de structure, de taille, de génome et de résistance aux conditions physico-chimiques. Moins de 1% de la diversité virale existante a été explorée.

Parmi cette immense population, les virus entériques humains constituent un groupe d'intérêt en santé publique. Ces virus, parasites intra-cellulaires obligatoires, possèdent des capacités de résistance très importantes et peuvent re-infecter l'Homme après des semaines de persistance dans l'environnement (Metcalf *et al.*, 1995). Au sein de ce groupe, notre attention s'est portée sur un petit virus à ARN, présent dans le monde entier, le norovirus (NoV), agent majeur de gastroentérite virale aiguë chez l'homme.

Les NoVs sont parfois détectés dans l'environnement littoral, où ils parviennent suite à des rejets humains (rejet direct, pluie torrentielle...), avec pour conséquence majeure leur accumulation par les huîtres, élevées en zone littorale. Leur consommation peut ensuite être source de nouvelles épidémies, montrant la capacité de résistance de ces virus qui restent infectieux après ce cycle de transfert par l'environnement.

La contamination du littoral représente donc une problématique de santé publique et la caractérisation des virus entériques humains présents dans cet environnement peut aider à la prévention de la transmission de ces virus et la protection du consommateur.

C'est dans cette optique de '*one world one health*' que s'insère le projet Européen H2020 COMPARE qui a pour ambition d'améliorer et de faciliter la détection et la prise de décision en lien avec les épidémies chez l'Homme ou l'animal en utilisant les nouveaux outils du séquençage génomique. Cette approche novatrice permet de détecter et d'identifier tous les génomes contenus dans un échantillon.

Ma thèse, s'incluant dans ce projet, a comme objectif le développement méthodologique pour l'application de la métagénomique à l'analyse d'échantillons de l'environnement tels que les eaux usées et les huîtres. L'approche métagénomique est séduisante pour l'étude d'échantillons environnementaux car elle offre la possibilité d'identifier d'autres virus entériques humains éventuellement présents. Cette technique ouvre de nouvelles perspectives comme l'utilisation des eaux usées comme outil de surveillance (Nieuwenhuijse and Koopmans, 2017). En effet, le séquençage agnostique d'un échantillon d'eau usée peut révéler la diversité des virus en circulation dans la population locale, et être plus représentatif que des échantillons d'individus. Les difficultés majeures rencontrées lors des études de métagénomique virale sont souvent liées à l'échantillon qui contient une grande diversité d'autres virus, de phages, de bactéries et de parasites, constituant un bruit de fond important. Lors du choix de la méthode de séquençage à utiliser et de la préparation de l'échantillon, il faut prendre en considération la matrice et le niveau de concentration virale.

En effet, cette abondance et la diversité du matériel génomique compliquent la détection des NoV, souvent présents en faibles concentrations, d'autant plus que ces virus possèdent un génome relativement petit en comparaison avec d'autres micro-organismes.

Parallèlement aux progrès technologiques du séquençage à haut débit, une autre avancée majeure a écrit un nouveau chapitre dans l'histoire des NoV: le développement d'un système de culture cellulaire permettant la réplication virale. Découverts dans les années 1970, leur étude a été limitée par les difficultés de cultiver ce virus *in vitro*.

Pendant longtemps, la culture de cellules épithéliales intestinales humaines n'a pas été concluante pour leur multiplication, justifiant l'utilisation de virus modèles ou des études sur volontaires humains (Estes *et al.*, 2019). Les travaux réalisés sur la culture d'épithélium intestinal humain à partir de cellules souches (Sato and Clever, 2013) a permis la réplication de plusieurs souches de NoVs (Ettayebi *et al.*, 2016), technique reproduite depuis dans plusieurs laboratoires. Pouvoir étudier la réplication virale permet de mieux comprendre la biologie du virus mais également la persistance des particules dans l'environnement, point majeur pour le développement des stratégies pour la prévention ou le contrôle des infections par les NoVs.

Dans ce manuscrit la première partie comprend une revue bibliographique exposant les caractéristiques des NoVs, et l'impact du rejet de ces virus sur la contamination de l'environnement littoral et des huîtres. Les différentes approches de métagénomique ainsi que les méthodes de culture des entéroïdes, sont également exposées. La seconde partie comprend des travaux originaux avec les premiers deux chapitres présentant des résultats préliminaires, données acquises mais incomplètes qui devront être finalisées, et les articles publiés ou soumis.

Ce mémoire se termine par une discussion sur l'intérêt de ces travaux et sur les perspectives envisageables pour compléter cette étude.

1. NOROVIRUS

1.1 HISTORIQUE

Pendant longtemps les épidémies de gastroentérite aiguë non-bactériennes étaient simplement connues comme « grippe hivernale » ou « winter vomiting disease », sans identification d'un agent pathogène, mais avec l'observation d'une saisonnalité récurrente (pendant la saison froide). L'histoire des NoVs commence en 1972, quand un virologue, Albert Kapikian, mit en évidence pour la première fois en microscopie électronique la présence de petites particules de symétrie icosaédrique, dans une selle d'un enfant impliqué dans une épidémie survenue dans une école de la ville de Norwalk, Ohio, Etats-Unis (Figure 1), d'où le premier nom attribué à la première souche identifiée, le virus de Norwalk.

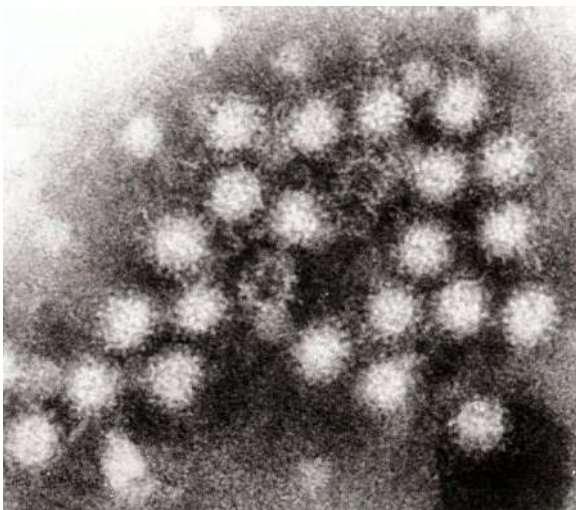


Figure 1 Agrégation des particules du diamètre de 27nm observées par immunomicroscopie électronique (Kapikian *et al.*, 1972)

C'est dans les années 90, grâce à l'évolution de la biologie moléculaire que les études sur les NoV progresseront. Le premier événement marquant est le séquençage du génome complet de la souche Norwalk dans l'équipe de M.K. Estes (Jiang *et al.*, 1992).

A l'heure actuelle les NoV humains sont reconnus comme la première cause de gastroentérite aiguë d'origine virale toute classe d'âge confondue dans le monde.

1.2 LES CALICIVIRUS

La famille des *Caliciviridae* est constituée de virus à ARN simple brin, dépourvus d'enveloppe (également appelés virus nus) leur conférant une résistance importante dans l'environnement. L'enveloppe virale, étant lipoprotéique, est en effet thermolabile, sensible aux enzymes et à de nombreux agents chimiques et constitue un facteur de fragilité chez un virus. Selon la dernière parution du Comité International de Taxonomie des Virus (ICTV, International Committee on Taxonomy of Viruses) les calicivirus sont organisés en cinq genres : trois sont responsables de maladie affectant les animaux : les *Lagovirus*, les *Nebovirus* et les *Vesivirus* et deux affectent l'Homme et les animaux : les *Norovirus* et les

Sapovirus (Desselberger, 2019). D'autres virus ayant des caractéristiques morphologiques et physicochimiques similaires à celles des calicivirus pouvant être regroupés dans le genre *Recovirus* sont en attente de classement (Tableau 1).

Tableau 1 Classification des calicivirus incluant les genres et espèces types (et abréviations utilisées) (Desselberger, 2019)

Genus	Type Species
Established:	
<i>Norovirus</i> (NoV)	<i>Norwalk virus</i> (NV)
<i>Sapovirus</i> (SaV)	<i>Sapporo virus</i> (SV)
<i>Lagovirus</i> (LaV)	<i>Rabbit hemorrhagic disease virus</i> (RHDV) <i>European brown hare syndrome virus</i> (EBHSV)
<i>Vesivirus</i> (VeV)	<i>Vesicular exanthema of swine virus</i> (VESV) <i>Feline calicivirus</i> (FCV) <i>San Miguel sea lion virus</i> (SMSV)
<i>Nebovirus</i> (NeV)	<i>Newbury-1 virus</i> (NBV)
Proposed:	
<i>Recovirus</i> (ReV)	<i>Tulane virus</i> (simian)
<i>Valovirus</i> (VaV)	<i>St Valérian virus</i> (porcine)
<i>Bavovirus</i> (BaV)	<i>Bayern virus</i> (avian)
<i>Nacovirus</i> (NaV)	<i>Novel avian calicivirus</i> (chicken, turkey)
<i>Salovirus</i> (SaV)	<i>Atlantic salmon calicivirus</i> (salmon)
<i>Minovirus</i> (MiV)	<i>Fathead minnow calicivirus</i> (minnow)

Parmi les calicivirus infectant l'Homme, les sapovirus provoquent également des gastro-entérites aiguës, et ont souvent été identifiés au Japon (Oka *et al.*, 2015). L'augmentation du nombre d'épidémies impliquant ces virus suggère qu'ils doivent être considérés dans les problématiques de santé publique (Torner *et al.*, 2016). Ces virus sont également fréquemment retrouvés dans l'environnement et les coquillages, confirmant leur présence chez l'Homme (Fusco *et al.*, 2017).

L'absence d'un modèle animal ainsi que les difficultés rencontrées dans la mise au point d'une méthode de culture des NoVs *in vitro*, ont justifiés la recherche de modèles d'étude ou de virus de substitution en utilisant des calicivirus comme prototype. Parmi les modèles les plus utilisés pour l'analyse de la biologie des NoV, nous pouvons citer les calicivirus félins (FCV), premiers calicivirus cultivables en laboratoire, le NoVs murin (MNV), découvert en 2003 et, plus récemment, le virus tulane (TV) (Farkas *et al.*, 2008; Thorne and Goodfellow, 2014).

Ce virus a été isolé pour la première fois en 2008 à partir de selle de singe rhésus. L'analyse de son génome et de ses caractéristiques physicochimiques ont conduit à la proposition d'un nouveau genre, les *Recovirus*, au sein de la famille *Caliciviridae* (Farkas *et al.*, 2008). C'est notamment sa similitude aux NoVs humains lors des premières étapes de l'infection (reconnaissance de mêmes ligands), et la possibilité de le cultiver *in vitro* sur cellules de rein de singes qui renforcent l'intérêt du TV comme virus modèle pour étudier la physiologie des

NoVs (Farkas, 2015). Pour notre thématique, ce virus est intéressant car sa distribution tissulaire suite à des contaminations d'huître en bassin est identique à celle des NoVs (Drouaz *et al.*, 2015). Sa longue persistance dans les tissus, comparable à celle observée pour les NoVs, nous a confirmé sa pertinence comme virus modèle (Polo *et al.*, 2018).

1.3 STRUCTURE DU GENOME VIRAL

Le génome des NoVs est constitué d'un ARN linéaire monocaténaire d'environ 7600 bases (Figure 2). Les extrémités du génome viral contiennent des structures conservées pour permettre l'interaction avec les cellules hôtes et faciliter la réplication et la production de nouveaux virus (Thorne and Goodfellow, 2014). La portion 5' du génome présente la protéine VPg (Viral Protein linked to Genome), liée par covalence et essentielle à la réplication virale. La portion 3' se termine avec une séquence poly A impliquée dans la stabilité du génome viral et la traduction (Robilotti, Deresinski and Pinsky, 2015).

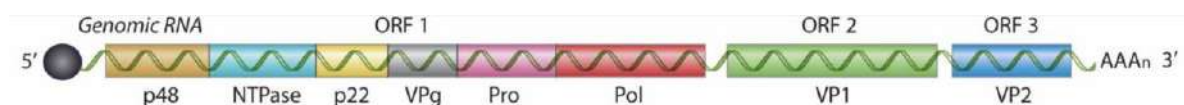


Figure 2 Représentation du génome de NoV (Robilotti, 2015)

Le génome des NoV est constitué de trois cadres de lecture (Open Reading Frame, ORF). L'ORF1 code les protéines non structurales, l'ORF2 code la protéine structurale majeure VP1 et l'ORF3 code la protéine structurale mineure VP2 contribuant à la stabilité de la capsid. Les noms et les fonctions des protéines virales sont résumés dans le tableau suivant (Tableau 2), considérant que certaines fonctions ne sont pas encore totalement élucidées.

Tableau 2 Nomenclature des protéines structurales et non structurales qui composent le génome des NoVs, et leurs fonctions (Hardy, 2005; Thorne and Goodfellow, 2014)

Nomenclature des protéines virales	Fonction
P48 (N-term)	Protéine terminale, forme un complexe pendant la réplication virale. Séquence et longueur variables selon les génogroupes.
NTPase (Nucléoside triphosphate)	RNA hélicase / NTPase.
p22	Formation d'un complexe pendant la réplication.
VPg	Protéine liée à l'ARNm de façon covalente, impliquée dans la traduction et la réplication du génome viral.
Pro (3C-like)	Unique protéase codée par le génome de NoV.
RdRp	RNA-dépendante RNA polymérase.
VP1 (Viral Protein 1)	Protéine majeure de la capside virale, participant à l'assemblage de la capside.
VP2 (Viral Protein 2)	Protéine mineure de la capside, rôle probable dans la production de virus infectieux et dans l'encapsidation du génome.

1.4 STRUCTURE DE LA CAPSIDE VIRALE

Les virions sont constitués d'une seule protéine de capside VP1 organisée en 90 dimères assemblés avec une symétrie icosaédrique pour former une capside de 28-35 nm de diamètre. VP1 est organisée en deux domaines principaux : le domaine S (shell en anglais), qui forme une sorte de coque intérieure et le domaine P (Protruding) formés de deux sous-domaines, P1 et P2 (**Erreur ! Source du renvoi introuvable.**), proéminentes. Les interactions cellulaires sont majoritairement localisées sur le sous-domaine P2 à la surface du domaine S, ce qui se traduit par une variabilité importante du génome (Almand *et al.*, 2017). L'interaction entre les domaines S et P joue un rôle dans l'assemblage de la capside mais aussi dans sa stabilité et le contrôle de sa taille (Choi *et al.*, 2008).

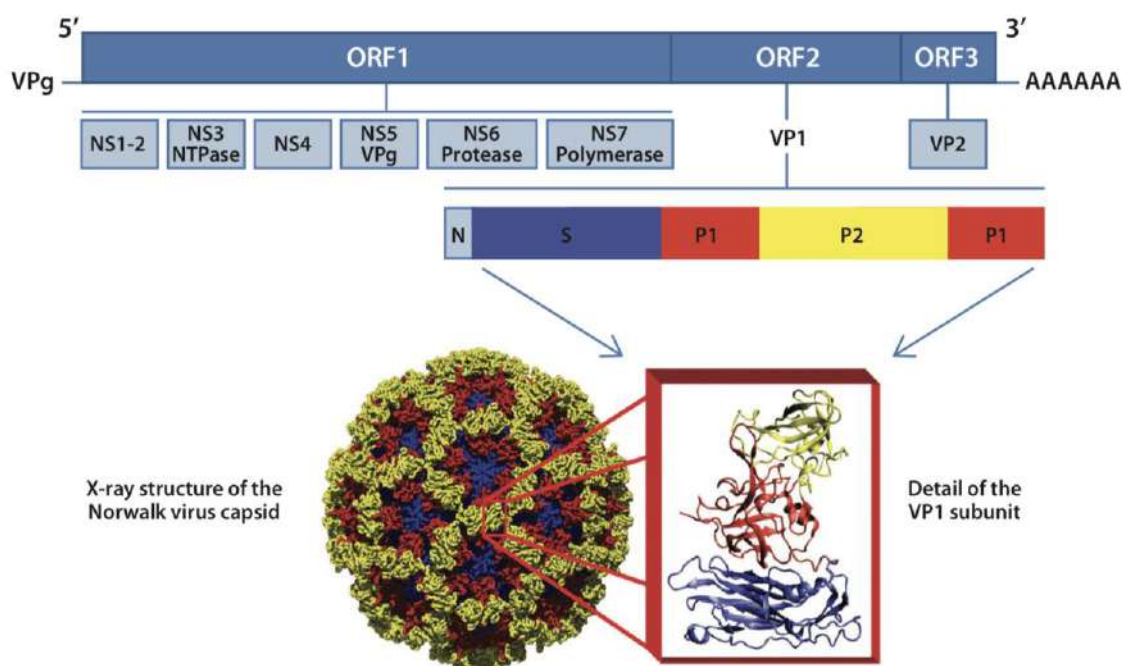


Figure 3 Structure de la capside virale (Atmar *et al.*, 2005)

La capacité d'auto-assemblage des VP1 a été démontrée *in vitro* (Jiang *et al.*, 1992). Cette protéine structurale forme de façon spontanée des Virus Like Particles (VLPs) (ou pseudo-particules en français), particules identiques aux virions en terme de morphologie et d'antigénicité, mais dépourvues d'ARN. Les VLPs ont permis des avancées notables dans la compréhension de l'architecture des NoVs, de leur capacité de liaison aux récepteurs cellulaires et de la réponse du système immunitaire (Prasad *et al.*, 2000; Ruvoën-Clouet *et al.*, 2013; Tan and Jiang, 2014). Ces VLPs ont également trouvé un intérêt pour étudier les interactions entre les NoVs et les tissus de coquillages telles que l'expression des ligands ou leur localisation tissulaire (Le Guyader *et al.*, 2006; Maalouf *et al.*, 2010).

1.5 TAXONOMIE ET CLASSIFICATION

Basée sur une analyse phylogénétique de la séquence de la protéine structurale VP1, la classification des NoVs est actuellement organisée en 7 génogroupes (G), de GI à GVII, eux-mêmes divisés en génotypes. Les NoVs infectent divers hôtes : les GI, GII et GIV infectent l'Homme, le GIII les bovins et les ovins, le GV les murins, les GVI et GVII les chiens (De Graaf, Van Beek and Koopmans, 2016; Atmar *et al.*, 2018). Parmi les souches humaines, plus de 30 génotypes ont été identifiés et de nouveaux variants émergent environ tous les deux ou trois ans, entraînant une augmentation des cas de gastro-entérites dans le monde (Kroneman *et al.*, 2013; Atmar *et al.*, 2018).

Pour être classées dans un même génogroupe, les souches doivent partager au moins 60 % des séquences en acides aminés de la VP1, alors qu'entre les deux génogroupes GI et GII cette similarité est inférieure à 50 %. Au sein d'un même génotype, les NoVs doivent avoir une homologie minimale de 80 % avec la souche de référence (Kroneman *et al.*, 2013; Vinjé, 2014) (Figure 4).

Les NoVs étant sujets à recombinaison (voir paragraphe 1.7), un double système de nomenclature a été proposé, basé sur la séquence de la polymérase (P) dans l'ORF1 et de la capsid (VP1) dans l'ORF2 (Kroneman *et al.*, 2013). Cette nomenclature inclue également des informations concernant l'espèce hôte, les abréviations du nom de la ville et du pays d'identification de la souche, ainsi que l'année de la première identification, le génogroupe et le génotype (ORF1 et ORF2) et, en cas de nouveaux variants, le nom du lieu d'identification et l'année de la 1ère identification. Pour exemple, la nomenclature pour la souche de Norwalk, prototype des NoV humains est : GI/Hu/US/1968/GI.P1-GI.1/Norwalk, indiquant qu'il s'agit d'une souche du génogroupe I, dont l'hôte est l'Humain (Hu), identifiée aux Etats-Unis (US), au cours de l'année 1968, dans la ville de Norwalk. GI.P1-GI.1 se réfère respectivement au classement de la polymérase (ORF1) et la capsid (ORF2).

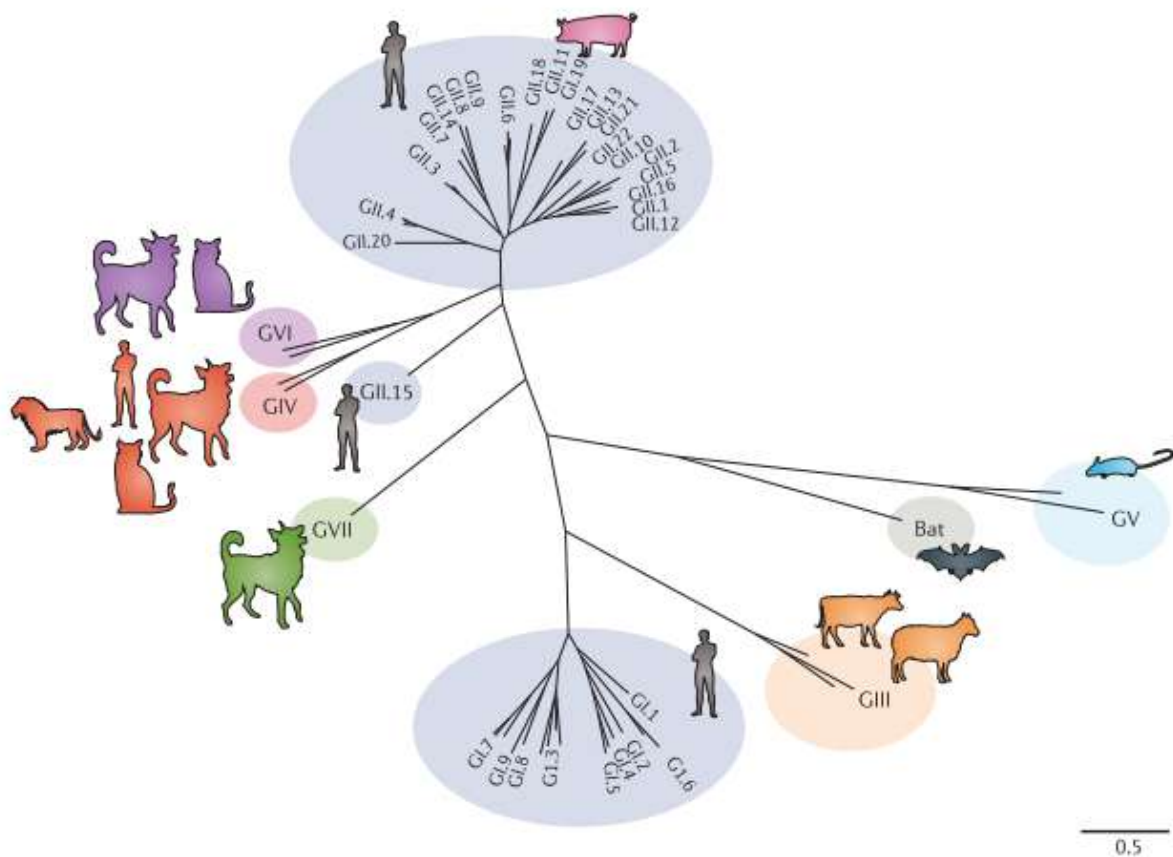


Figure 4 Classification des NoVs en sept génogroupes selon la séquence en acides aminés de la protéine structurale VP1.

L'arbre phylogénétique représente la diversité spatiale et temporelle des souches de NoV. Les NoV appartenant aux GI, GII et GIV infectent l'homme, à exception des virus GII.11, GII.18 et GII.19, qui infectent le porc, et les GIV.2 le chien. Les souches GII.15, détectées uniquement chez l'homme font l'objet d'une proposition d'un nouveau génogroupe. La barre d'échelle reflète le nombre de substitutions d'acide aminé par site (de Graaf et al., 2016).

Depuis la dernière mise à jour de la classification des NoVs en 2013, l'utilisation croissante du séquençage génomique a conduit à l'identification de plusieurs séquences candidates pour être classifiées comme nouveaux génogroupes et génotypes de NoV. Afin de fournir une classification uniforme face aux recombinaisons et pour éliminer le système de dénomination des souches « orphelines » pour l'ORF1 (par exemple, GI.Pa, GI.Pb, GII.Pa, GII.Pe, etc.), une nouvelle classification a été récemment proposée regroupant les séquences nucléotidiques de la région partielle RdRp en (P)-groupes et (P)-types indépendamment de leurs génogroupes et de leur génotypes (basés sur la classification des capsides) (Chhabra et al., 2019). Selon cette nomenclature, la séquence GI.P1-GI.1 est désormais désignée GI.1[P1] et les séquences orphelines telles que GI.Pa, GI.Pb ont été intégrées et normalisées (par exemple GI.P10 et GI.P11, respectivement).

Ce manuscrit de thèse, arrivant dans un moment de transition, a été rédigé selon la classification de Kroneman et al., 2013.

1.6 MUTATION ET EVOLUTION

Comme pour de nombreux autres virus à ARN, l'accumulation des mutations est un des mécanismes à la base de la grande variabilité des NoVs. Parmi les génotypes de NoVs infectant l'Homme, les NoVs GII.4 ont montré la plus grande diversité de variants, répandus

dans le monde entier (Figure 5). Tous les 2 ou 3 ans depuis 2002, un nouveau variant GII.4 devient prédominant remplaçant la souche précédente et causant des nouvelles épidémies (Belliot *et al.*, 2010). Une comparaison entre l'évolution des souches GII.4 et les souches non-GII.4, indique que les souches non-GII.4 sont moins soumises à la pression adaptative et que les variants pandémiques GII.4 ont un taux d'erreur de la polymérase cinq à 40 fois supérieur à celui des souches non pandémiques, engendrant une augmentation dans la vitesse d'évolution de cette souche (Belliot *et al.*, 2010 ; Ruvoën-Clouet *et al.*, 2013; De Graaf *et al.*, 2016).

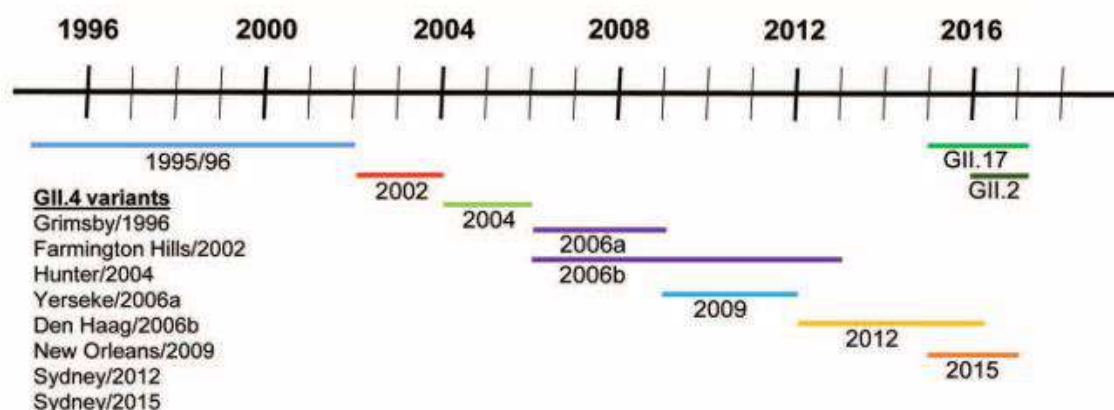


Figure 5 Evolution temporelle de l'apparition d'apparition des variants GII.4 et d'autres génotypes récents (Atmar *et al.*, 2018)

Un certain nombre de mécanismes déterminent l'évolution de cette souche, comme la possibilité de se lier à un plus large panel d'antigènes tissulaires de groupe sanguin (HBGA, Histo Blood Group Antigens), et aussi des taux de réplication et de mutation plus élevés, ce qui leur donne une plus grande capacité d'évolution que les autres génotypes de NoVs (Belliot *et al.*, 2010; White, 2014; Atmar *et al.*, 2018).

Deux modèles d'évolution distincts ont été proposés chez les NoVs : statique et évolutif (Figure 6) (Parra *et al.*, 2017). Ces auteurs proposent le terme d'immunotype, pour différencier entre les virus statiques (principalement souches non-GII.4), ciblant essentiellement les enfants (plus exposés et infectés par des souches différentes en raison d'un système immunitaire encore immature), et les virus évolutifs, virus accumulant continuellement des mutations (souches GII.4 mais aussi GII.17) (Chan *et al.*, 2016). Ces derniers infecteraient préférentiellement les individus plus âgés ayant acquis une immunité contre un certain nombre de génotypes statiques (Mori *et al.*, 2017 ; Parra *et al.*, 2017). Ce modèle est en accord avec les observations de Simmons *et al.* (2013) qui estiment la durée d'une immunité souche-dépendante jusqu'à 9 ans.

Model of Norovirus Reinfection

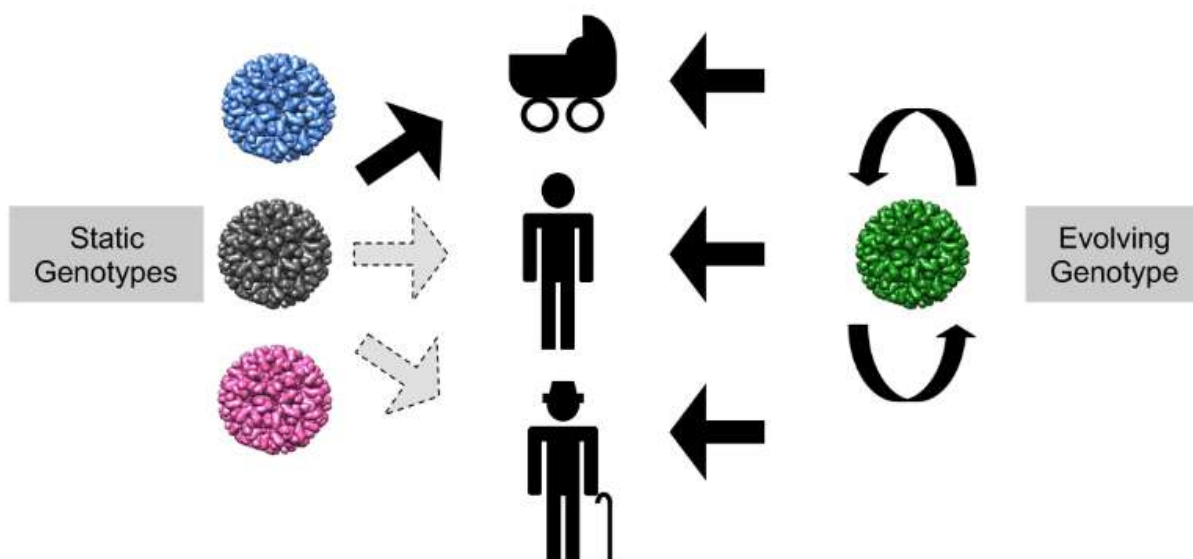


Figure 6 Modèle d'évolution de NoVs.

Les souches GII.4 acquièrent des modifications phénotypiques de leurs principales protéines de capsid au fil du temps (génotype en évolution), tandis que les virus non-GII.4 conservent une protéine de capsid hautement conservée pendant des décennies (génotype statique). Ce modèle prédit que les enfants sont constamment exposés et infectés par des souches de chacun des différents immunotypes (jusqu'à ce qu'une immunité étendue se développe), tandis que les individus plus âgés sont plus susceptibles de tomber malades à cause de l'évolution des génotypes. Ce modèle expliquerait les différences épidémiologiques rapportées dans la distribution des génotypes de norovirus chez les enfants et les adultes (Parra et al., 2017).

1.7 RECOMBINAISONS GÉNÉTIQUES

La recombinaison est la formation d'un génome chimère à partir des génomes parentaux d'origines différentes. Ce mécanisme participe à l'évolution continue des NoVs et à leur diversité avec un impact sur la virulence de la nouvelle souche, la capacité d'évasion au système immunitaire de l'hôte et la résistance dans l'environnement et aux agents antiviraux (Simon-Loriere and Holmes, 2011; White *et al.*, 2014). Les recombinaisons ont essentiellement lieu entre génotypes et sont induites par un saut de l'enzyme polymérase lors de la réplication, généralement au niveau de la jonction entre ORF1 et ORF2. La région du génome impliquée dans la recombinaison est une des portions les plus conservées, ce qui explique cet échange de matrice entre deux souches et la formation des nouveaux variants homologues qui amène à un assortiment entre protéines structurales et non structurales (Simon-Loriere and Holmes, 2011; Vinjé, 2014; Rocha-Pereira *et al.*, 2016). Par ailleurs, comme pour de nombreux autres virus à ARN, la réplication du génome entraîne des erreurs liées à l'activité de la polymérase (De Graaf, Van Beek and Koopmans, 2016).

2. NOROVIRUS CHEZ L'HOMME

2.1 SENSIBILITE GENETIQUE AUX NOV_s

Les souches de NoVs reconnaissent des motifs glycaniques, similaires aux antigènes de groupe sanguin, ce qui induit une différente sensibilité aux infections (Ruvoën and Le Pendu, 2013). Les HBGA sont présents sur les érythrocytes mais également sur la surface des cellules épithéliales de différents tissus dont le système digestif, et sous forme libre ou complexées dans certains liquides biologiques comme la salive et le lait. Les HBGA sont des sucres complexes, appelés glycannes, synthétisés par des glycosyltransférases, enzymes que catalysent le transfert d'un monosaccharide (comme le fucose ou le galactose) pour former des oligosaccharides. Parmi les gènes qui codent pour les glycosyltransférases, trois sont impliqués dans la formation des glycannes qui interagissent avec les NoVs : *Fucosyltransferase 2* (FUT2), *Fucosyltransferase 3* (FUT3) et ABO, chacun présentant un important polymorphisme (Le Pendu *et al.*, 2014). La présence ou l'absence de l'une ou de l'autre de ces enzymes génère une diversité du profil de glycosylation dans la population humaine. Comme schématisé dans la Figure 7, les individus sécréteurs, porteurs du gène codant pour l'enzyme FUT2, représentent environ 80% de la population et sont sensibles aux infections par les NoVs. Le restant de la population, non porteurs de ce gène, sont globalement résistants à l'infection par les NoVs bien que certaines souches les infectants aient été caractérisées. De façon similaire, la présence/absence de l'enzyme FUT3 permet de distinguer les individus Lewis positifs des Lewis négatifs, alors que la variabilité du gène ABO divise la population en individus A, B, AB et O. La combinaison des FUT2 et ABO détermine la sensibilité ou la résistance aux NoVs (Ruvoën and Le Pendu, 2013). Ces antigènes ABH et Lewis étant présents dans les mucines salivaires, un test utilisant des salives et des VLPs permet de discriminer leur sensibilité éventuelle à l'infection en considérant la capacité de fixation des VLPs (Le Pendu *et al.*, 2006).

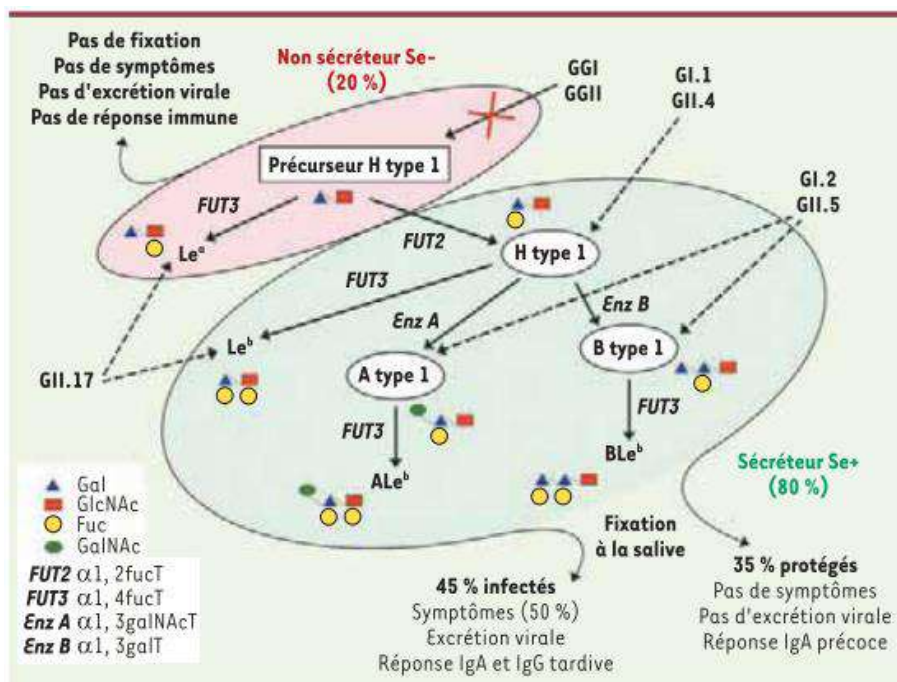


Figure 7 variabilité du profil de glycosylation dans la population humaine et son impact sur la sensibilité aux NoVs (Rougemont *et al.*, 2010).

2.2 FACTEURS INFLUENCANT L'ÉPIDÉMIE

Fortement contagieux, rapidement et efficacement excrétés, très résistants dans l'environnement et en évolution continue, les NoVs ont acquis un caractère épidémique qui fait de ces virus la première cause de gastroentérite aiguë non bactérienne dans le monde, toutes classes d'âges confondues (Hall, 2012).

Les raisons de son succès sont multiples :

- *Excrétion virale*

La quantité de virus excrétée par les individus infectés est importante et varie de 10^5 à 10^{11} particules virales par gramme de selles et cette excrétion peut se prolonger au moins 15 jours après la rémission des symptômes (Aoki *et al.*, 2010; Atmar, 2010; Hall, 2012).

La présence de plusieurs particules de NoVs au sein de vésicules membranaires (diamètre <200 nm) libérées dans les selles des personnes infectées, pourrait contribuer à leur diffusion (Santiana *et al.*, 2018). Ces vésicules pourraient protéger les virus lors de leur passage dans l'appareil gastro-intestinal, et favoriser leur propagation en transportant plusieurs particules virales collectivement vers l'hôte suivant, améliorant la MOI (*Multiplicity Of Infection*) et la gravité de la maladie.

- *Dose infectieuse*

La DI_{50} correspond à la dose de virus qui entraîne des signes cliniques chez 50 % des individus infectés. L'incapacité de cultiver les NoVs humains *in vitro* et l'absence d'un modèle animal ont conduit à l'utilisation des volontaires humains pour étudier la pathogenèse virale et l'immunité. Considérant l'impossibilité de différencier entre particules infectieuses et non, la PCR de l'inoculum a toujours été la méthode sélectionnée pour quantifier les particules virales (Teunis *et al.*, 2008; Atmar *et al.*, 2014).

Une des caractéristiques des NoVs est la dose infectieuse extrêmement basse, moins de 18 particules virales suffisent pour transmettre l'infection (Teunis *et al.*, 2008). Toutefois, la compréhension de la dose infectieuse est en outre compliquée par la sensibilité génétique individu-dépendant et par l'agrégation des particules virales dans l'inoculum (Hall, 2012 ; Atmar *et al.*, 2014). Les NoVs font partie des agents pathogènes les plus infectieux.

- *Individus asymptomatiques*

Selon leur sensibilité génétique, les individus infectés peuvent soit développer la maladie dans les 12-48 heures soit rester asymptomatiques. Ces porteurs asymptomatiques, ignorant leur infection, peuvent favoriser la transmission à d'autres individus en négligeant les règles d'hygiène élémentaires. De plus, l'excrétion des virus par les individus asymptomatiques est presque aussi importante que celle des individus malades, avec des variations importantes (de 10^5 à 10^9 particules virales/g de selle), et une durée estimée entre 8 et 60 jours (Teunis *et al.*, 2015).

2.3 PATHOGENÈSE

Les difficultés rencontrées pour développer un modèle de culture *in vitro* des NoVs, a freiné la compréhension des mécanismes liés à l'infection virale. Pour pallier à ces difficultés, des études physiologiques et histologiques sur des volontaires humains et des modèles animaux

en utilisant des virus modèles ont été réalisées (Cheetham *et al.*, 2006; Troeger *et al.*, 2009; Karst, 2010).

Les étapes clés du processus d'infection des NoVs peuvent se résumer ainsi : suite à l'introduction dans l'organisme, les particules virales infectent les entérocytes du petit intestin (duodénum, jéjunum et iléum). Les cellules du système immunitaire (cellules B, macrophages et cellules dendritiques), comme suggérées dans certaines études, ne semblent pas jouer un rôle majeur, en raison de l'absence de réplication du virus dans ces cellules (Karandikar *et al.*, 2016; Karst and Tibbetts, 2016). Des changements histopathologiques sont observés également dans le petit intestin simultanément à une diminution de l'activité des enzymes de la bordure en brosse entraînant une malabsorption (Karst *et al.*, 2010).

Le microbiote pourrait également jouer un rôle en facilitant le passage des NoV dans le lumen intestinal (Karst and Wobus, 2015), de plus il a été démontré que certaines bactéries entériques reconnaissent spécifiquement les NoVs (Miura *et al.*, 2013). Après pénétration dans la cellule hôte et multiplication, la mort cellulaire conduit à la libération des particules virales néoformées. Des lésions intestinales spécifiques ont été mises en évidence, comme le raccourcissement des microvillosités intestinaux, la vacuolisation des cytoplasmes et l'œdème intercellulaire (Karandikar *et al.*, 2016; Atmar *et al.*, 2018). La mort des entérocytes provoque une inflammation intestinale, et la réponse immunitaire avec la production d'anticorps se met en place. Pendant cette phase de multiplication virale les symptômes apparaissent avec de la fièvre, des nausées souvent accompagnées de vomissements violents et de diarrhées aqueuses (Atmar *et al.*, 2018). La pathogénicité des NoVs semble être souche-dépendante avec certaines souches GII.4 plus souvent associées à des vomissements, diarrhées prolongées et une excrétion dans les selles de particules plus importantes en comparaison aux autres génotypes de NoVs GII. La pathogénicité des NoVs et la quantité de virus excrétés pourraient également être en relation avec l'âge et l'état de santé de l'individu infecté (De Graaf *et al.*, 2016).

2.4 VOIE DE TRANSMISSION

La transmission se fait majoritairement de personne à personne (féco-orale) due à un manque d'hygiène. Cette voie de transmission représente environ 90% des cas et concerne essentiellement les souches GII.4. Ainsi, de nombreuses épidémies impliquant le génotype GII.4 ont été décrites dans les établissements de soins et de santé, dû au contact direct entre personnes (Ahmed *et al.*, 2013; De Graaf *et al.*, 2016; Atmar *et al.*, 2018). Les autres souches sont transmises de façon plus diverses comme montré par les données d'épidémiologie moléculaire. Par exemple, les souches du GI sont fréquemment identifiées suite à la consommation d'eau, d'aliments, lors d'activités de loisirs, ou encore par ingestion des particules projetées par aérosol lors des vomissements (Makison Booth, 2014; Verhoef *et al.*, 2015; De Graaf *et al.*, 2017). Les souches du GI sont également plus fréquemment transmises par les coquillages ou les huîtres contaminées (Le Guyader *et al.*, 2012; Yu *et al.*, 2015). Hormis les souches GII.4, les souches plus communément reconnues comme cause d'épidémie de gastroentérite sont les GII.3 (6.1%), GII.6 (5.7%) et GII.2 (4.6%) (Burke *et al.*, 2018).

La transmission par les aliments est une voie de dissémination importante des NoVs à

travers le monde, dû à la globalisation de la chaîne alimentaire. Les NoVs constituent les agents les plus fréquemment impliqués dans les épidémies d'origine alimentaire aux États-Unis (53.0%) et en Europe (32.2%) (Li *et al.*, 2018). La contamination des aliments peut survenir lors des étapes de la production primaire jusqu'à la commercialisation mais également pendant la manipulation des aliments. Parmi les 14% d'épidémies d'origine alimentaire, 10% sont causées par une souche GII.4, 27% par tous les autres génotypes et le 37% par des mélanges de GII.4 et d'autres NoVs (Verhoef *et al.*, 2015). La compréhension exacte du rôle des NoVs dans les épidémies d'origine alimentaire est difficile à estimer en raison des disparités des méthodes de surveillance selon les pays. La présence d'un système européen d'alerte rapide (RASFF) permet de partager plus efficacement les informations sur les risques pour la santé liés aux denrées alimentaires et aux aliments pour animaux. Il est important de noter que ce réseau d'alerte est basé sur la détection de virus dans des produits alimentaires sans lien avec des données épidémiologiques ou cliniques. En 2018, des NoVs et des virus de l'hépatite A ont été détectés 56 fois (33 en 2017) dans des aliments variés (11 produits différents), produits ou importés dans 30 pays (Figure 8).

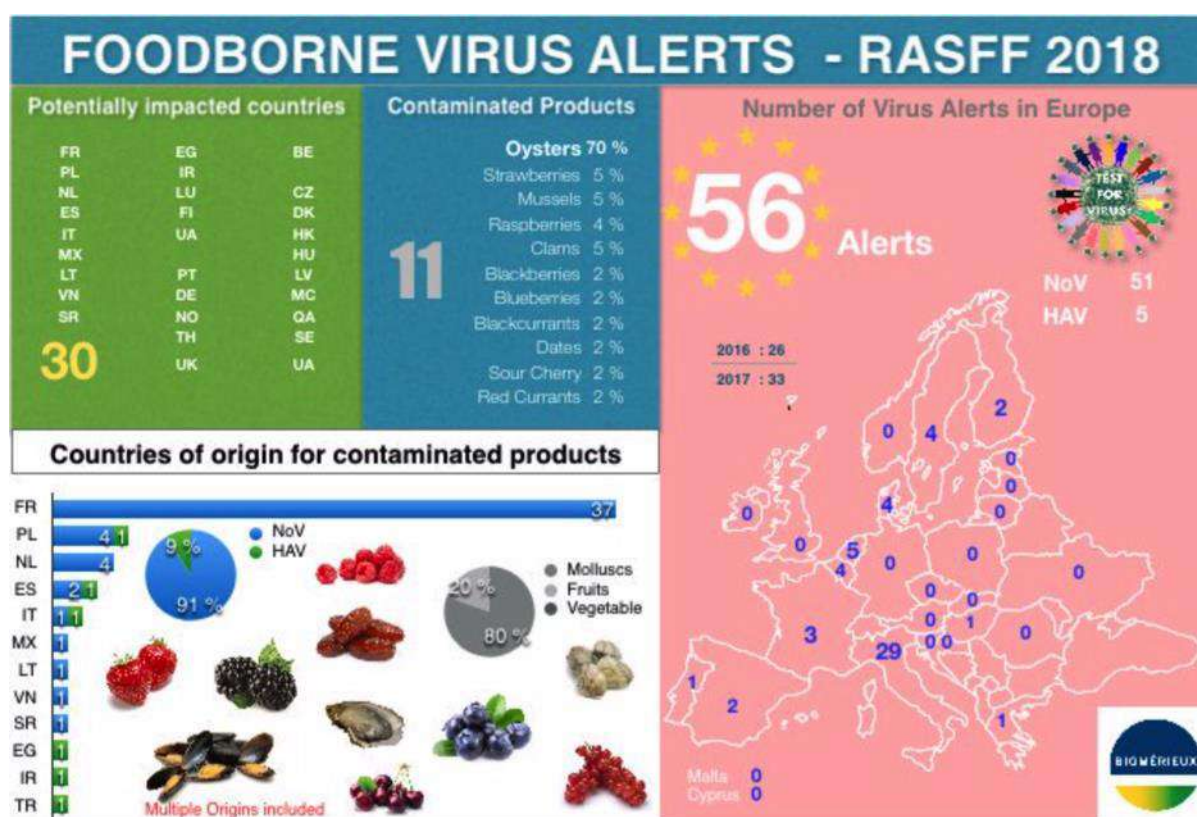


Figure 8 Alerte RASFF (<https://twitter.com/ceeramtools>)

2.5 CYCLE DE REPLICATION

La réplication virale est un procédé de biosynthèse au cours duquel le virus entre en interaction avec une cellule hôte afin d'y introduire son génome viral, détourne la machinerie cellulaire à son profit, pour fabriquer de nouveaux virus. Les étapes de réplication varient selon les caractéristiques du virus. Comme de nombreux autres virus à ARN, le génome des NoVs permet une réplication rapide, mais non encore totalement élucidée. La première étape dans le cycle de réplication est l'entrée en contact du virus et de la cellule (résumé dans la figure 9). Pour les NoVs cette phase d'attachement se fait par

reconnaissance de structures carbohydrates connues comme les HBGA, selon une spécificité liée aux génogroupes (Ruvoën-Clouet *et al.*, 2013), et vraisemblablement également de façon aspécifique par reconnaissance d'acides sialiques ou d'autres structures aujourd'hui encore inconnues qui permettraient la pénétration du virus à l'intérieur de la cellule (Almand *et al.*, 2017). La capsid virale est ensuite dégradée et seul le génome est libéré dans le cytoplasme cellulaire. Les données disponibles sont essentiellement basées sur la réplication du virus murin et les travaux pour comprendre la réplication des souches humaines se poursuivent (Estes *et al.*, 2019). La réplication des NoVs se situe dans les entérocytes comme montré par la détection d'antigènes et d'ARN viraux dans des biopsies de patients infectés. Ensuite, même si les détails précis ne sont pas encore élucidés après production des divers éléments constitutifs du virion, la capacité d'auto-assemblage des VP1 et les interactions avec VP2, les génomes sont encapsidés avant libération des nouveaux virus, probablement par apoptose cellulaire (Morillo and Timenetsky, 2011; Thorne and Goodfellow, 2014).

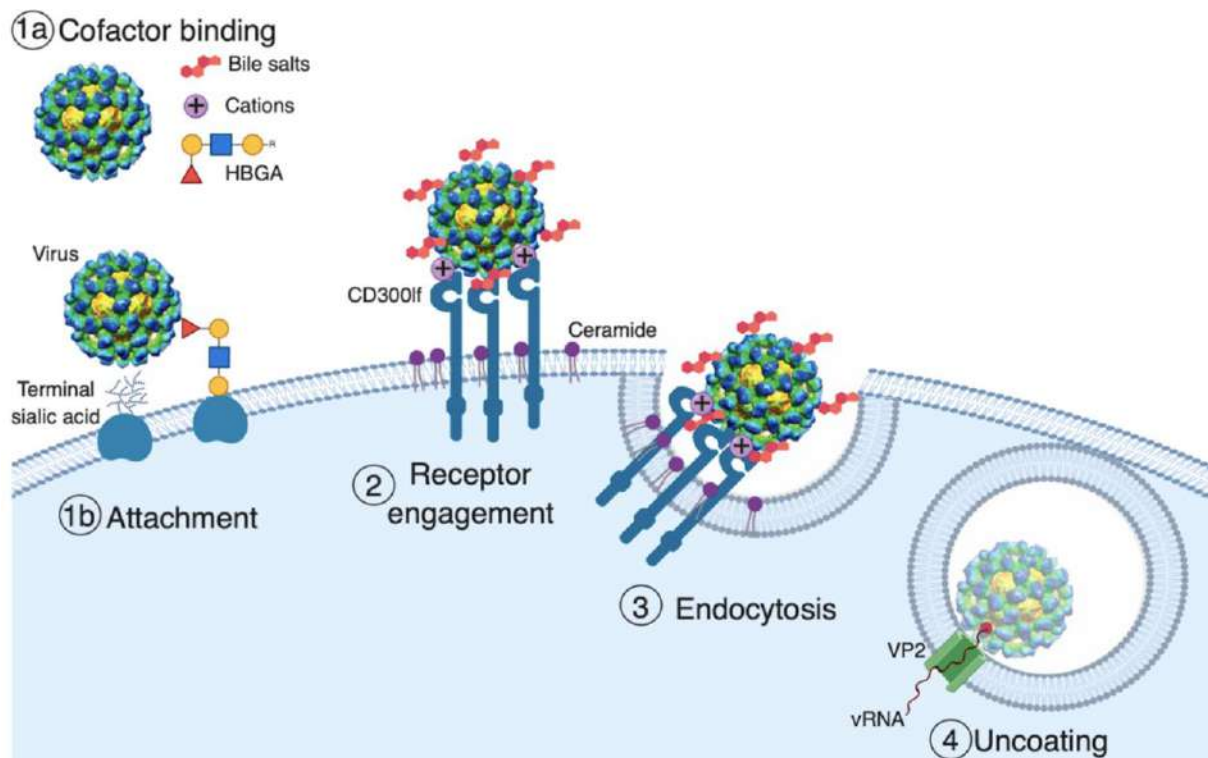


Figure 9 Modèle d'entrée du NoVs.

La première étape est l'attachement du virus à la surface de la cellule, qui peut être facilité par les glycanes hôtes associés aux cellules, y compris les HBGA. La deuxième étape est l'engagement du récepteur, protéine membranaire contenant un domaine d'immunoglobuline (Ig) encore inconnu pour les NoVs humains. Après l'engagement du récepteur, le virus pénètre par endocytose. Si le calicivirus félin constitue un bon modèle, la liaison au récepteur déclenche la protéine de capsid mineure VP2 pour former un portail membranaire susceptible de permettre la libération du génome viral dans le cytosol (Graziano *et al.*, 2019).

2.6 IMMUNOLOGIE

Pendant longtemps l'immunité induite par une infection à NoVs est restée difficile à comprendre. La réponse immunitaire est généralement de courte durée et souche-dépendante rendant possible une nouvelle infection chez un même individu (Dolin *et al.*, 1975; Ruvoën-Clouet *et al.*, 2013; Zhu *et al.*, 2013). En cas de déficit immunitaire, l'infection par les NoVs peut perdurer pour des longues périodes, voir plus d'un an (Nilsson *et al.*,

2003). La chronicité de la maladie étant due, à l'impossibilité du système immunitaire d'éliminer les virus (De Graaf *et al.*, 2016). Chez l'hôte immunocompétent la réponse immunitaire suite à une infection est de courte durée et souche-dépendante (Atmar *et al.*, 2018). Les nouvelles souches échappent plus facilement à la réponse immunitaire de l'organisme hôte, qui induit une pression de sélection antigénique sur le sous-domaine P2 de la VP1, connu pour sa variabilité (Atmar, 2010; Vinjé, 2014).

2.7 EPIDEMIOLOGIE

Tous les groupes d'âge sont susceptibles d'être contaminés par les NoVs, la majorité de la morbidité et de la mortalité concerne la population pédiatrique et les personnes âgées (Robilotti *et al.*, 2015). Les gastro-entérites à NoVs sont aussi communément appelées «*winter vomiting disease*», en raison de la prévalence plus élevée d'infections à NoVs enregistrées pendant les mois d'hiver, principalement de novembre à avril. Pour expliquer cette tendance, il semble que la température, l'humidité et les conditions climatiques en général pourraient jouer un rôle (Hall, 2012). Dans l'hémisphère nord, rassemblant les pays pour lesquels le plus de données sont disponibles, la saison associée à la pluie et à la natalité semblent constituer les facteurs déterminants (Ahmed *et al.*, 2013). La transmission est facilitée par les activités intérieures (Grøndahl-Rosado *et al.*, 2014). Une corrélation avec le début de l'année scolaire a également été mise en évidence pour les épidémies chez les enfants (Kraut *et al.*, 2017).

Depuis 1999 les données concernant les épidémies causées par les virus entériques ont été collectées dans 13 pays de l'Union Européenne et partagées dans un réseau informel appelé FBVE (Foodborne Viruses in Europe). En 2008 les bases des données ont été séparées selon les différents agents pathogènes et NoroNet a pris le relais pour les NoVs avec un nombre croissant de pays participants (<https://www.rivm.nl/en/noronet>). J'ai représenté les données collectées par ce réseau entre janvier 2005 et janvier 2019 à l'aide du logiciel RStudio v1.2.13 dans le but d'étudier la tendance temporelle des épidémies aux NoVs (Figure 10) et la variabilité des souches impliquées (Figure et Figure). Ce suivi met en évidence les variations épidémiologiques. Par exemple, l'hiver 2012-2013 se différencie avec un pic très important de cas de gastroentérites, liés à l'apparition d'un recombinant de la souche GII.4 avec la polymérase GII.Pe (Figure 11 et 12). Une évolution soudaine similaire a été rapportée par la souche GII.17 dans l'hiver 2015-2016, confirmant la définition des *souches évolutives* de Parra et collègues (2017). Si cette souche s'est rapidement répandue dans des nombreuses villes d'Europe, États-Unis et Australie, elle n'a pas remplacé la souche GII.4 (De Graaf *et al.*, 2016). Bien que la majorité des épidémies soient liées aux GII, l'hiver 2016-2017 a montré une importante augmentation des infections liées à des souches de NoVs GI.

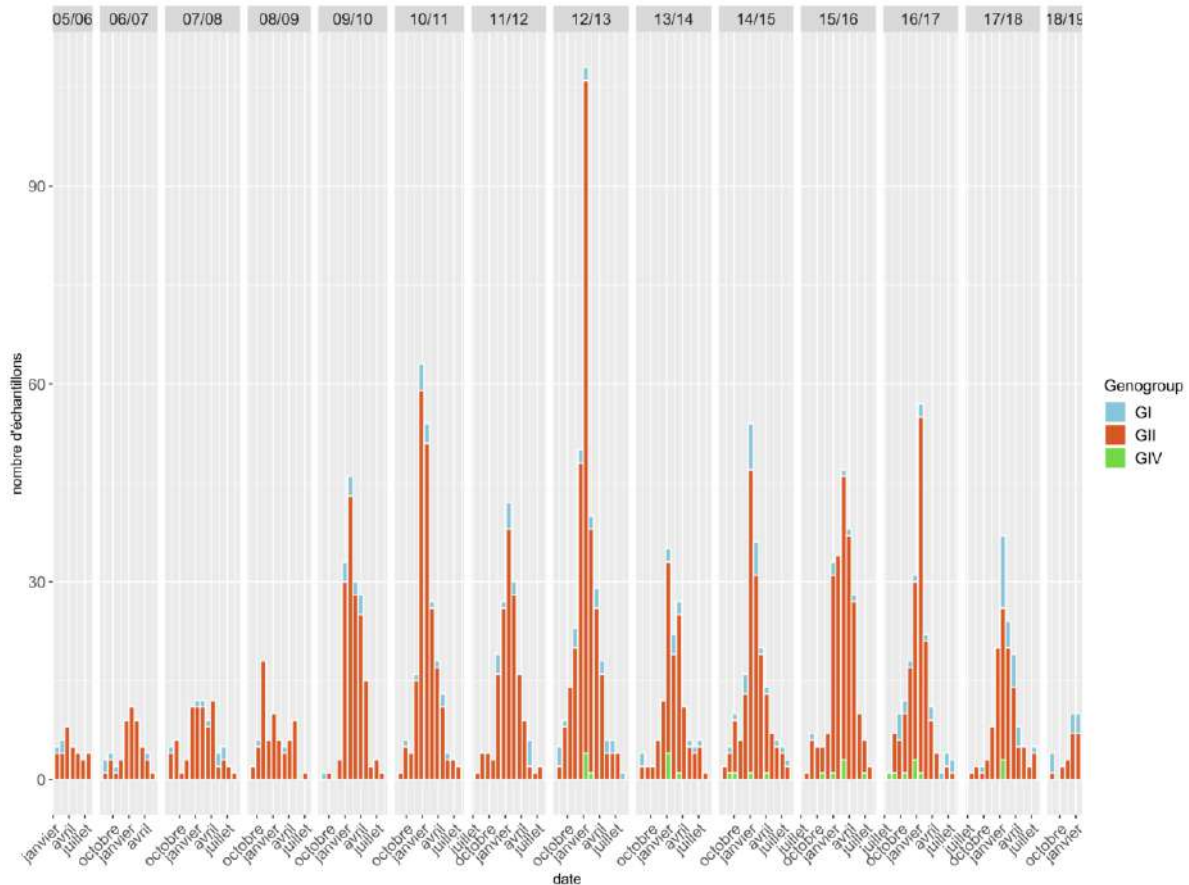


Figure 10 Variabilité temporelle de NoVs GI (bleu), GII (rouge) et GIV (vert) à partir de janvier 2005 (05) jusqu'à janvier 2019 (19).

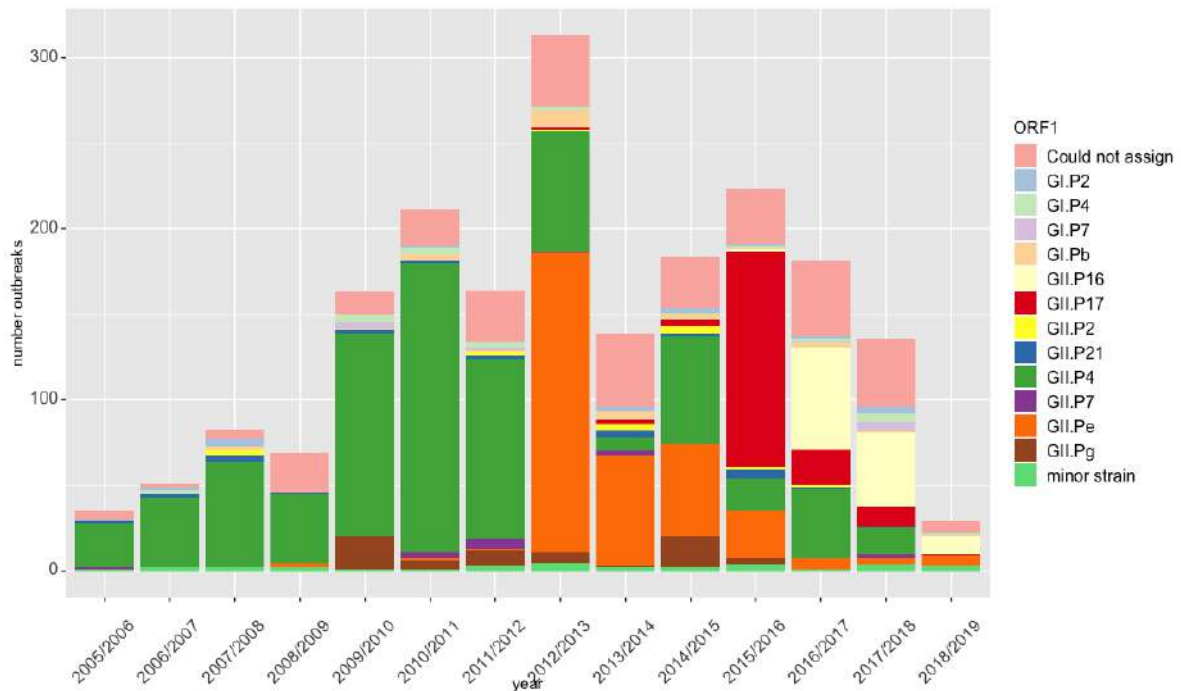


Figure 11 Variabilité de la région ORF1 des souches impliquées dans des épidémies entre janvier 2005 et janvier 2019. Could not assign : souches pour lesquels le typage n'a pas été possible. Minor strain : Nombre d'épidémies reportées < 10.

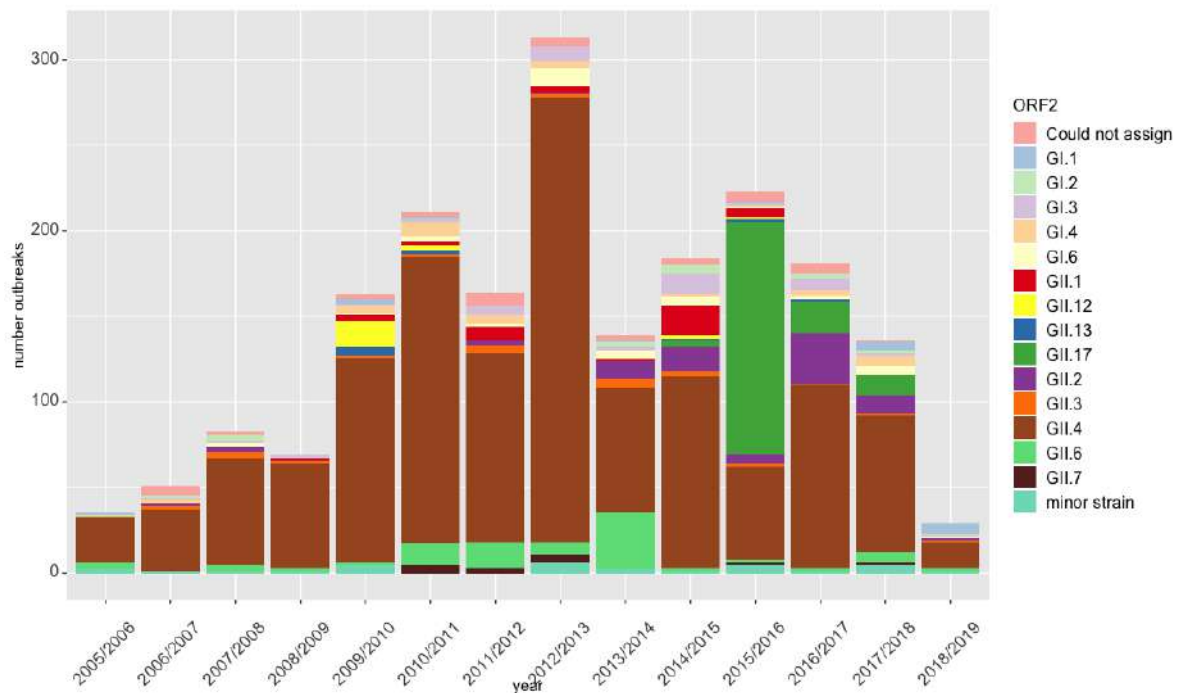


Figure 12 Variabilité de la region ORF2 des souches impliquées dans des épidémies entre janvier 2005 et janvier 2019. Could not assign : souches pour lesquels le typage n'a pas été possible. Minor strain : Nombre d'épidémies reportées < 10.

3. LES VIRUS ENTERIQUES HUMAINS ET L'ENVIRONNEMENT

3.1 VIRUS ENTERIQUES

Les rejets humains contiennent un mélange de matières, de débris et une variété de produits chimiques issus des activités résidentielles, commerciales et industrielles (Okoh *et al.*, 2010). Le plus souvent ces rejets vont être acheminés par un réseau collectif vers les stations d'épuration. Ces rejets, appelés plus largement eaux usées ou eaux brutes, représentent, d'un point de vue microbiologique, la diversité des micro-organismes circulant dans la population locale, pouvant représenter un danger pour l'homme (organismes pathogènes) ou pas (flore intestinale, portage sain de virus..) (Fernandez-Cassi *et al.*, 2017; Osunmakinde *et al.*, 2018).

Les virus entériques humains sont ainsi appelés pour leur cycle de multiplication qui comprend l'excrétion dans les selles des individus infectés. Ces virus sont essentiellement des petits virus (autour de 30 nM), non enveloppés et possédant un génome à ARN. Leur taille et l'absence d'enveloppe (la présence d'une enveloppe constituant un facteur de fragilité) leurs procurent des capacités de résistance importante et favorisent leur transmission par l'environnement (Metcalf, 1995). En particulier les poliovirus, responsable d'épidémies dévastatrices autrefois dans nos pays et malheureusement toujours présents dans certains pays, ont été les premiers à démontrer l'importance de l'environnement pour leur transmission (George, 1916; Metcalf, 1995; Sadeuh-Mba *et al.*, 2019). Après contamination féco-orale et multiplication dans l'intestin, ces virus migrent vers d'autres organes (comme le système nerveux central) et sont responsables d'un large éventail d'infections chez l'homme présentant divers syndromes cliniques (paralyse, cardiopathie, méningite...) (Montazeri *et al.*, 2015).

Les virus excrétés dans les selles humaines appartiennent à de nombreuses familles virales et induisent des pathologies variées (Tableau 3) :

- Gastroentérite : *Caliciviridae* (NoVs et sapovirus), *Astroviridae*, *Rotaviridae*, *Adénoviridae*, *Picornaviridae*.
- Hépatite : *Picornaviridae* (virus de l'hépatite A) et *Hepeviridae* (virus de l'hépatite E).
- Maladies diverses : *Picornaviridae* (poliovirus, coxsackievirus, échovirus...).

Tableau 3 Principaux virus entériques humains rejetés dans l'environnement (Adapté de Parnaudeau et al., 2014)

Nom	Capside	Taille capside	Génome	Taille génome	Saison
Calicivirus	icosaèdre	27-32nM	ARNsb*	7642pb*	hiver
Astrovirus	Icosaèdre	27-32nM	ARNsb	6797pb	hiver
Rotavirus	Icosaèdre, triple	70nM	ARNdb**	11 gènes de 3302-667pb	Hiver
Adénovirus	Complexe	70nM	ADNdb	35900pb	Toute l'année
Aïchi virus	Icosaèdre	27-32nM	ARNsb	8251pb	Toute l'année
V. Hépatite A	Icosaèdre	27-32nM	ARNsb	7478pb	Toute l'année
V. Hépatite E	Icosaèdre	27-32nM	ARNsb	7500pb	Toute l'année
Entérovirus	icosaèdre	20-30nM	ARNsb	7200pb	été

*sb : simple brin

**db : double brin

***pb : pair de base

3.2 NOV_s DANS L'ENVIRONNEMENT

Les concentrations dans les eaux usées changent notablement au cours de l'année, avec des concentrations plus faibles pendant l'été (entre $< 10^3$ et 10^6 copies d'ARN/L) et des pics en hiver (entre 10^5 et 10^9 copies d'ARN/L) (Kazama *et al.*, 2016; Schaeffer *et al.*, 2018). En corrélation avec l'épidémiologie, la concentration de NoVs GII est toujours significativement plus élevée que la concentration de NoVs GI, phénomène indépendant de la région géographique et de la saison (Figure 13) (Eftim *et al.*, 2017). Le traitement des eaux usées est donc important pour la santé de toute communauté. Cet objectif de diminuer les NoVs doit être conjugué à l'objectif de l'élimination des polluants chimiques et biologiques afin de réduire les impacts négatifs pour l'environnement, et la contamination des aliments (Hoque *et al.*, 2019).

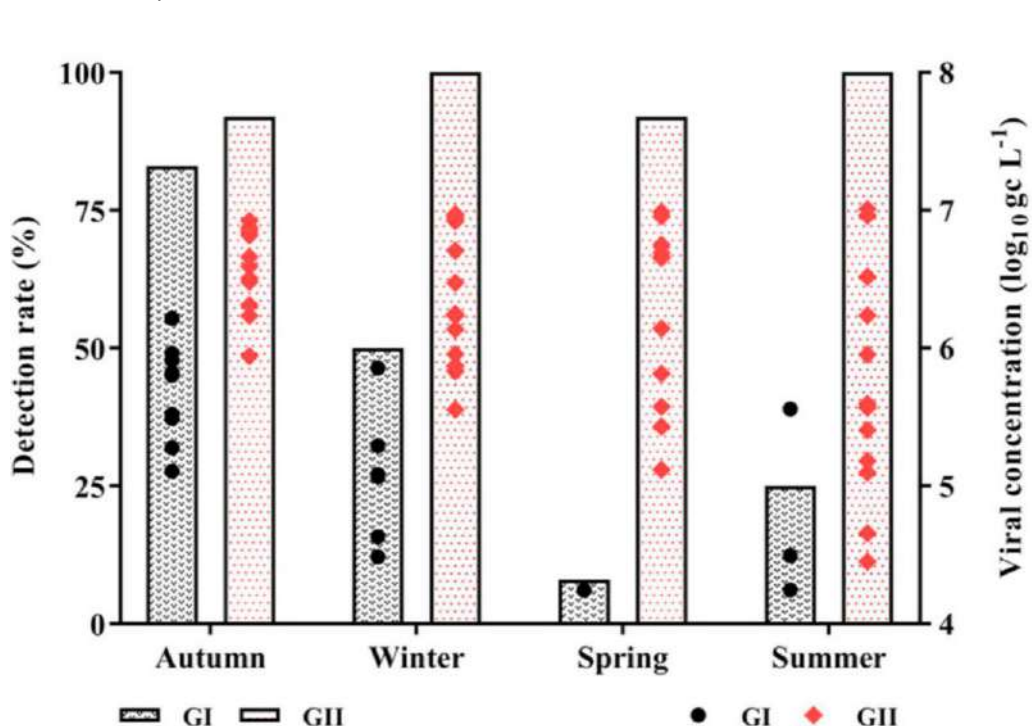


Figure 13 Distribution saisonnière et concentration de NoVs GI (noir) et GII (rouge) dans les échantillons d'eaux usées (Adapté de Fumian *et al.*, 2019)

En effet, de par leur capacité à s'infiltrer dans les sols ou leur résistance importante, les NoVs peuvent contaminer les eaux souterraines, les rivières et estuaires (Adriaenssens *et al.*, 2018; Kauppinen *et al.*, 2018; Sedji *et al.*, 2018). Outre cet apport par les bassins versants, les NoVs peuvent atteindre l'écosystème marin via des fosses septiques défaillantes, des débordements de station d'épuration ou encore des inondations (Maalouf *et al.*, 2010). Une fois rejetées dans l'environnement côtier, les particules virales sont soumises à un processus de sédimentation et/ou de dilution dû aux marées et aux conditions physico-chimiques du milieu marin, telles que l'irradiation par le soleil, la salinité, la température et le pH (Pommepuy *et al.*, 2005; Cook *et al.*, 2016).

3.3 PRINCIPES ET FONCTIONNEMENT DES SYSTEMES D'EPURATION

L'épuration des eaux usées est un ensemble de techniques qui font appel à différents traitements physiques, chimiques et biologiques afin de recycler ou purifier les eaux usées avant de les réintroduire dans l'environnement (le plus souvent dans les eaux de surface ou parfois en milieu côtier) (Figure 14).

Ils existent différents procédés de traitement des eaux usées : individuels (pour les habitations situées en dehors des agglomérations) ou collectifs. Parmi les traitements collectifs, le lagunage naturel tend à disparaître en raison des difficultés d'entretien, du manque de place disponible et de sa faible efficacité pour l'élimination des micro-organismes dont les virus entériques humains (Schaeffer *et al.*, 2018).

Les étapes principales d'épuration comprennent une étape préliminaire d'élimination des déchets plus volumineux par passage de l'eau à travers des grilles ou des filtres à sable (dégrillage ou dessablage) suivie par le traitement primaire par décantation des eaux permettant le dépôt et l'élimination des parties solides (boues), et secondaire dans des bassins d'aération où le développement d'organismes biologiques dégrade la pollution carbonée et azotée par voie oxydative.

Les eaux traitées sont rejetées dans le milieu naturel alors que les boues en excès peuvent être utilisées pour l'épandage en agriculture ou encore dans la production de biogaz, après une étape d'élimination de l'eau et d'inactivation des particules virales à la chaux (<http://www.natura-sciences.com/environnement/station-depuration-biologique.html>).

Plusieurs traitements tertiaires basés sur le principe de la filtration sont possibles pour diminuer les matières en suspension et la pollution organique avant que les eaux usées soient rejetées dans l'environnement ou réutilisées pour l'irrigation agricole et l'arrosage d'espaces verts. Les procédés les plus courants sont la microfiltration (MF), l'ultrafiltration (UF), la nanofiltration (NF), l'osmose inverse (OI) et la filtration sur milieu granulaire (sable, anthracite...) (Boutin, 2009). L'ultrafiltration est un des procédés les plus utilisés ; elle utilise la séparation membranaire en utilisant la pression pour éliminer les particules en suspension dans l'eau (Abdessemed, Nezzal and Aïm, 1999). Si les principaux avantages sont la rétention des principaux composants des eaux usées (azote, du phosphate total, des coliformes totaux..), la simplicité et le respect de l'environnement (car ne prévoit pas l'utilisation de substances chimiques), le problème principal est le colmatage de la membrane qui provoque une diminution du flux de perméat dans le temps et une limitation de l'efficacité de séparation (Salladini *et al.*, 2007).

Les différents procédés d'épuration

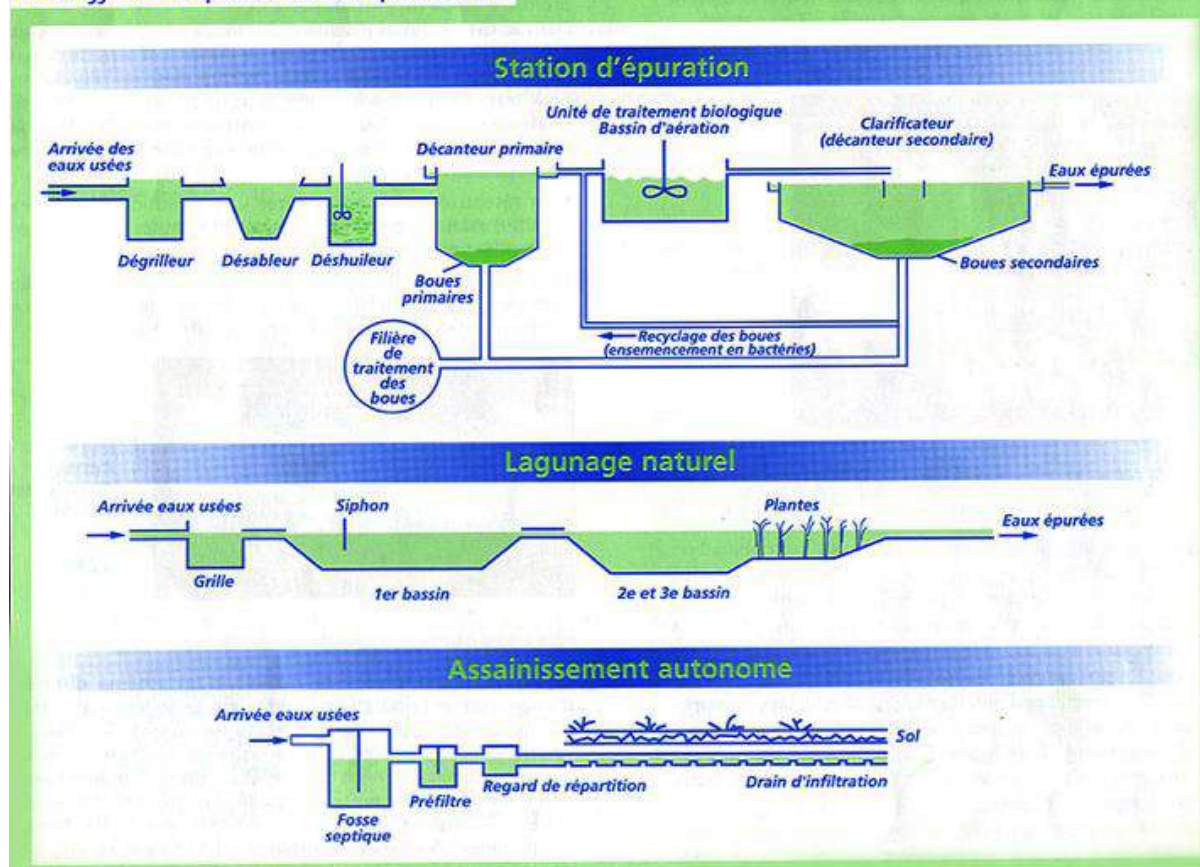


Figure 14 Schéma des différents procédés d'épuration

3.4 EFFICACITE DES TRAITEMENTS D'EPURATION

L'efficacité des traitements d'épuration est actuellement contrôlée par l'utilisation d'indicateurs bactériens, définies en trois classes principales selon le rapport de la *World Health Organisation* sur les indicateurs de la qualité microbiologique de l'eau (Ashbolt, Grabow and Snozzi, 2001) :

Indicateurs microbiens généraux de processus : groupes d'organismes qui démontre l'efficacité d'un processus, telles que les bactéries hétérotrophes totales ou les coliformes totaux pour la désinfection au chlore.

Indicateurs de contamination fécale : tels que les groupes bactériens, coliformes thermotolérants ou *E. coli*. Leur présence signale que des agents pathogènes peuvent être présents.

Organismes Index et organismes modèles : groupe/espèce indicative de la présence et du comportement de l'agent pathogène, tels que *E. coli* en tant qu'indice pour *Salmonella* et les coliphages ARN-F cultivables en tant que modèles de virus entériques humains.

Si les traitements mis en place dans les stations d'épuration réduisent le nombre des bactéries indicateurs de contamination fécale, avec une efficacité de 98,76% pour les coliformes totaux et de 99,29% pour les coliformes thermotolérants (Branco *et al.*, 2016), des virus infectieux ont été isolés dans des environnements aquatiques répondant aux normes des indicateurs bactériens (Okoh, Sibanda and Gusha, 2010; Bisseux *et al.*, 2018). Les virus sont résistants au traitement UV et à la chloration, deux stratégies courantes de

désinfection dans le traitement secondaire des eaux usées et leurs agrégations ou leurs adhésions aux particules en suspension réduit l'efficacité des traitements (Montazeri *et al.*, 2015). Ces systèmes d'épurations montrant leurs limites pour l'élimination des virus, ils sont souvent complétés par un traitement tertiaire d'ultrafiltration (Sima *et al.*, 2011; Schaeffer *et al.*, 2018). Cependant aucun système de traitement des eaux usées ne peut garantir l'élimination totale de ces virus entériques humains (Sano *et al.*, 2016; Gerba *et al.*, 2018).

4. NOROVIRUS ET HUITRE

“J'adore les huîtres : on a l'impression d'embrasser la mer sur la bouche.”

Léon-Paul Fargue

4.1 LES HOMMES ET LES HUITRES, UNE HISTOIRE ANCIENNE

Depuis toujours la mer représente pour l'homme un milieu de ressources. De nombreuses traces fossiles témoignent du rôle des coquillages comme source d'alimentation pour les peuples anciens ou primitifs habitants le littoral. Les huîtres représentaient pour les hommes préhistoriques un aliment facilement accessible, abondant et riche d'un point de vue nutritionnel. Au sud de la péninsule ibérique, des restes de coquillages consommés par les hommes de Néandertal, il y a 150 000 ans ont été trouvés (Cortés-Sánchez *et al.*, 2011). Après consommation, les coquilles pouvaient être utilisées pour élaborer des objets destinés à la vie quotidienne, rituelle ou religieuse. Nombreux sont les restes de coquilles retrouvés dans les tombes afin d'orner, nourrir et protéger les êtres chers après leur mort (Prieur, 2008; Bonnardin, 2012).

Les grecques conféraient aux huîtres des propriétés aphrodisiaques et utilisaient les coquilles comme bulletin de vote en y gravant le nom de l'homme politique qu'ils souhaitaient bannir de la cité (d'où le terme ostracisme) (<http://www.ostreiculteurtraditionnel.fr/histoire/histoire.html>).

Les premiers ostréiculteurs semblent remonter à plus de 3000 ans et étaient situés au sud de l'actuel Extrême-Orient russe, où des amas de coquilles, principalement *Crassostrea gigas* (huître creuse dite japonaise), ont été retrouvés sur les rives d'une lagune d'eau salée. En Europe, les Romains sont les fondateurs de l'ostréiculture en créant le premier parc à huîtres et en maîtrisant tout le cycle de vie. De la fin de l'Empire romain à la naissance de l'ostréiculture moderne française en 1860, l'huître plate *Ostrea edulis* était la seule espèce consommée (Lescroart, 2017). Plusieurs épizooties successives ont fortement réduit la production d'huîtres plates à partir de 1920, date à partir de laquelle l'huître portugaise *Crassostrea angulata* est devenue l'espèce très majoritairement produite. En 1970 une nouvelle épizootie éradique l'huître portugaise des côtes françaises. L'huître japonaise est introduite en France notamment grâce à l'opération « Résur » en 1971 (Bernard, 2011). Aujourd'hui, *Crassostrea gigas* représente 99% de la production ostréicole française, qui se situe au premier rang Européen et au troisième rang mondial (<http://www.ostrea.org>).

Rapidement les huîtres sont apparues comme un vecteur possible de maladies infectieuses et une note dans les Comptes Rendus de l'Académie des Sciences en 1912, décrit les bassins de purification construits à Paris pour les purifier (Fabre-Domergue, 1912). En 1913, des restaurateurs parisiens fondent l'AEIO (Association d'Encouragement des Industries Ostréicoles), pour inciter les producteurs à purifier leurs produits et à réaliser les premiers contrôles sanitaires. A compter de 1923, l'état Français prend le premier décret rendant le contrôle sanitaire des huîtres obligatoire (Garry, 2015). Depuis 1939, il existe une obligation de classement des zones de production selon leur qualité microbiologique. Aujourd'hui, le chapitre II de l'annexe II du règlement (CE) n° 854/2004 prévoit un classement de l'ensemble des zones de production de coquillages en trois catégories selon leur qualité microbiologique (A, B ou C). Ce classement est établi selon des critères de concentration de

la bactérie indicatrice de contamination fécale *Escherichia coli* dans les coquillages (Piquet, 2018). Les huîtres peuvent être commercialisées directement si elles proviennent de zones classées A, mais doivent être purifiées dans des bassins à terre si elles proviennent de zone classées B. Pour les zones classées C, les coquillages doivent être reparquées pour une longue durée ou traités thermiquement avant de pouvoir être consommés.

4.2 ANATOMIE DE L'HUÎTRE

L'huître est un bivalve et sa coquille est constituée de deux valves articulées par une charnière et qui s'ouvrent et se ferment sous l'action du muscle adducteur (Figure 15). Ce système permet à l'huître de se protéger des prédateurs et de maintenir sa coquille fermée lors des marées basses, quand elle se retrouve hors de l'eau. Le manteau, constitué d'un tissu conjonctif, adhère à la coquille et participe à sa formation en sécrétant du carbonate de calcium (Miossec et *al.*, 2009). L'huître, est un organisme filtreur, filtrant l'eau de mer pour capter l'oxygène et les particules nécessaires à son alimentation. Les mouvements des branchies dirigent les particules vers les palpes labiaux qui vont effectuer un premier tri des particules, certaines étant rejetées directement dans les pseudo-fèces, les autres, présentant un intérêt nutritif, étant dirigées vers la bouche. Les tissus digestif (TD), comprennent l'estomac et un intestin, et se terminant avec un anus qui permet l'élimination des déchets. De mai à août, lors de la phase de reproduction, la masse gonadique est largement diffusée dans toute la chair et atteint jusqu'à 70% du poids total de l'animale adulte, l'huître est alors dite « *laiteuse* ». Le cœur envoie l'hémolymphe, liquide circulant dont le rôle est analogue au sang, vers les branchies où il s'enrichit d'oxygène (Lescroart, 2017).

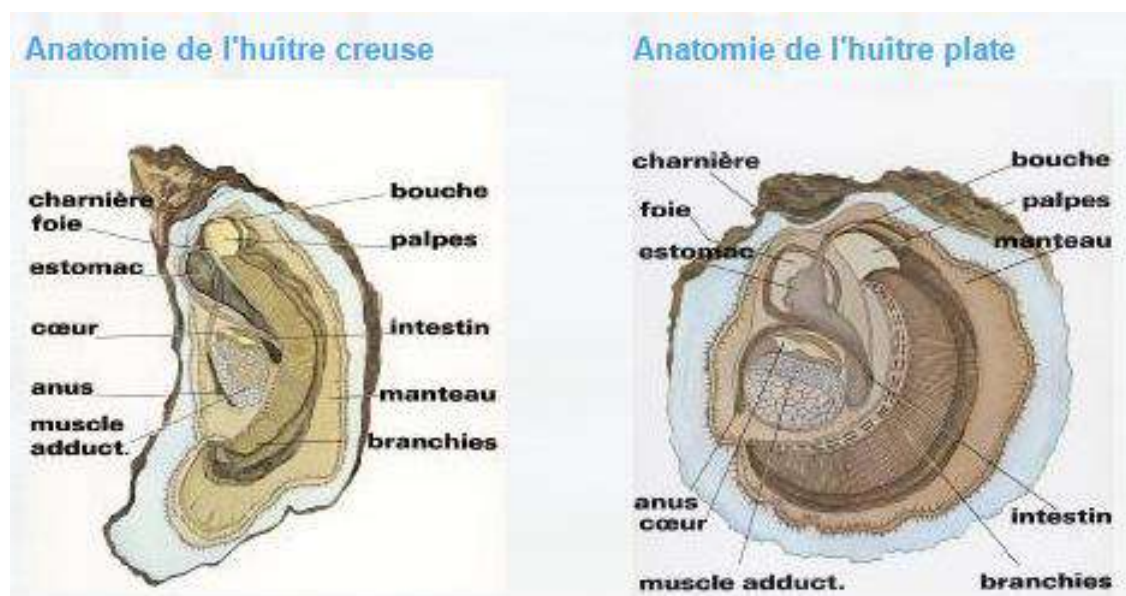


Figure 15 Anatomie de l'huître creuse et de l'huître plate (Lescroart et al., 2017).

4.3 LES HUÎTRES ET LES EPIDEMIES DE GASTROENTERITE

Si la mise en place de la réglementation Européenne basée sur l'indicateur bactérien a permis de diminuer les épidémies bactériennes suite à la consommation d'huîtres, les épidémies de gastro-entérites virales persistent et les NoVs constituent l'agent pathogène le plus fréquemment impliqué dans ces épidémies. Pour cela plusieurs facteurs y contribuent comme leur consommation le plus souvent crues, leur capacité à accumuler les NoVs et leurs zones d'élevage situées le plus souvent à proximité des côtes (McLeod *et al.*, 2017). Une revue systématique des épidémies de gastroentérites survenues entre janvier 2003 et juillet 2017 a été menée afin d'identifier les aliments couramment associés aux épidémies de NoVs. Les aliments les plus fréquemment impliqués étaient les fruits de mer (61%), dont 89% étaient des huîtres; les salades et les framboises étaient également reportées (Hardstaff *et al.*, 2018).

Des chercheurs de l'université de Shanghai ont étudié la diversité des séquences de NoVs impliquées dans des épidémies liées à la consommation d'huître entre 1983 et 2014 au niveau mondial (Yu *et al.*, 2015). En analysant les séquences déposées dans les banques de données du National Center for Biotechnology Information (NCBI) et NoroNet suite à la survenue d'épidémies liées à la consommation d'huître, ils ont montré une très grande diversité de séquences de NoV démontrant que plus de 80% des souches de NoVs rejetées par l'homme ont par la suite été détectées dans les huîtres. Les pays côtiers, traditionnellement producteurs et consommateurs d'huîtres sont les principaux pays décrivant ces épidémies. Leurs études montrent également une fréquence plus élevée de NoV GI dans les épidémies liées à la consommation d'huîtres par rapport aux autres modes de transmission confirmant les travaux développés au laboratoire (Le Guyader *et al.*, 2012). Les rejets humains insuffisamment traités entraînent la contamination des huîtres éventuellement par plusieurs souches de NoVs ce qui pourrait favoriser l'apparition de souches recombinantes chez les consommateurs (Le Guyader *et al.*, 2012; McIntre *et al.*, 2012; Yu *et al.*, 2015) (Figure 16). Pour mieux comprendre les enjeux à la base de ce mécanisme, une étude de modélisation a été menée considérant une population côtière de 2000 individus avec 50% de consommateurs et 50% de non consommateurs pendant la période hivernale (Thebault *et al.*, 2014). Pour les individus sécréteurs, le risque d'avoir 50% de la population malade (DI_{50}) est de 32 copies (médiane) pour GI et de 5 copies pour GII, les non sécréteurs présentant une probabilité d'infection de 100 à 1000 fois plus faible.

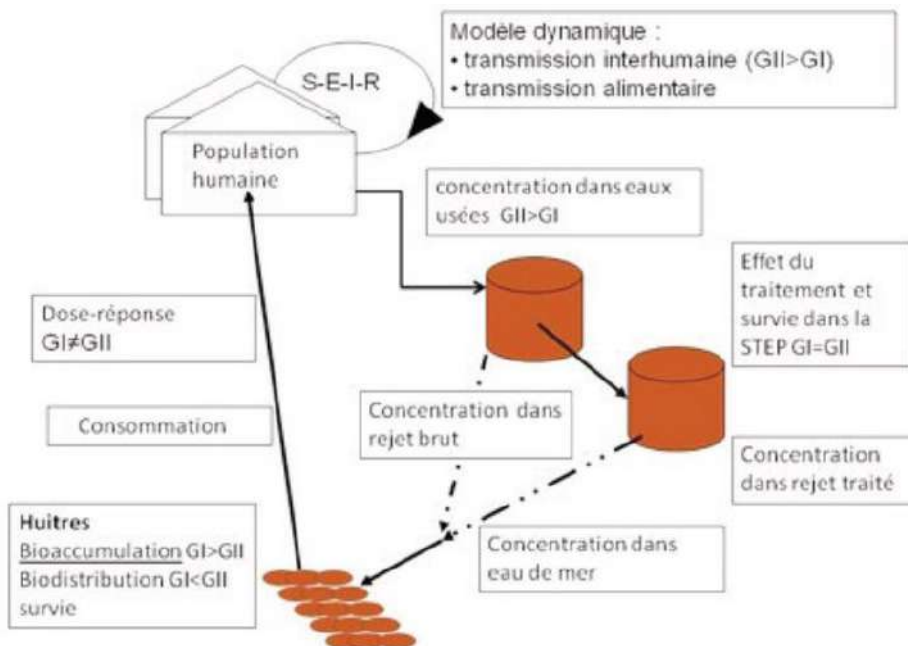


Figure 16 Représentation du modèle probabiliste de type SEIR qui considère qu'un individu passe par plusieurs états : Susceptible (S), Infecté (E), Infectieux (I) et Immunisée (R) (Thebault et al., 2014)

Les NoVs peuvent persister plusieurs semaines dans les huîtres et résister aux traitements de purification qui sont efficaces seulement pour éliminer les bactéries, mais inefficaces pour les virus (Hassard *et al.*, 2017). La purification est un traitement post-récolte impliquant le placement des coquillages dans des eaux de mer propres afin de diminuer la contamination bactérienne. Le temps de purification n'est pas réglementé, seul le résultat final importe et doit correspondre aux critères réglementaires, (<230 *E.coli* pour 100g de chair et liquide intervalvaire (CLI)) (Règlement (CE) n° 853/2004). La durée varie selon la charge initiale en bactérie entre 24 à 48 heures, mais parfois des durées beaucoup plus longues peuvent être appliquées (Lees, 2010; EFSA, 2019). Si des études ne montrent pas de réduction dans la concentration en NoVs, d'autres travaux ont montré qu'un temps plus long (entre 9 et 45,5 jours) permettait d'obtenir une réduction d'un log (Nappier *et al.*, 2008; Polo *et al.*, 2014; McLeod *et al.*, 2017). L'augmentation de la température de l'eau de mer a été suggérée pour accélérer l'élimination des NoVs (Doré *et al.*, 2010).

4.4 DES LIGANDS SPECIFIQUES POUR LES NoVs DANS LES HUITRES

La longue persistance, la prédominance de certaines souches dans les huîtres supposant une sélection par l'huître par rapport aux souches de NoVs rejetées par l'homme, et la mise en évidence chez l'homme de ligand de type glycanes a amené le laboratoire à analyser plus en détails ce qui pouvait se passer dans l'huître. Après des contaminations au laboratoire en utilisant des souches identifiées nous avons pu ainsi localiser précisément les particules virales (Figure 17) (Thèse de F. Loisy, travaux réalisés avec l'équipe de J. Le Pendu).

Par la suite, les travaux réalisés lors de la thèse d'H. Maalouf ont ensuite permis de caractériser des ligands différents dans les tissus d'huîtres selon les souches comme observés chez l'homme et une variation saisonnière de l'expression de ces ligands (Maalouf *et al.*, 2010 ; Ruvoën *et al.*, 2013). Ces travaux ont également permis de mettre en évidence une efficacité de bioaccumulation différente selon les souches de NoVs disponibles au laboratoire ou lors d'étude sur le terrain (Zakhour *et al.*, 2010; Maalouf *et al.*, 2011). Ces

observations nous ont permis de conclure que l'huître était capable de sélectionner spécifiquement un pathogène humain (Le Guyader *et al.*, 2012).

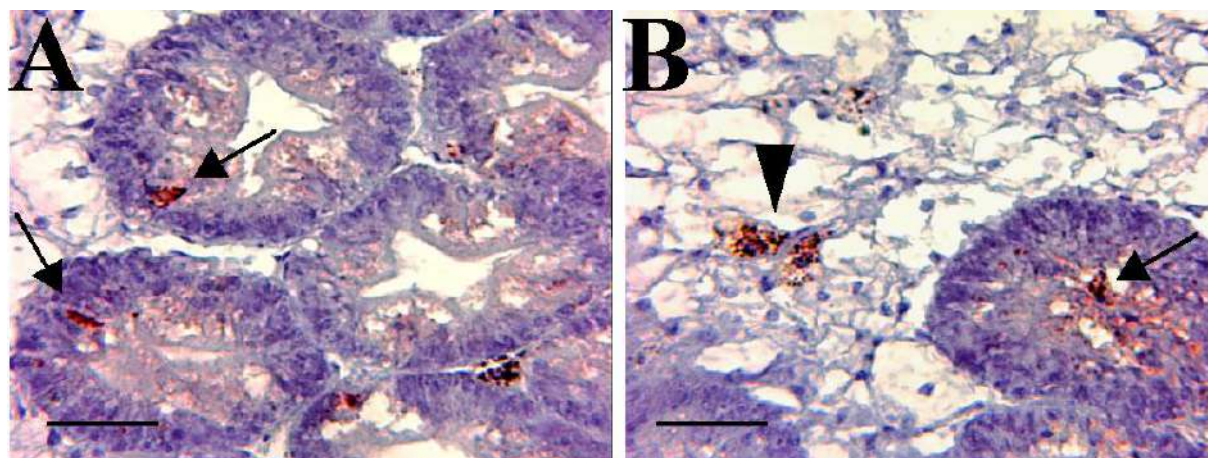


Figure 17 Localisation des NoVs GI.1 au niveau du tissu digestif de l'huître. Des huîtres ont été mise en contact avec des pseudo-particules virales (VLPs) de la souche Norwalk (GI.1). Après fixation du tissu digestif en paraffine et préparation de coupes de tissus, la présence de particules virales a été révélée par un anticorps monoclonal, puis coloration (technique d'immunohistochimie). A) Les flèches montrent la présence de particules virales (coloration rouge) détectées dans les cellules intraépithéliales des tubules digestifs (A) et dans le tissu conjonctif et dans la lumière d'un tubule digestif (B) (Modifié d'après Le Guyader *et al.*, 2006).

4.5 EPIDEMIES DE NOVs ET HUITRES

Dans les épidémies de NoVs liées à la consommation d'huîtres, plusieurs souches peuvent être détectées chez les patients infectés et chez les huîtres correspondantes.

La contamination par plusieurs souches différentes peut s'expliquer par une contamination par les eaux usées (Le Guyader *et al.*, 2006). La persistance des NoVs dans l'eau (jusqu'à 30 jours) et leur capacité d'être transportés sur des longues distances (>10Km) facilite la contamination des zones de production ostréicoles par les NoVs ainsi que l'accumulation d'une variété de souches différentes dans les tissus d'huître (Hassard *et al.*, 2017). Selon le site de production et la localisation des points de rejet, un modèle mathématique pourrait aider à prévenir les épidémies en prédisant les périodes à fort risque de contamination pour les huîtres (Wang and Deng, 2016).

La présence de NoVs dans les coquillages a fait l'objet de nombreuses études depuis des années. Le Tableau 4 synthétise les résultats d'études au niveau mondial, en sélectionnant les études présentant un nombre d'échantillons analysés relativement important ($n > 100$) et la méthode de détection utilisée. Ces résultats soulignent la nécessité d'une surveillance régulière afin de limiter le risque de contamination virale.

Tableau 4 Exemple d'études sur la prévalence des NoVs dans les coquillages

Pays	Coquillages	Période étudiée	Nombre échant.	Echant. positif en		Référence
				NoV GI	NoV GII	
Maroc	Huîtres Palourdes	Nov. 2015-fév. 2017	102	Pas fait	7	(El Moqri <i>et al.</i> , 2019)
UK	Huîtres	Mars 2015-mars 2016	630	433*		(Lowther <i>et al.</i> , 2019)
Chine	Huîtres ¹	2016	463	2	94	(Tan <i>et al.</i> , 2018)
Australie	Huîtres	Juil. 2014-août 2015	300	0	0	(Torok <i>et al.</i> , 2018)
Irlande	Huîtres	Jan. 2015-avril 2016	120	108*		(Rupnik <i>et al.</i> , 2018)
France	Huîtres, moules et coques	Fév. 2013-janv. 2015	150	33*		(Rincé <i>et al.</i> , 2018)
Italie	Moules et palourdes	Mars-sept 2015	384	Pas fait	259	(La Rosa <i>et al.</i> , 2017)
Japon	Huîtres	Oct. 2015-fev. 2016	200	9	72	(Imamura <i>et al.</i> , 2017)
Italie	Moules	Janv. 2014-déc. 2015	108	10	22	(Fusco <i>et al.</i> , 2017)
Thaïlande	Huîtres, moules, coques ²	Aout 2011-juil. 2012	300	29**	14**	(Kittigul <i>et al.</i> , 2016)
Espagne	Moules, palourdes et coques	Janv. 2011-juin 2012	168	54	43	(Polo <i>et al.</i> , 2015)
Australie	Huîtres ³	Fév. 2010-déc. 2012	120	0	2	(Brake <i>et al.</i> , 2015)
UK	Huîtres	Mai 2009-avril 2011	844	578**	467**	(Lowther <i>et al.</i> , 2012)
Espagne, Grèce et Finlande	Moules	Mai-déc. 2010	153	1	25	(Diez-Valcarce <i>et al.</i> , 2012)

1: méthode ISO (15216-1) et kit Qiagen

2: broyage de 2 g de TD, élution glycine, purification chloroforme et concentration sous vide, extraction AN kit Qiagen

3: méthode ISO (15216-1) et kit Roche

* : résultat publié sans mention du génogroupe,

** : certains échantillons sont positifs pour les deux génogroupes.

Pour aller plus loin dans l'analyse de la diversité génétique des souches de NoVs en circulation Yu *et al.*, (2015) ont étudié toutes les souches de NoVs repertoriées dans les banques de données GenBank (NCBI) et NoroNet entre 1983 et 2014 (Tableau 5).

Sur 1077 séquences, 665 ont été retenues suite au contrôle qualité appliqué et ont été soumises à génotypage et analyse phylogénétique. Plus de 80% des génotypes de NoVs humains ont été détectés dans des échantillons d'huîtres.

Tableau 5. Variabilité génétique (ORF1 et ORF2) et prévalence des souches NoVs GI et GII identifiées dans les huîtres impliquées dans des épidémies de gastroentérite (Adapté de Yu *et al.*, 2015).

ORF1	PREVALENCE	ORF2	PREVALENCE
GI.PNA	3.77%	GI.5	1.75%
GI.PA	3.77%	GI.7	5.85%
GI.P3	3.77%	GI.2	5.85%
GI.P7	5.66%	GI.NA	8.19%
GI.P4	11.32%	GI.6	8.77%
GI.P1	11.32%	GI.1	11.11%
GI.PB	13.21%	GI.3	13.45%
GI.P2	41.51%	GI.4	45.03%
GII.P2	3.16%	GII.NA	1.74%
GII.P7	5.26%	GII.8	2.90%
GII.P3	5.26%	GII.7	3.19%
GII.PNA	6.32%	GII.13	4.93%
GII.P12	6.32%	GII.12	5.80%
GII.P21	7.37%	GII.1	6.38%
GII.PE	8.42%	GII.5	8.12%
GII.PG	13.68%	GII.2	8.41%
GII.P4	32.63%	GII.6	9.57%
		GII.3	16.23%
		GII.4	29.28%

Les nouvelles souches de NoVs sont rapidement détectées dans les huîtres, comme démontré en Corée du sud avec la mise en évidence des GII.17 et GII.21 dans les eaux usées et les huîtres (Koo *et al.*, 2017). Lors de l'émergence du variant GIIP17-GII.17, un lien épidémiologique direct a été établi entre la consommation d'huîtres françaises importées et des consommateurs danois (Rasmussen *et al.*, 2016). Cet épisode montre l'impact que les échanges d'aliments entre pays peut avoir sur l'épidémiologie de ces virus entériques.

En Europe, considérant le nombre d'épidémies liées à la consommation d'huître, une étude sur la prévalence des NoVs dans les huîtres prélevées en zone de production ou dans les centres d'expédition a été réalisée à la demande de la Commission Européenne. Au total, 2180 échantillons ont été prélevés dans des zones de production et 2129 dans des centres d'expédition, pendant deux années consécutives. La France étant le premier producteur européen, 937 échantillons ont été prélevés dans les zones de productions de notre littoral et 1809 échantillons dans les centres d'expédition français. La prévalence dans les zones de production était estimée à 34,5 % (IC : 30,1 à 39,1 %) à l'échelle Européenne et à 13% en France. Pour les lots provenant des centres d'expédition cette prévalence était de 10,8 % (IC : 8,2 à 14,4 %) pour la totalité de l'étude et inférieure à 7% en France. Comme attendu un impact saisonnier a été observé avec une contamination plus fréquente en hiver, lié au rejet du virus par la population (Figure 18) (Efsa, 2019).

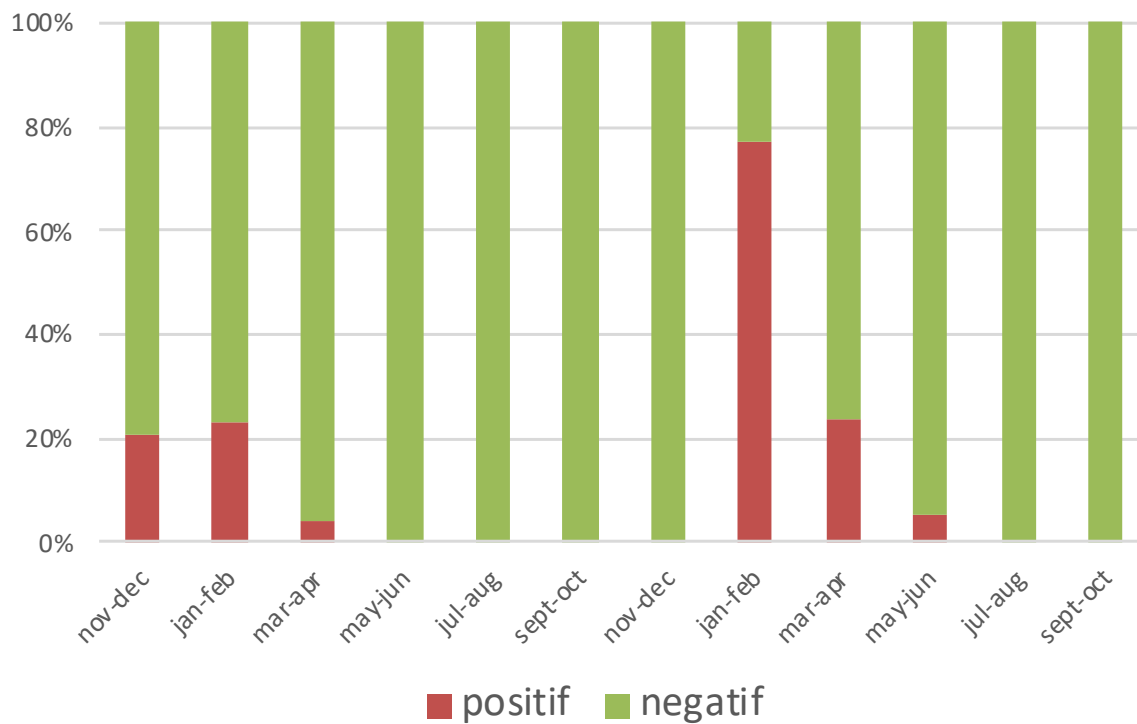


Figure 18 Détection des NoVs dans les zones de production française. Données recueillies entre novembre 2016 et octobre 2018, avec des prélèvements bimensuel (Données Efsa, non publiées)

5. METHODES DE DETECTION

5.1 ELUTION VIRALE

I- EAU

L'analyse d'échantillons environnementaux nécessite une méthode suffisamment sensible pour détecter un faible nombre de particules virales ayant une petite taille et diluées dans les eaux de surface, usées ou de mer. De façon générale, les méthodes utilisées reposent sur les caractéristiques physico-chimiques du virus cible qui varient selon les protéines constituant la capsid virale et qui affectent la taille et la charge des protéines de la capsid. Les diamètres des virus entériques varient de 28-30 nM (enterovirus et NoVs) à 100 nM (adénovirus) et les points isoélectriques oscillent entre pH 3 pour le virus de l'hépatite A et $\text{pH} \geq 9.5$ pour les NoVs (Vongpunsawad *et al.*, 2013).

a) Concentration

Les volumes des échantillons analysés peuvent varier de quelques mL à plusieurs centaines de litres selon les protocoles utilisés (Ikner *et al.*, 2012). Afin de réduire les volumes de départ pour leur analyse, des étapes de concentration primaire et secondaire sont appliquées (Figure 19). Une bonne méthode de concentration, devrait satisfaire certains critères : être techniquement simple, rapide, permettre une récupération virale élevée, être adaptée à un large éventail de virus entériques, fournir un faible volume de concentré et être peu coûteuse (Bosch *et al.*, 2009). Toutefois, aucune méthode ne répond à tous ces critères et il n'existe pas de méthode de référence.

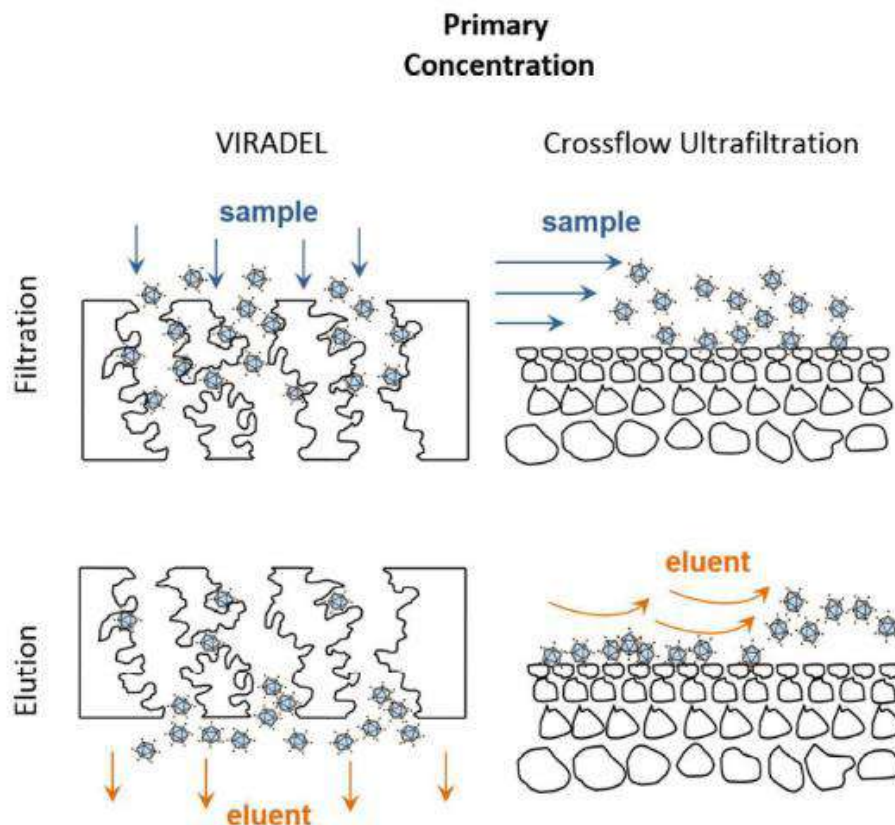


Figure 19 Représentation schématisée des étapes de concentration primaire avec la méthode VIRADEL (VIRus Adsorption-Elution) et ultrafiltration tangentielle (adaptée de Shi *et al.*, 2017)

Concentration primaire

Cette étape a pour but de réduire de très larges volumes d'eau en passant généralement par une phase de filtration et une d'éluion par ajustement du pH (selon le point isoélectrique du virus ciblé). La concentration primaire représente l'obstacle majeur en raison d'un processus souvent long et de la perte possible des virus d'intérêts (Shi *et al.*, 2017). Les deux méthodes les plus souvent utilisées sont schématisées dans la Figure :

Les **méthodes d'adsorption/éluion**, qui reposent sur les interactions de charge ionique entre les virus et la surface des filtres. Considérant que les virus entériques présents dans l'eau ont généralement une charge négative, des filtres dits électronégatifs ou électropositifs peuvent être utilisés. Les premiers requièrent un conditionnement de l'échantillon avec ajustement de la conductivité et du pH, représentant donc une limite sur le volume de l'échantillon à traiter; les seconds, pour lesquels le pré-conditionnement de l'échantillon n'est pas nécessaire, permettent la filtration de plus de 1000L d'eau mais sont plus coûteux. La nature de l'échantillon est une des limites les plus importantes dans l'application des méthodes d'adsorption/éluion. La filtration d'eaux turbides ou riches en microparticules, comme les eaux usées, ne sont guère envisageable pour de tels volumes car elles provoquent rapidement l'obstruction des filtres (Cashdollar *et al.*, 2013).

L'**ultrafiltration** s'appuie sur l'exclusion par la taille. Le principe repose sur le passage de grands volumes d'eau à travers des capillaires, des fibres creuses ou encore à travers des membranes plates avec des pores de taille variable. Un flux tangentiel limite le colmatage des filtres. L'eau et les molécules de bas poids moléculaire (30-100 kDa) passent à travers les fibres alors que les substances plus larges sont arrêtées et constituent le retentât. Bien que l'efficacité de récupération virale soit variable (entre 2 et 98%), une étude utilisant un NoVs murin a montré un taux de récupération jusqu'à 74% dans 100 L d'eau de surface (Gibson and Schwab, 2011). Toutefois, un débit de filtration très lent et l'équipement coûteux constituent des désavantages importants (John *et al.*, 2011). De plus, une perte importante de la concentration des NoVs peut s'observer dû à leur capacité d'adsorption sur la matière particulaire ou sur les bactéries (données du laboratoire non publiées).

Concentration secondaire

La concentration secondaire permet une réduction encore plus importante des volumes avant la suite de l'analyse. Parmi les stratégies les plus communément utilisées on peut citer:

La **floculation**, processus impliquant un coagulant ajouté à l'eau pour faciliter la liaison entre les particules, créant des agrégats plus gros et donc plus faciles à séparer. Les coagulants peuvent être chimiques (par exemple à base de fer) ou organiques (en utilisant des protéines de lait pré-floculées) (Calgua *et al.*, 2008; John *et al.*, 2011). Les particules virales précipitées forment un culot qui sera remis en suspension dans des solutions tampons. Cette technique s'adapte à des échantillons de petit volume (entre 0.5 et 1 L) présentant une faible turbidité. L'inconvénient de l'utilisation de coagulants à base de fer repose essentiellement dans le risque d'interférences lors de l'extraction des acides nucléiques (Jacobsson and Schultz, 2014). Les coagulants organiques ont permis de détecter une large variété de famille virale, incluant les *Caliciviridae*, à partir d'échantillons d'eaux usées, mais

en utilisant des volumes limités (500 mL). En effet la même méthode appliquée sur de plus grands volumes (10 L) a entraîné la concentration d'inhibiteurs affectant l'amplification par PCR ou les enzymes utilisées pour le séquençage par métagénomique (Fernandez-Cassi *et al.*, 2017; Hjelmsø *et al.*, 2017).

La **précipitation chimique** des protéines permet la récupération de virus infectieux, comme cela a été rapporté pour les rotavirus, entérovirus et pour le virus de l'hépatite A (Lewis and Metcalf, 1988). Parfois en raison de la persistance éventuelle de cytotoxicité, des étapes supplémentaires de purification sont nécessaires pour la mise en évidence de virus infectieux (Falman *et al.*, 2019). Cette méthode utilise le polyéthylène glycol (PEG)/chlorure de sodium (NaCl), pour former un maillage où les particules virales seront retenues selon leur taille et précipitées suite à une centrifugation réfrigérée. Le culot est ensuite remis en suspension dans une solution tampon (PBS ou glycine) (Bosch *et al.*, 2008 ; Petterson *et al.*, 2015). Le PEG/NaCl est utilisé pour concentrer une grande variété de virus, mais son utilisation est limitée à de faibles volumes (jusqu'à 100 mL). Cette méthode est efficace pour détecter et quantifier les virus par PCR et/ou pour des analyses métagénomiques. Le PEG peut aussi être utilisé avec un autre polymère de différent poids moléculaire, comme par exemple le dextran. Cette solution interagit avec l'échantillon d'eau et, après agitation vigoureuse, se sépare sur la base des différents poids, entraînant les virus de poids moléculaire similaire (Matrajt *et al.*, 2018). L'Organisation Mondiale de la Santé recommande cette méthode de concentration/séparation pour l'identification du poliovirus (World Health Organization, 2003).

b) Elution

Suite aux étapes de concentration des particules virales, l'élution constitue une phase critique car elle peut avoir un impact très important sur la quantité finale de virus récupéré et sur la concentration des inhibiteurs de la PCR (Shi *et al.*, 2017). A cet effet des solutions alcalines sont utilisées pour favoriser la répulsion électrostatique entre la surface de la membrane filtrante et les particules virales qui seront ainsi détachées. Le taux de récupération virale dépend de plusieurs facteurs :

- la nature de l'éluant utilisé, qui peut être organique (extrait de bœuf 1.5-3%, pH de 7.5 à 9.5) ou chimique (glycine 0.05M, pH de 9 à 11.5). Les composants organiques peuvent inhiber les réactions de PCR (Shi *et al.*, 2017) ;
- la rapidité du flux d'élution. Une élution lente (0.5 mL/min) favorise une meilleure récupération virale (Borrego *et al.*, 1991) ;
- le nombre des passages de l'éluat. Un bon taux de récupération des NoVs (environ 99%) a été obtenu en réalisant un second passage de l'éluat par rapport à la 1^{ère} élution (0.1%) (Gibbons *et al.*, 2010).

II- HUÎTRES

a) Dissection

L'élution et l'extraction des particules virales à partir des tissus d'huître demandent une préparation en amont afin d'isoler les tissus ciblés pour l'analyse. Les huîtres, le plus souvent entre 6 à 10 individus pour éviter les variabilités individuelles, sont lavées, ouvertes, et la chair totale est pesée. Selon les analyses envisagées, différents tissus peuvent être disséqués (branchies, manteaux, palpes labiaux...) à l'aide d'un scalpel stérile. Pour la détection des virus entériques humains il a été démontré que les tissus digestifs (TD) concentrent 90% de la contamination, facilitant ainsi leur détection (Atmar *et al.*, 1995). Les travaux réalisés dans l'équipe ont par la suite confirmé cette observation pour les NoVs GI et GII (Maalouf *et al.*, 2010 ; Maalouf *et al.*, 2011). Pendant ma thèse les TD ont donc été sélectionnés comme matrice pour l'élution virale. Après dissection, les TD sont coupés en petits morceaux (environ 1-2mm) et congelés (-20°C) sous forme d'aliquots de 2 g ou directement analysés (Le Guyader *et al.*, 2009).

b) Elution

Les tissus peuvent être traités par :

Digestion enzymatique en utilisant une protéase telle que la protéinase K afin de décomposer la matière organique associée à la matrice (Jothikumar *et al.*, 2005). L'activité enzymatique est favorisée par un traitement thermique (Uhrbrand *et al.*, 2010) et les débris tissulaires sont sédimentés par centrifugation et éliminés, alors que le surnageant, contenant les particules virales, est récupéré et utilisé pour l'extraction des acides nucléiques (Jothikumar *et al.*, 2005; Le Mennec *et al.*, 2017).

Cette technique a été validée comme méthode ISO (15216 partie I et II) pour la détection des NoVs dans les mollusques bivalves (Lowther *et al.*, 2019).

Précipitation chimique, parallèlement au protocole adopté dans les échantillons d'eau, cette technique de concentration vise à réduire le volume de l'échantillon et à éliminer une partie de la matrice sans perte de particules virales. Une phase primaire d'homogénéisation de l'échantillon (par exemple avec un potter ou un ultra-turax) est appliquée pour broyer les tissus, suivie par une étape d'élution virale par variation de pH et/ou des conditions ioniques de l'échantillon. L'étape de concentration virale utilise le plus souvent une précipitation chimique par le PEG 6000. Le culot sera ensuite remis en suspension dans du PBS ou dans de l'eau stérile pour l'extraction des acides nucléiques viraux (Atmar *et al.*, 1995; Le Guyader *et al.*, 2009). Bien que cette méthode nécessite plus d'étapes, avec un risque potentiel de perte de particules virales, elle présente l'avantage de ne pas endommager la capsid virale par l'utilisation d'une enzyme ou par un traitement thermique et permet la récupération des particules virales infectieuses (Polo *et al.*, 2018).

5.2 ELUTION CIBLEE DES NOV_s

Comme rapidement présenté, plusieurs méthodes montrent des avantages selon la matrice, le virus cible ou la technique d'analyse prévue, sans qu'aucune ne soit parfaite pour répondre à toutes les applications. Des combinaisons de différentes étapes peuvent parfois se révéler efficaces (Farkas *et al.*, 2018).

Les méthodes présentées ci-dessous ont été utilisées depuis de nombreuses années au sein du laboratoire pour détecter les NoV essentiellement par PCR (Figure 20). Pendant ma thèse, souhaitant appliquer des techniques novatrices de séquençage métagénomique (chapitre 6), une adaptation de ces protocoles a été nécessaire, comme rapporté dans la section « travaux originaux » de ce manuscrit.

Les échantillons d'eaux sont généralement concentrés par précipitation au PEG à partir d'un volume de 40mL préalablement ajusté en pH (4) et conductivité (>2000 μ S) (Mullendore *et al.*, 2001). Un tampon alcalin (glycine à pH 9) est ensuite utilisé pour remettre en suspension le culot et favoriser l'élution virale. Ensuite, la phase d'extraction des acides nucléiques est précédée par un traitement avec une solution de chloroforme-butanol afin de rompre les membranes lipidiques, en diminuant la charge bactérienne, les inhibiteurs de PCR et les virus enveloppés (Da Silva *et al.*, 2011; Conceição-Neto *et al.*, 2015; Hjelmsø *et al.*, 2017).

La digestion enzymatique des TD d'huître nécessite une double incubation, à 37°C et 60°C pour un bon taux de récupération (Jothikumar *et al.* 2005).

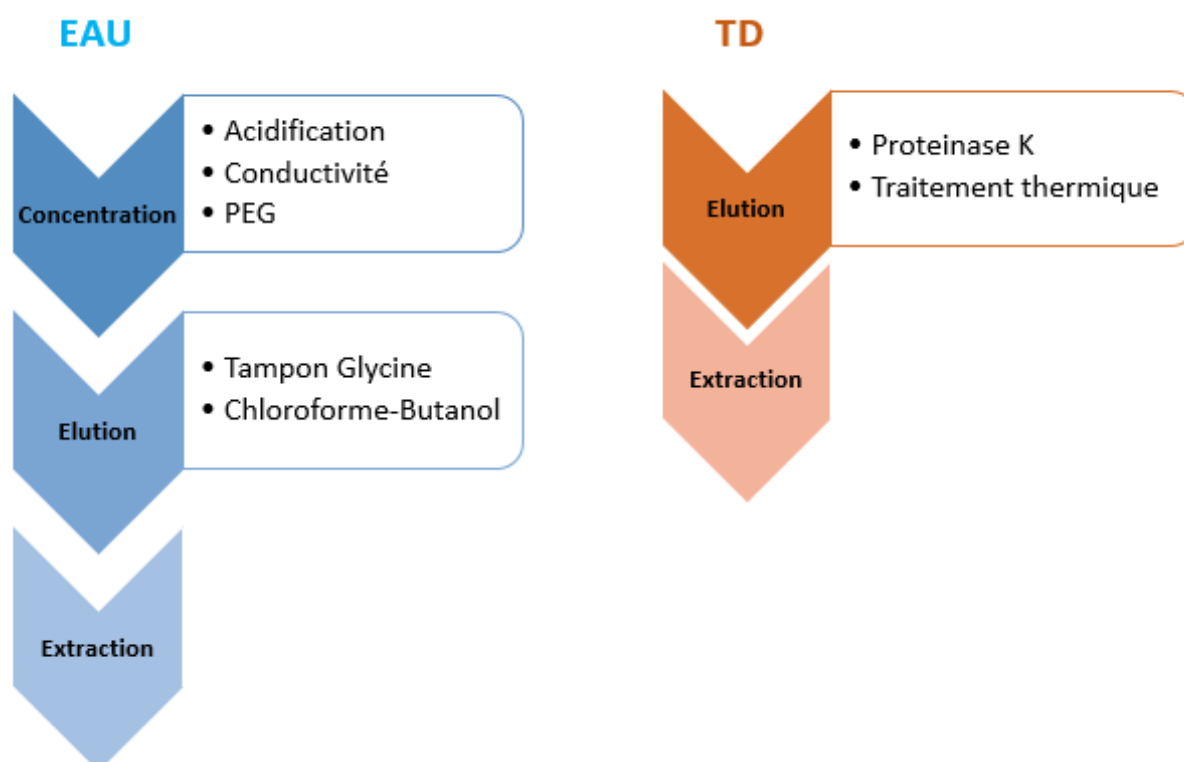


Figure 20 Résumé des étapes clés pour l'élution des particules virales de NoVs à partir d'échantillons d'eaux et d'huîtres.

5.3 EXTRACTION DES ACIDES NUCLÉIQUES ET QUANTIFICATION

L'ARN viral peut être extrait de la capsidie par traitement thermique (Schwab et al., 2001), enzymatique (Atmar *et al.*, 1995) ou chimique, à l'aide d'un agent dénaturant tel que la solution de guanidinium thiocyanate (Boom *et al.*, 1990). Le développement d'un kit commercial basé sur cette technologie facilite les étapes d'extraction et de purification des acides nucléiques. En raison de sa bonne efficacité, cette méthode d'extraction fait partie de la norme ISO (Le Guyader *et al.*, 2009; Lowther *et al.*, 2019).

Grâce aux avancements dans les techniques de biologie moléculaire, la quantification des copies génomes (cg) présents dans un échantillon est désormais devenue une technique préconisée marquant un pas important entre l'analyse qualitative, capable seulement de donner une information de présence/absence de l'agent pathogène ciblé, et l'analyse quantitative, malgré l'absence de standard universel.

La **RT-PCR en temps réel (rRT-PCR)** est une méthode sensible et spécifique disponible pour détecter l'ARN des NoVs dans des échantillons environnementaux, où la concentration virale est généralement assez faible. Après la transcription de l'ARN en ADN complémentaire (ADNc), l'enzyme Taq-polymérase amplifie la zone cible en effectuant un nombre défini de cycles d'amplifications. La présence d'une sonde marquée avec une molécule fluorescente permet de détecter le moment auquel la fluorescence franchit les seuils de sensibilité (*threshold*, en anglais) donnant la valeur de *Ct*. Cette valeur, associée à une courbe standard, permet de quantifier la concentration virale dans l'échantillon (Bustin *et al.*, 2009 ; Loisy *et al.*, 2005). En raison de la variabilité génomique, le problème pour les NoVs était de sélectionner une zone du génome conservée. De courtes régions conservées dans la jonction entre ORF1 et ORF2 ont été identifiées et permettent une détection efficace des souches des différents génogroupes constituant la base des différents tests développés jusqu'à présent (Kageyama *et al.*, 2003; Le Guyader *et al.*, 2009; Lees, 2010; Masago *et al.*, 2016). La contrainte majeure de cette technique est la production de gamme standard car aucune référence n'existe et les variations selon les laboratoires peuvent être importantes.

La **PCR digitale (dPCR)** adopte une approche différente pour quantifier le nombre de génomes viraux dans l'extrait d'AN. En effet la réaction d'amplification s'effectue en point final sous forme de milliers de petites réactions indépendantes (généralement dans l'ordre de nanolitres) (Figure 21). Idéalement, chaque mélange réactionnel contient une seule molécule cible, ou aucune, de façon à pouvoir déterminer le nombre de réactions positives et négatives et de mesurer une quantité absolue calculée sur base statistique, selon la loi de Poisson (Kuypers and Jerome, 2017). Le partitionnement peut être réalisé en utilisant des microplaques, des capillaires ou des émulsions dans l'huile, dans tous les cas un ratio entre puits négatif et positifs est nécessaire, il faut donc respecter un seuil de dilution des échantillons (Zhang *et al.*, 2015; Polo *et al.*, 2016; Persson *et al.*, 2018). En évitant les courbes standards et étant probablement moins sensible à l'inhibition, cette technique se révèle plus précise pour estimer des faibles concentrations (Monteiro and Santos, 2017).

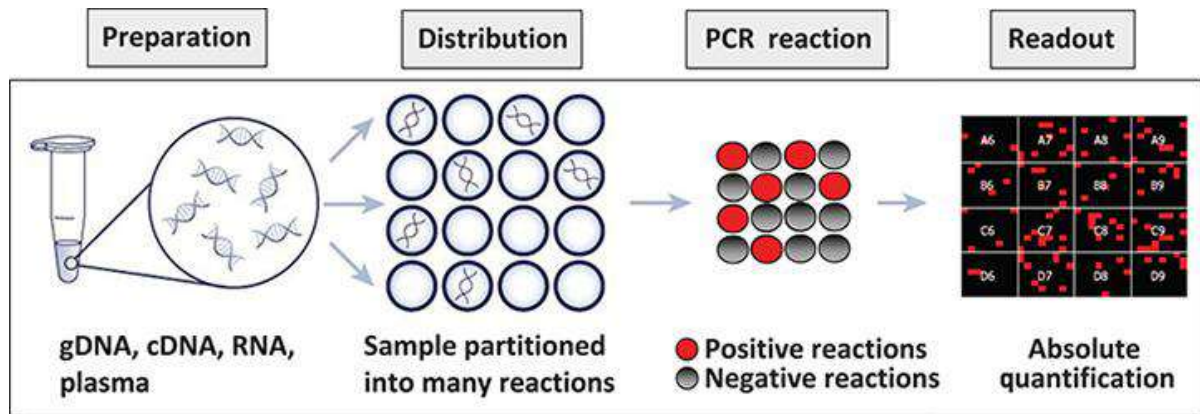


Figure 21 Schéma de fonctionnement de la digital PCR (<https://www.thermofisher.com/fr/fr/home/life-science/pcr/digital-pcr.html>)

6. SEQUENCAGE HAUT DEBIT

La répétition et recombinaison de seulement quatre nucléotides est le mécanisme à la base de la diversité, unicité et individualité de tous les organismes vivants. Les génomes peuvent alors être comparés à un livre écrit seulement avec quatre lettres, A, C, G, T et le séquençage haut débit (HTS, de l'anglais *High Throughput Sequencing*) à la lecture du contenu de ce livre.

6.1 LE SEQUENCAGE, DE GENERATION EN GENERATION

Les premiers séquenceurs de type Sanger, aussi appelés **séquenceurs de première génération**, développés il y a environ 40 ans permettaient la lecture de longs fragments (jusqu'à 1000 pb) avec un taux d'erreur relativement bas. Cette méthode nécessitait de disposer d'un ADN simple brin servant de matrice pour la synthèse enzymatique (DNA polymérase) du brin complémentaire. La détection des fragments ainsi synthétisés était déterminée par l'incorporation d'un traceur radioactif dans l'ADN synthétisé (Sanger and Coulson, 1975). Dans les années '90 les séquenceurs capillaires basés sur le passage des échantillons dans un tube capillaire (d'où leur nom) se sont développés, présentant l'avantage de l'utilisation d'un marquage fluorescent mais un temps d'analyse relativement long (Strazzulli *et al.*, 2017).

Plus performants, les séquenceurs de **deuxième génération** peuvent générer l'analyse concomitante de fragments plus courts (de 150 à 300 pb) en très grand nombre (jusqu'à 20 milliards de fragments à la fois). Différentes technologies font partie de ce groupe, chacune utilisant une méthode de séquençage différente (Tableau 5). Cependant, le principe général reste le même. Une préparation de l'échantillon est nécessaire afin de découper l'ADN en fragments de taille homogène et constituer ce que l'on appelle des *librairies*. Des séquences courtes appelées index (ou barcode) sont ajoutées aux extrémités pour l'identification des échantillons lors de leur séquençage simultané. Puis l'ajout d'adaptateurs permet de fixer les fragments au support de séquençage (flowcell) et d'initier leur clonage par la formation d'un brin d'ADN complémentaire afin d'augmenter le signal. L'amplification peut être réalisée sur un support solide (flowcell) ou en émulsion (sur billes préalablement préparées afin de capturer les adaptateurs). Pendant le séquençage, des cycles de lavages et de lecture se succèdent (wash & scan) avec incorporation successive des nucléotides et émission d'un signal (qui peut être lumineux ou chimique) à chaque incorporation (Schadt *et al.*, 2010). La succession des signaux permet de lire les séquences d'ADN, appelées *lectures* ou plus communément *reads*. Chaque fragment sera séquencé en parallèle résultant en un volume important des données en sortie (Shendure and Ji, 2008; Metzker, 2010; Buermans and den Dunnen, 2014).

Tableau 5 Comparaison des techniques de séquençage HTS de deuxième génération

Plate-forme	Technologie	Pré-amplification	Enzyme	Signal d'incorporation
IonTorrent® Termofisher	Détection ions H+	En émulsion	polymérase	Changement pH et libération H+
Illumina®	Séquençage par synthèse	Sur plaque	polymérase	Reconnaissance signal lumineux
Roche 454®	Pyrosequençage	En émulsion	polymérase	Reconnaissance signal lumineux
SOLid® Termofisher	Ligation	En émulsion	ligase	Reconnaissance signal lumineux

Les séquenceurs de **troisième génération**, sont plus sensibles et permettent la détection d'une seule molécule, diminuant l'intérêt de l'étape de pré-amplification. De plus, le séquençage est plus fluide, sans cycles d'interruption et d'incorporation de bases. Une lecture en temps réel du signal permet d'augmenter le débit de séquençage et la longueur des reads (entre 1000 et 10000 pb), réduisant ainsi la complexité de la préparation des échantillons et les coûts (Liu *et al.*, 2014). Cependant, la qualité et la concentration d'ADN demandé représentent un challenge pendant l'extraction des acides nucléiques (> 1µg ADN) comparé à ce qu'est demandé par le système Illumina (environ 500 ng ADN) (van Dijk *et al.*, 2018).

Les stratégies technologiques plus utilisées et toujours en cours de développement sont :

- le séquençage par synthèse, développé et commercialisé par Pacific Biosciences (PacBio) en 2011. Permet le séquençage en temps réel d'une seule molécule d'ADN par action d'une molécule d'ADN polymérase enduite sur la surface de séquençage. L'incorporation de chaque nucléotide marqué déclenche la libération d'un signal lumineux différent qui permet le codage de la séquence analysé (Eid *et al.*, 2009) ;
- le séquençage sur minION introduit en 2014 par Oxford Nanopore Technology (Jain *et al.*, 2015) est basé sur la détection de la perturbation du flux d'un courant électrique lors du passage des molécules d'ADN simple brin, ou d'ARN, à travers des nanopores biologiques (membranes protéiques) ou des pores synthétiques. Chaque nucléotide modifie le flux ionique de façon différente selon leur taille et génère une modulation de courant spécifique, avec une vitesse de séquençage augmentée de 30 bases/sec avec les premiers modèles, à 450 bases/sec avec les derniers développements (Brown *et al.*, 2016).

Le taux d'erreur élevé (entre 13% et 15%) est un inconvénient majeur pour ces technologies, mais la rapidité de préparation des bibliothèques, les faibles coûts et la possibilité de séquencer des reads longs (plusieurs dizaines de Kb et jusqu'à 1 Mb avec nanopore) en temps réel, rendent ces technologies extrêmement intéressantes et elles permettent d'explorer les génomes à une résolution sans précédent (van Dijk *et al.*, 2018).

De plus, la portabilité du système Nanopore (Figure 22) a permis pour la première fois un séquençage sur le terrain, point important, pour l'étude des épidémies dans certains pays (Quick *et al.*, 2016; Batovska *et al.*, 2017).



Figure 22 Dispositif MinION d'Oxford Nanopore (<https://research.a-star.edu.sg/articles/highlights/expanding-the-power-of-pores/>)

Malgré l'émergence des technologies de 3^{ème} génération pour les études de routine et de surveillance, ces technologies de 2^{ème} et 3^{ème} génération devraient être considérées comme complémentaires plutôt que successives. Par exemple, si la taille réduite des reads produits par les séquenceurs de 2^{ème} génération constitue une limite dans l'assemblage des régions hautement répétitives, les reads produits par les technologies de 3^{ème} génération sont suffisamment longs pour contourner ce problème. Ils permettent ainsi de séquencer des gènes bactériens difficiles, avec plusieurs régions répétées et un taux d'erreur de séquençage élevé (Sims *et al.*, 2014; Ronholm *et al.*, 2016). Une option possible serait de générer dans un premier temps des séquences longues en utilisant une technologie de 3^{ème} génération, puis augmenter la couverture en utilisant une technique de 2^{ème} génération pour améliorer la précision de l'analyse génomique (Miclotte *et al.*, 2016). Cette approche d'un séquençage *hybride* nécessite un travail supplémentaire de préparation de deux librairies, un coût non négligeable et le traitement d'une quantité importante de données. Un résumé des différentes plates-formes HTS et leurs caractéristiques est présenté en Annexe I.

6.2 METAGENOMIQUE ET METABARCODING

L'approche **métagénomique** est une méthode globale (shotgun) utilisée pour séquencer de préférence des génomes entiers. Cette approche ne demande pas de connaissance *à priori* des micro-organismes ou autres cibles et donc, par exemple, des virus présents dans l'échantillon, d'où le terme de séquençage *agnostique*. Pour l'analyse du virome, les protocoles utilisés comprennent des étapes de purification de l'échantillon pour éliminer les acides nucléiques des bactéries, de l'hôte ou encore de la matrice alimentaire. Un avantage est que tous les virus (au-dessus d'un seuil de sensibilité) peuvent être représentés dans les séquences obtenues, fournissant ainsi des données importantes sur le virome et la co-infection, particulièrement importante dans les échantillons cliniques. Toutefois, les virus

présents en concentration trop faible peuvent ne pas être détectés. La sensibilité de cette approche de métagénomique dépend du niveau de couverture et du nombre d'échantillons séquencés dans un même run (*multiplexing*) (Cotten and Koopmans, 2016).

L'approche ***métabarcoding*** est une méthode ciblée qui permet de séquencer seulement des parties du génome d'intérêt à travers l'utilisation d'amorces spécifiques et des étapes d'amplification. Cette approche peut être utilisée en ciblant spécifiquement des régions permettant l'identification de la souche (par exemple la région codant la polymérase ou la région de la capsidie nécessaire pour classer les NoVs). Une seconde approche est d'utiliser des amplicons se chevauchant et, lorsqu'ils sont regroupés et séquencés, fournissent une couverture de reads courts mais permettant de couvrir l'intégralité du génome viral. L'avantage principal réside dans le fait que la méthode concentre les ressources de séquençage sur le virus cible, qui peut être détecté même pour de faibles concentrations grâce à l'étape d'amplification. L'inconvénient est que la méthode ne révèle que les virus connus ou relativement proches, alors que les autres souches, non reconnues par les amorces, restent invisibles (Cotten and Koopmans, 2016).

6.3 HTS DANS LE CONTEXTE DE LA METAGENOMIQUE VIRALE

Les technologies les plus communément utilisées pour générer des séquences de génome d'ARN viral sont soit Illumina ou IonTorrent en raison de leur performance et de leur coût relativement bas par giga base (Marine *et al.*, 2019). Le choix entre ces deux systèmes doit considérer le temps de séquençage et la quantité des reads obtenue. Le nombre élevé de reads produit par Illumina augmente les chances qu'un génome viral, même très peu abondant, soit suffisamment couvert, justifiant son utilisation pour la métagénomique d'échantillons complexes. La technologie IonTorrent, permettant un temps de séquençage plus rapide mais produisant moins de reads, semble mieux adapté pour l'analyse d'échantillons cliniques présentant de plus fortes concentrations virales (Nieuwenhuijse and Koopmans, 2017).

L'approche novatrice du séquenceur minION diminue significativement la durée du run de séquençage (environ 6 h) au détriment du taux d'erreur (entre 10 et 60%). Cette technologie semble prometteuse pour une identification du virus cible rapide lorsque les concentrations virales sont élevées ou lorsqu'une couverture faible peut suffire (Greninger *et al.*, 2015).

Pendant ma thèse nous avons utilisé la technologie Illumina (séquenceur MiSeq et son successeur NextSeq) pour l'analyse métagénomique de nos échantillons (section travaux originaux).

6.4 PREPARATION DES ECHANTILLONS

Bien que le séquençage métagénomique offre la possibilité d'analyser la totalité des génomes dans un échantillon, la majorité du matériel génétique contenu est souvent d'origine non virale, mais plutôt bactérienne et issu de la matrice (l'hôte, l'aliment...) (Nieuwenhuijse and Koopmans, 2017). Lorsque l'on cible un virus, ou une famille virale, des approches spécifiques peuvent aider à pallier ces inconvénients en augmentant sa récupération. Toutefois, l'absence d'une méthode unique pour le séquençage haut débit, complique l'analyse du virome, voire l'obtention de résultats reproductibles (Conceição-

Neto *et al.*, 2015).

La Figure 23 résume les étapes principales permettant de réduire la charge bactérienne et la présence d'acides nucléiques libres dans un échantillon. Centrifugation, filtration, traitement par le chloroforme et digestion par des nucléases, sont des étapes importantes pour enrichir l'échantillon en particules virales complètes. La plupart des méthodes publiées reposent sur la petite taille des virus et sur la stabilité de leur capsid (Lewandowska *et al.*, 2017). Le traitement par un détergent ou autre traitement chimique tel que le chloroforme, associé à un traitement enzymatique, endommage la structure de l'enveloppe et expose les acides nucléiques à l'action des nucléases favorisant l'enrichissement des virus non enveloppés (Vibin *et al.*, 2018).

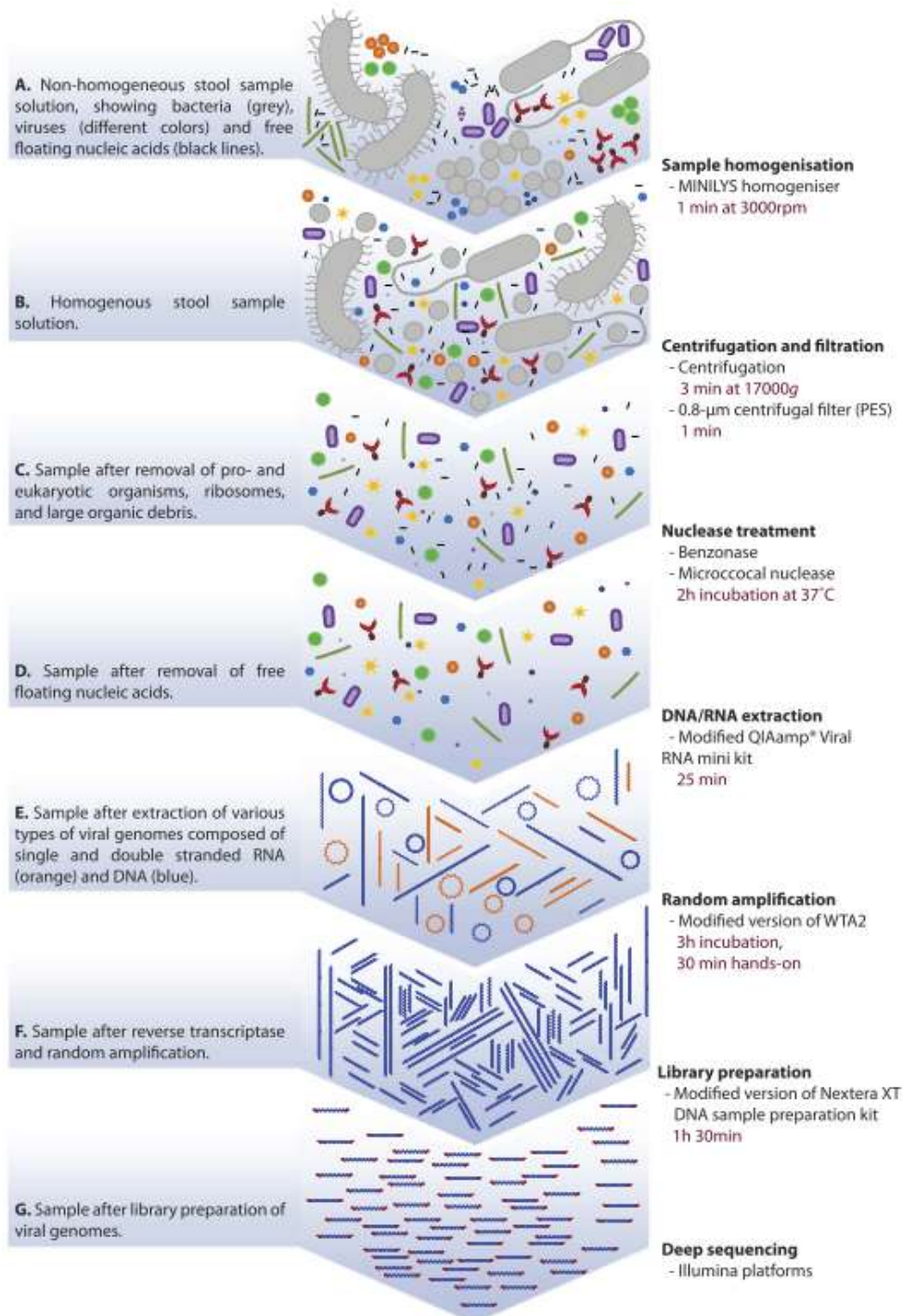


Figure 23 Schématisation du protocole pour l'analyse métagénomique des virus à partir d'échantillons de selles (Conceição-Neto *et al.*, 2015).

6.5 PREPARATION DES LIBRAIRIES

Divers protocoles et kits sont disponibles et leur nombre augmente constamment, variant non seulement dans le prix mais également dans les méthodes de sélection. Outre les approches classiques (dont les étapes sont présentées ci-dessus), d'autres approches existent combinant plusieurs de ces étapes et diminuant le temps de préparation et la perte du matériel génétique (Aigrain *et al.*, 2016). Des étapes de lavages sur billes magnétiques sont fréquemment proposées pour éliminer les débris et purifier la librairie (Tableau 6). Une étude a comparé la qualité des produits des librairies obtenues en utilisant différents kits commerciaux en évaluant le taux d'erreur après séquençage. Les kits NEB, Agilent et KAPA se sont révélés être les plus performants avec un taux d'erreur < 0.1% (Aigrain *et al.*, 2016).

Tableau 6 Description des étapes prévues pour différents kits utilisés pour la préparation des librairies Illumina (Aigrain *et al.*, 2016).

	End repair	Bead cleaning	A-tailing	Bead cleaning	Adaptor ligation	Bead cleaning	PCR & bead cleaning	Number of steps after shearing
NEBNext	x	x	x	x	x	x	x	8
NEBNext Ultra	2 in 1 ^a				x	x	x	5
SureSelect	x	x	x	x	x	x	x	8
Truseq Nano	x	x	x		x	x	x	7
Truseq DNA PCR-free	x	xx ^b	x		x			6
Accel-NGS 15 ^c	Adaptase		1st extension	x	2nd extension	x	x	7
Accel-NGS 25 ^c	4 different steps + 4 bead cleaning						x	10
KAPA Hyper ^d	2 in 1 ^a				x	x	(x)	3 (or 5)
KAPA HyperPlus ^{d,e}			x		x	x	(x)	3 (or 5)

^aBoth End-repair and A-tailing enzymes are combined in a single reaction mix

^bIllumina recommends performing an upper and lower bead clean-up selection after the end repair step

^cSwift Biosciences Accel protocols follow different chemical steps than the classical end-repair, A-tailing, adaptor ligation and PCR

^dKAPA Hyper and KAPA HyperPlus protocol don't always require a PCR amplification step

^eKAPA HyperPlus protocol starts with non-sheared DNA. The 1st step of the protocol corresponds to the enzymatic shearing of the DNA sample (fragmentase). This fragmentase step leaves blunt-ended DNA fragments which don't require End-repair and can go straight to A-tailing without any bead clean-up

a) Synthèse de l'ADNc à partir de l'ARN extrait

Des méthodes de séquençage direct des molécules d'ARN existent (Ozsolak *et al.*, 2009; Galalde *et al.*, 2018), mais les technologies les plus développées et performantes ont été optimisées pour le séquençage de l'ADN. La transcription de l'ARN en ADNc est donc nécessaire et représente une étape critique pouvant induire des biais lors de la préparation des librairies et donc avoir des répercussions sur le séquençage (Bustin *et al.*, 2015; Boone *et al.*, 2018).

Les ARNs extraits, le plus souvent cassés, sont soumis à une retro-transcription (RT) en utilisant une enzyme (transcriptase inverse) et des amorces constituées de courtes séquences aléatoires de quelques nucléotides pour la synthèse de l'ADNc. Pendant cette étape, les erreurs de transcription peuvent être assez importants (1/9000 – 1/30000), en raison de l'absence de système de correction (*proofreading*) des enzymes utilisées, qui ne corrigent pas un mauvais appariement des bases pendant la synthèse de ce nouveau brin d'ADN (Boone *et al.*, 2018). La variabilité de la fiabilité de la transcription pendant la phase de RT peut aussi dépendre de la qualité du matériel génétique de départ (des contaminants organiques peuvent affecter la qualité de l'extrait), être concentration-dépendant ou liés à l'opérateur (pipetage, manipulation de l'échantillon.) (Bustin *et al.*, 2015). Des répliques

techniques sont conseillés pour limiter l'impact de ces biais.

L'utilisation d'amorces aléatoires pendant cette étape est importante pour la métagénomique afin de permettre la rétro-transcription de tous les génomes présents dans l'échantillon. Parmi les amorces les plus souvent utilisées figurent :

- Les **hexamères random**, commercialisés par New England BioLabs (NEB) comme un mélange optimisé et prêt à l'utilisation. Ce mélange est composé d'amorces de 6 bases (d'où le nom d'hexamères) et d'oligo(dT)-*ancrées* (qui se terminent par un V, mélange de A, C et G, afin de faciliter la liaison avec la première base en amont de la queue poly(A) de l'ARNm) (<https://international.neb.com/-/media/catalog/datacards-or-manuals/s1330datasheet-lot0041208.pdf?rev=33bbe02bd0684b90983f259786047360>). Bien que l'exacte composition de ce mélange commercial ne soit pas connue, ils sont largement utilisés. Leur action aléatoire a été remise en cause par un travail suggérant l'absence d'une couverture uniforme et la forte récurrence d'un même motif sur les premières positions des reads obtenus (Hansen *et al.*, 2010).
- Les **hexamères non-ribosomiaux**, ont été sélectionnés par Endoh *et al.* (2005) à partir des hexamères random afin de limiter l'amplification de l'ARN ribosomal (ARNr) de l'hôte (rat et Homme en particulier) sans avoir un impact sur l'amplification des autres virus de mammifères. Le séquençage des molécules d'ARNr peut en effet réduire de façon important la sensibilité globale du séquençage (Boone *et al.*, 2018).

Le brin d'ADNc synthétisé sert ensuite à la synthèse du deuxième brin complémentaire grâce à l'action d'une ADN polymérase. Le fragment protéolytique (Klenow) de l'ADN polymérase I (Pol I) dérivée d'*E. coli* était la première polymérase utilisée pour le séquençage capillaire (Schirmer *et al.*, 2016). Aujourd'hui plusieurs kits commerciaux utilisent des recombinants de la polymérase mais sans précision.

b) Fragmentation

La fragmentation de l'ADN est une étape nécessaire lors de la construction des bibliothèques afin de rendre uniforme la taille des brins d'ADN et optimiser le séquençage. Les plateformes Illumina demandent des fragments de taille comprise entre 200 et 500 pb (Bronner *et al.*, 2013). Deux stratégies sont possibles :

- **Mécanique :**

Plusieurs approches sont possibles comme la sonication, les ultrasons, ou la nébulisation qui utilise l'air comprimé pour forcer le passage des molécules d'ADN à travers un pore dans une unité de nébulisation, la taille des fragments étant déterminée par la pression utilisée (Apone *et al.*, 2017). La technologie Acoustique Focalisée Adaptative (AFA) commercialisée par Covaris est une des méthodes les plus utilisées. Les ultrasons, focalisés dans un petit flacon en verre contenant l'échantillon, induisent la cavitation de l'échantillon par création et effondrement continu de microbulles. Cela crée des forces localisées qui fragmentent

l'ADN rapidement et de manière reproductible. Covaris utilise une fréquence de 15 à 30 fois supérieure à celle d'un commun sonicateur de laboratoire. Avec la sonication ou la nébulisation, la taille des fragments est souvent très variée et difficile à adapter, avec un risque d'échauffement de l'échantillon et une altération possible de l'ADN. L'approche Covaris est reconnue être la meilleure, avec une faible perte de l'échantillon, des gammes de taille facilement ajustables et une reproductibilité élevée (Bronner *et al.*, 2013 ; Boone *et al.*, 2018). Cependant ces trois méthodes induisent le même biais pendant le séquençage avec un clivage amplifié des zones du génome riches en nucléotides G ou C (Poptsova *et al.*, 2014).

- **Enzymatique :**

Cette méthode implique la digestion de l'ADN libre par des enzymes dont le choix est capital. Les fragments peuvent être obtenus par digestion simultanée des deux brins d'ADN (endonucléases de restriction) ou par génération de coupures sur chaque brin d'ADNdb (nucléases, fragmentases). Des kits contenant un mélange d'enzymes sont le plus souvent utilisés (Fragmentase, commercialisé par NEB ou Nextera tagmentation par Illumina) (Apone *et al.*, 2017; Ignatov *et al.*, 2019). Cette approche constitue une alternative attractive car aucun investissement n'est nécessaire, plusieurs échantillons peuvent être fragmentés en même temps, avec une faible perte. La taille des fragments peut être adaptée en modifiant le rapport entre concentration d'ADN et d'enzyme ainsi que le temps d'incubation (Boone *et al.*, 2018). L'inconvénient majeur réside dans le fait que les enzymes peuvent avoir des zones de clivage préférentielles ou induire des insertions ou délétions donnant lieu à une fragmentation non aléatoire (Ignatov *et al.*, 2019).

Indépendamment de la stratégie de fragmentation utilisée, la taille et la distribution des fragments doivent être vérifiées avant de passer à l'étape suivante. Cette analyse est réalisée par électrophorèse des fragments d'ADN, le plus souvent en utilisant un appareil type Bioanalyzer (Agilent) qui permet une visualisation du profil des bibliothèques en mesurant la concentration (Lee and Abdullah, 2014 ; Aigrain *et al.*, 2016).

c) Réparation, ligation, tailing

Suite à la fragmentation, une étape de réparation des extrémités des fragments (End-repair) et de phosphorylation est nécessaire pour obtenir des bouts francs et permettre la ligation des adaptateurs associés à l'index. Afin d'éviter la ligation des fragments d'ADN entre eux, un nucléotide (généralement l'Adénine) est ajouté en position 3' du fragment d'ADN, étape appelée *tailing* (Figure 24) (Bronner *et al.*, 2013). Les adaptateurs sont des séquences spécifiques qui varient selon les kits utilisés, toutefois une forme en « Y » est souvent utilisée et permet d'avoir une séquence différente sur chaque brin d'ADN réduisant la formation de produits secondaires et optimisant la fixation à la flowcell (Boone *et al.*, 2018). Chaque échantillon est associé à un index, associé à l'adaptateur, qui donne un code unique permettant d'identifier les échantillons après mélange pour le séquençage.

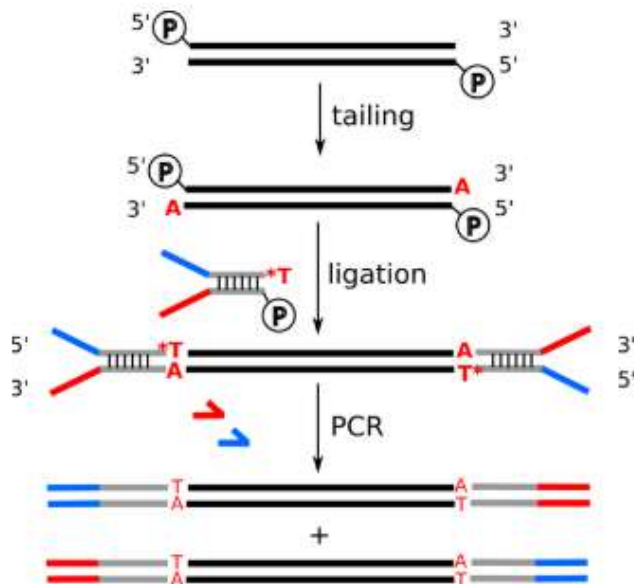


Figure 24 schéma de l'étape de tailing et ligation. Les index (gris) sont contenus dans les adaptateurs (double séquence, rouge et bleu) (Boone *et al.*, 2018)

d) Pré-amplification

Avant le séquençage, chaque fragment est amplifié afin d'augmenter le signal. L'amplification à pont (*bridge-PCR*) permet aux deux séquences contenues dans les adaptateurs de se fixer à la flow cell, support en lame de verre creuse à travers lesquelles s'écoulent les réactifs et l'ADN matrice (Bronner *et al.*, 2013). Suite à la formation d'un pont attaché, les fragments sont amplifiés formant des *clusters* (Figure 25). A chaque cycle, des étapes de synthèse et dénaturation se succèdent pour produire, à la fin de la PCR, environ 1000 copies de chaque cluster, avec plusieurs millions de clusters amplifiés simultanément (Shendure and Ji, 2008). Le profil des librairies est ensuite visualisé, la concentration mesurée avant réalisation du séquençage haut débit.

La pré-amplification, nécessaire pour les échantillons faiblement contaminés, peut introduire des biais, comme des mutations ponctuelles, souvent difficiles à différencier des SNPs (Single Nucléotide Polymorphisms ou polymorphismes d'un seul nucléotide). En outre, l'action des polymérases peut induire des erreurs, comme l'amplification préférentielle de certains fragments par rapport à d'autres, selon leur composition nucléotidique : les molécules d'ADN trop riches ou trop pauvres en GC sont plus difficiles à amplifier (donc moins représentés), alors que les fragments les plus courts peuvent être surreprésentés (Aigrain *et al.*, 2016 ; Boone *et al.*, 2018).

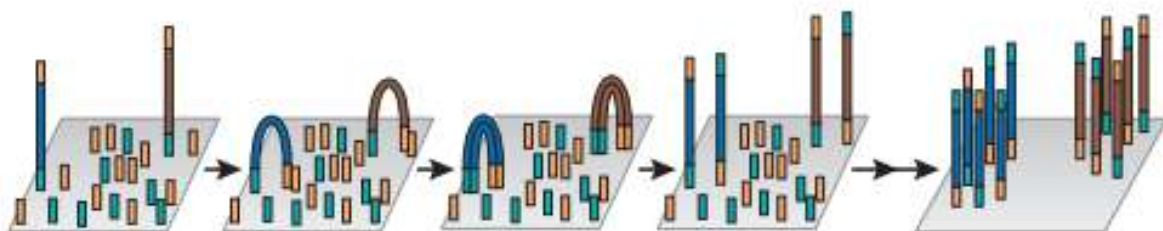


Figure 25 Amplification à pont des librairies avant séquençage per Illumina (Adapté de Shendure and Ji, 2008)

6.6 CAPTURE ET ENRECHISSIMENT

La métagénomique par son approche culture-indépendante et agnostique, est un outil intéressant pour caractériser les populations virales. Cependant, les virus présents en faible concentrations et possédant un génome court, seront plus difficile à identifier par cette méthode. Différentes approches pour enrichir la composante virale dans les échantillons ont été proposées. Par exemple, Wylie *et al.* (2015) ont sélectionné un ensemble de sondes (appelé ViroCap) reconnaissant 34 familles virales (à ADN et à ARN) susceptibles d'infecter des hôtes vertébrés. Sur ce même principe, une puce de capture (appelée VirCapSeq-VERT) ciblant 207 taxons viraux infectant des vertébrés a été décrite (Briese *et al.*, 2015). Ces deux approches sont basées sur une étape d'hybridation qui intervient à la fin de la préparation de la librairie.

Les sondes sont des séquences oligonucléotidiques biotinylées à base de solutions conçues pour hybrider et capturer des régions cibles. Leur taille, plus longue que les amorces de PCR, permet aux sondes de tolérer la présence de plusieurs mésappariements dans le site de liaison sans pour autant interférer avec l'hybridation sur la région cible. Comparé à l'approche métabarcoding, la capture par hybridation permet le séquençage de régions voisines éventuellement non reconnues par des amorces spécifiques, et cible la totalité du génome (Jennings *et al.*, 2017). De plus, ces approches présentent l'intérêt de réduire l'impact de la matrice (humain, animal, aliment...) et ainsi d'augmenter les reads viraux de 100 à 10,000 fois (Briese *et al.*, 2015 ; Wylie *et al.*, 2018). Cependant, des limites peuvent restreindre l'utilisation de cette approche. Ainsi, le temps nécessaire à la préparation des échantillons peut être considérablement allongé (l'hybridation des sondes de capture aux librairies peut demander jusqu'à 72h). Egalement des génomes très divergents ou non identifiés dans les bases de données, ne seront pas enrichi augmentant la probabilité qu'ils ne soient pas détectés. Enfin, le coût de cette technologie est élevé, justifiant un multiplexage important pour optimiser le coût de l'analyse (Gaudin and Desnues, 2018; Wylie *et al.*, 2018). A l'issue de cet enrichissement, une visualisation du profil des librairies et la mesure de la concentration doivent être vérifiées (Aigrain *et al.*, 2016).

6.7 PIPELINE BIOINFORMATIQUE POUR L'ANALYSE DES DONNEES DE SEQUENCAGE

Suite au séquençage haut débit, plusieurs centaines de milliers à plusieurs millions de reads sont produits pour chaque librairie. Le traitement et l'analyse d'une telle quantité de données pose des problèmes de calcul et nécessite des connaissances en bio-informatique afin de sélectionner les outils les plus performants et les mieux adaptés tout en réduisant les temps d'analyse.

De nombreux outils ont été développés et des programmes sont disponibles sur le net et « prêts à l'emploi » rendant l'analyse de séquences métagénomiques accessible aux scientifiques. Parmi les outils automatisés nous pouvons citer RIEMS (Scheuch, Höper and Beer, 2015), SLIM (Cotten *et al.*, 2014), Genome Detective (<https://www.genomedetective.com/app/typingtool/virus/introduction>). Ces outils demandent néanmoins une étude post-analyse pour évaluer, interpréter et analyser les résultats obtenus pour vérifier les virus identifiés dans l'échantillon (Brinkmann *et al.*, 2019). L'analyse bio-informatique prévoit une succession d'étapes (pipeline) utilisant une combinaison de logiciels informatiques avec des paramètres personnalisés à chaque étape.

Tableau 7 Estimation de la qualité des séquences obtenues

Score de qualité Phred	Probabilité d'une identification incorrecte	Précision de l'identification d'une base
10	1 pour 10	90 %
20	1 pour 100	99 %
30	1 pour 1000	99.9 %
40	1 pour 10000	99.99 %
50	1 pour 100000	99.999 %

https://fr.wikipedia.org/wiki/Score_de_qualit%C3%A9_phred

- Filtration des reads non viraux

Cette étape facultative, prévoit l'élimination des reads qui composent le bruit de fond (ou *background*) à l'issu du séquençage tels que les reads liés à la matrice, aux bactéries, mais encore tous les reads sans intérêt pour l'objectif de l'analyse. Le filtrage, possible grâce à des outils tels que BWA (Burrows-Wheeler Aligement tool) et Bowtie2, permet d'augmenter la rapidité d'analyse lors des étapes suivantes et la spécificité en diminuant le nombre de faux positifs (Brinkmann *et al.*, 2019).

De plus, la complexité du jeu de donnée dépend de l'abondance de reads obtenus et de la variabilité des génomes viraux présents. Lorsque la complexité est importante l'étape d'assemblage devient un véritable défi. Pour limiter la quantité de reads et la redondance d'information, la normalisation *in silico* est possible (**BBnorm**). Cette stratégie permet de réduire la complexité de l'échantillon en filtrant la couverture en excès par élimination des séquences contenant des informations redondantes (Rose *et al.*, 2016). Un sous-échantillonnage basé sur la décomposition des reads en motif clés (kmers) élimine de façon aléatoire un certain pourcentage de reads afin de réduire la couverture moyenne, le risque d'erreurs pendant l'assemblage et ainsi de gagner en rapidité d'analyse (<https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbnorm-guide/>).

- Assemblage

Un des défis majeurs pendant l'analyse métagénomique est l'assemblage des reads en séquences plus longues (*contigs*) ou, idéalement, en génomes complets. L'abondance de données de départ et la variabilité des séquences font de cette étape un vrai challenge computationnel avec des durées d'analyses parfois longues et la nécessité de technologies informatiques performantes.

Deux approches sont possibles:

- **L'assemblage *de novo*** : les reads sont reliés ou assemblés entre eux grâce à la présence de chevauchements des nucléotides permettant la construction des contigs. Selon l'abondance de reads, chaque position du génome est couverte par plusieurs reads, ce nombre de reads constitue la couverture. L'importance et l'homogénéité de la couverture permet de différencier les « vrai contigs » et les faux positifs.

L'assemblage des contigs entre eux permet d'obtenir un *scaffold*, en incluant dans la séquence des *gaps* (trou), induits par l'absence d'information (Figure 27). L'estimation de la taille des gaps est possible en utilisant l'approche *Paired-End*. Ce terme se réfère à la possibilité de séquencer les deux extrémités du même fragment d'ADN (2 x 150 nt ou 2 x 300 nt) permettant un alignement des reads plus précis.

- **Alignement avec référence** : les reads sont comparés à des séquences de référence, permettant leur assemblage. Cette méthode peut être sensible à un biais selon la séquence de référence choisie, et assemble en général peu de reads, diminuant la qualité des informations sur les variations génétiques obtenues lors de l'assemblage (Lin *et al.*, 2017).

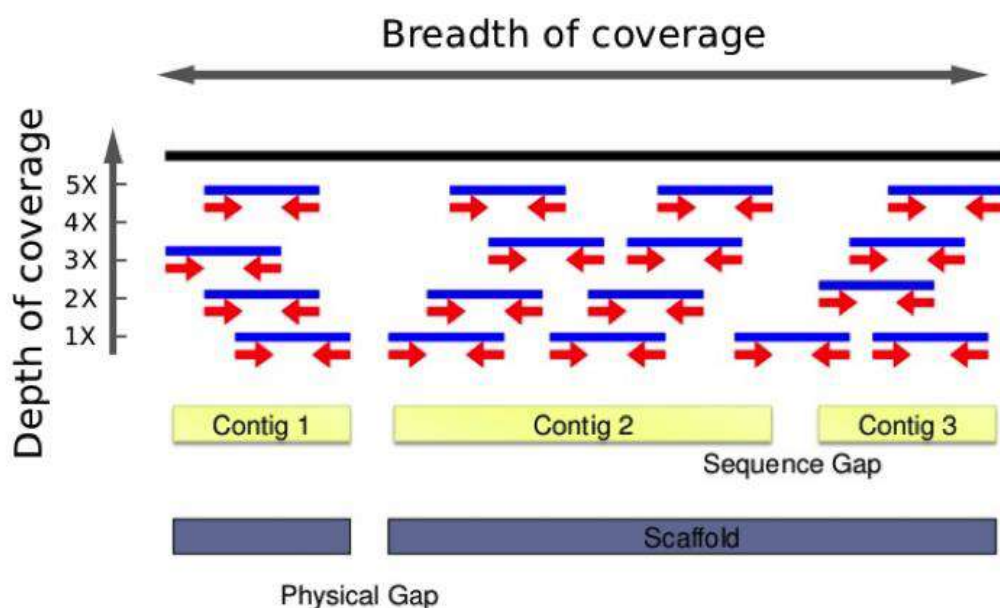


Figure 27 Schématisation de l'assemblage des reads.

Les reads paired-end sont représentées en bleu ; les deux flèches rouges montrent les deux sens de lecture. Le chevauchement des reads sur la même zone du génome donne une couverture et permet l'assemblage des contigs. Les scaffolds résultent de l'assemblage de plusieurs contigs pouvant inclure des régions avec très peu de reads. La couverture du génome est exprimé 1X, 2X.. et est calculé à partir du nombre de bases de chaque read qui mappe sur le génome divisé par la taille du génome. (Modifié de <https://github.com/Ecological-and-Evolutionary-Genomics/eeg2016/wiki/Mar-21-Exercise-7----SPAdes-assembler>)

Dans les approches de métagénomique virale, l'assemblage *de novo* est privilégié car les informations génétiques obtenues sont plus complètes et plus précises par rapport à l'approche sur référence (Nooij *et al.*, 2018). Plusieurs programmes d'assemblage existent et des algorithmes différents ont été développés mais leur comparaison en termes de sensibilité et précision est difficile (White *et al.*, 2017).

Les données de métagénomique virale représentent un challenge pour les outils d'assemblage. En effet, les régions répétées au sein des génomes viraux sont nombreuses, ainsi que les régions hypervariables associées à l'interaction avec l'hôte. De plus chez les virus à ARN, le taux de mutation élevé entraîne une augmentation de la complexité du métagénome et une variation des souches conduisant à des assemblages fragmentés limitant l'analyse (Sutton *et al.*, 2019). Afin d'aborder la complexité et la taille des jeux de données métagénomiques, les assembleurs utilisent typiquement le graphe de Bruij, utilisé entre autre par SPAdes (Bankevich *et al.*, 2012). Les reads sont découpés en k-mers de tailles

Il faut toutefois noter que ces deux approches peuvent entraîner une simplification de l'information en sortie et potentiellement limiter l'étude des variantes.

- Recherche dans la base de données

Après ces diverses étapes, l'identification des séquences virales obtenues représente le dernier défi de l'analyse métagénomique. Dans cette étape, les séquences sont associées à une base de données de référence. Plusieurs outils sont proposés pour aligner les séquences protéiques (DIAMOND) ou nucléotidiques (BWA ou Bowtie2) ou les deux (BLAST), permettant l'assignation taxonomique et donc l'identification des séquences obtenues (Nooij *et al.*, 2019). Toutes les méthodes de recherche reposent sur des bases de données de référence, telles que NCBI GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) ou RefSeq (<https://www.ncbi.nlm.nih.gov/refseq/>). Les données de ces bases conditionnent la qualité des résultats de cette dernière étape.

Avec l'avènement du séquençage haut débit, la rapidité à laquelle des nouveaux virus sont découverts et ajoutés à la base de données augmente rapidement, passant d'une moyenne annuelle de 0,34 espèce/jour en 2010 à 2,5 espèces/jour en 2015 (Rose *et al.*, 2016). Ce taux de croissance souligne d'un côté l'importance de mises à jour récurrentes, et de l'autre un problème de qualité des séquences parfois incomplètes ou erronées soumises à ces bases. De plus, l'homologie entre plusieurs génomes de référence peut conduire à une mauvaise assignation des séquences (Nieuwenhuijse and Koopmans, 2017).

Un autre problème réside dans l'existence de bases des données incomplètes. Entre 40 et 90% des séquences obtenues ne peuvent pas être alignés sur les séquences de références existantes et ne sont donc pas identifiables, alimentant ce qui est communément appelée la *matière noire* (black matter) (Krishnamurthy and Wang, 2017; Santiago-Rodriguez and Hollister, 2019).

- Analyse post-traitement

Cette dernière étape inclue la recherche de l'identification taxonomique la plus probable et l'interprétation des données obtenues. La phylogénie vient souvent en support ainsi que des outils de représentation graphique des données de sortie de BLAST, comme MEGAN (Huson *et al.*, 2016). Une connaissance approfondie du virus cible est importante pour pouvoir différencier des séquences faussement positives ou pour relier des échantillons (par exemple en cas de suivi d'épidémies) (Nooij *et al.*, 2018). Dans le cas des souches de NoV, l'outil NoV Genotyping Tool est utilisé pour étudier la variabilité des souches basée sur une analyse phylogénétique des séquences nucléotidiques. Des séquences longues sont nécessaires pour pouvoir attribuer un génotype selon le typage des zones ORF1 et ORF2. Cet outil utilise un ensemble de séquences de référence régulièrement mis à jour rendant possible l'identification des nouvelles souches et la comparaison des séquences trouvées dans différents pays (Kroneman *et al.*, 2011).

6.8 BIAS ET ERREURS DE SEQUENCAGE

Bien que les taux d'erreurs produits pendant le séquençage avec Illumina soient faibles (0.4%) comparés à ceux obtenus avec IonTorrent (1.78%) ou avec les séquenceurs de troisième génération (>13%), des erreurs persistent et leur connaissance peut aider dans l'interprétation des données (Quail *et al.*, 2012). Les erreurs les plus fréquents sont ici résumées :

- La présence des **nucléotides G et C** empêche une bonne pré-amplification qui se traduit par des erreurs de séquençage, avec un impact sur la couverture (Figure 29). Les molécules d'ADN trop riches ou trop pauvres en GC sont généralement moins représentées car plus difficiles à amplifier (Aigrain *et al.*, 2016 ; Boone *et al.*, 2018).

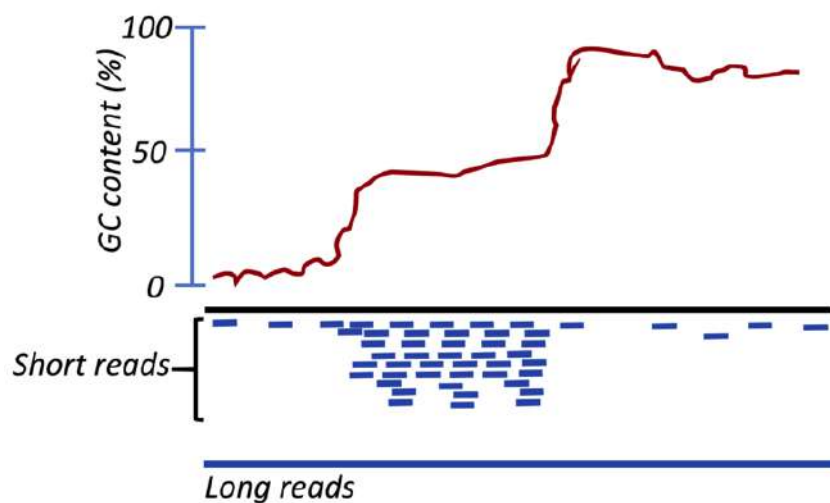


Figure 29 Impact de la présence des séquences trop riches ou trop pauvres en nucléotides G et C sur le séquençage haut débit (Adapté de van Dijk *et al.*, 2018)

- Présence d'erreurs suite à des séquences d'**homopolymères** > 20 nt. La succession du même nucléotide pendant plusieurs cycles successifs peut induire en erreur lors de la capture d'image. Toutefois ce taux d'erreur est relativement bas comparé à la technologie IonTorrent pour laquelle des erreurs sont observés avec des homopolymères de plus 8 nt et de séquences de plus de 14 nt résultent en absence de reads (Quail *et al.*, 2012).
- **Chimères**: contigs générés à partir de reads de différentes origines assemblés entre eux en raison de similarités génétiques (Nooij *et al.*, 2018).
- **Substitutions** : erreurs les plus communes en lien avec la technologie Illumina, observées le plus souvent dans les 10 premières paires de base et dans les 50 dernières. Cette erreur est due non pas à une erreur d'incorporation des bases mais à une discrimination insuffisante des fluorochromes pendant les cycles de lecture (Schirmer *et al.*, 2015). Par exemple lors de l'utilisation d'appareils Illumina MiSeq, l'incorporation préférentielle des nucléotides G et T a été mise en évidence (Schirmer *et al.*, 2016).
- **Insertion / délétion (Indel)** : Ces erreurs comprennent l'insertion ou la délétion de nucléotides sur des portions du génome, et insèrent la distribution de nucléotides

inconnus (Ns). Les erreurs d'*indel* sont environ 100 fois moins fréquents que les erreurs de substitutions (Schirmer *et al*, 2015).

- **Motifs** : séquence de 3 nucléotides qui forment un codon correspondant à un acide aminé. Des taux d'erreurs allant jusqu'à 16.2% sont associés aux motifs « GGG » et « CGG » (Schirmer *et al*, 2016).
- **Contamination avec des séquences de PhiX** : PhiX est un bactériophage connu pour avoir été le premier génome d'ADN séquencé par Sanger (1977). En raison de sa petite taille et de son génome court (environ 5300 nt) bien caractérisé, PhiX est couramment utilisé comme contrôle de séquençage dans les plateformes Illumina. L'addition de ce phage PhiX, à des concentrations variant entre 1% et 40% selon la diversité génomique des échantillons pendant la préparation des bibliothèques ou au moment du séquençage, permet d'augmenter la diversité génomique pendant le run. Cet effet est surtout notable pendant les dix premiers cycles quand une diversité de séquence est nécessaire pour obtenir des données de bonne qualité. Les séquences PhiX doivent être éliminées dans les premières étapes de l'analyse de bioinformatique afin qu'elles ne soient pas intégrées aux génomes identifiés. Mukherjee *et al.* (2015) ont constaté qu'environ 10% des séquences publiées et entrées dans les bases de données, sont contaminées par des séquences de PhiX.

6.9 CONTROLES ET RISQUES DE CONTAMINATIONS

Le séquençage haut débit est une technique sensible permettant des avancées importantes dans la découverte de nouveaux virus et dans l'étude du virome. Cependant cette sensibilité favorise cette technique à capter éventuellement les contaminants modifiant la composition de l'échantillon en amont (pendant l'échantillonnage ou les premières manipulations) ou en aval, lors du traitement de l'échantillon pour l'extraction des acides nucléiques ou encore pendant le séquençage (Sangiovanni *et al.*, 2019).

Les contaminations constituent un défi pour l'interprétation des données métagénomiques car elles peuvent passer inaperçues pendant l'assemblage *de novo* en raison d'une faible couverture ou d'un traitement non approprié et conduire à des résultats faux (Laurance *et al.*, 2014).

Pour limiter la contamination, au laboratoire les pièces dans lesquelles les échantillons sont préparés sont souvent séparés de celles dans lesquels les acides nucléiques sont amplifiés et celles où les bibliothèques sont réalisées. Les équipements sont traités aux UV et nettoyés à l'eau de javel, l'utilisation de cônes à filtre et de matériel à usage unique est indispensable (Gruber, 2015). De plus, alterner les codes-barres utilisés pour le multiplexage des échantillons lors des différents runs de séquençage réduit la contamination induite par les essais précédents (Nieuwenhuijse and Koopmans, 2017). Les contaminations peuvent aussi provenir des réactifs utilisés en laboratoire (Gruber, 2015 ; Laurance *et al.*, 2014).

Il est recommandé d'inclure plusieurs contrôles dans un même run de séquençage :

- contrôle négatif : échantillons sélectionnés sur leur faible probabilité de contenir des virus (Nieuwenhuijse and Koopmans, 2017) ;
- contrôle positif : échantillon contenant des quantités connues d'une variété de virus (Lusk *et al.*, 2014) ;

- contrôle technique : réplicats de librairie à partir du même échantillon qui sera de ce fait séquencé plusieurs fois dans le même run. Cette approche semble pertinente pour compenser les biais du pipetage, et de la RT (Cotten *et al.*, 2014 ; Bustin *et al.*, 2015) ;
- réplicats biologiques : échantillon prélevé plusieurs fois et analysé selon le même protocole (Kim *et al.* 2017).

Des outils bioinformatiques tels que DecontaMiner (Sangiovanni *et al.*, 2019) ont été développés pour vérifier les signaux de contamination régulièrement détectés en laboratoire dans les données de séquençage.

6.10 NOVs : UN CHALLENGE METAGENOMIQUE

Le séquençage haut débit présente un intérêt indéniable pour l'étude des NoVs permettant d'établir les relations phylogénétiques entre les souches, identifier les nouveaux recombinants et identifier les souches présentes dans les échantillons de l'environnement ou des aliments. En intégrant l'approche environnementale, la relation entre diversité virale, infection et transmission est une question scientifique importante. A cet égard, la capacité de suivre les épidémies et d'en identifier les sources d'introduction représente une importante application du séquençage haut débit dans les domaines cliniques et de santé publique (Cotten and Koopmans, 2016). Un autre potentiel intéressant réside dans les recherches sur le microbiome intestinal et le virome. Le séquençage agnostique des échantillons fécaux, associé à un échantillonnage stratégique au niveau de la population, révèle la diversité des virus en circulation dans la population échantillonnée. Cette approche appliquée à des échantillons de l'environnement, a permis de nouveaux développements comme par exemple l'utilisation des eaux usées comme outil de surveillance (Nieuwenhuijse and Koopmans, 2017). Cependant, le choix de la méthode et de la préparation de l'échantillon dépend de la matrice à analyser et du niveau de concentration en virus dans l'échantillon. En effet les échantillons cliniques sont caractérisés par des concentrations virales élevées, avec des valeurs de Ct variant entre 10-20, alors que les échantillons environnementaux et les matrices alimentaires ont des concentrations faibles (valeurs de Ct variant entre 30 et 40), proches de la limite de détection par PCR. Il est donc important de préparer l'échantillon pour augmenter la concentration en virus ou du moins obtenir une profondeur suffisante pour permettre une interprétation correcte des résultats (Cotten *et al.*, 2014; Nasheri *et al.*, 2017).

Le séquençage ciblé en utilisant des amorces spécifiques est une approche efficace pour typer la protéine VP1 de la capsid virale, en particulier le sous-domaine P2 sur lequel repose l'identification des souches de NoVs, (Imamura *et al.*, 2016; Oshiki *et al.*, 2018; Sabrià *et al.*, 2018). Mais il est important également de considérer la région codant la polymérase située à l'extrémité de l'ORF1 et le début de l'ORF2. En effet seul l'obtention de la séquence de ces deux régions permet d'identifier précisément une souche de NoVs. Cotten et collègues (2014) ont proposé une stratégie pour séquencer des génomes entiers de NoVs en utilisant trois paires d'amorces sélectionnées sur la base de toutes les séquences disponibles publiquement. Cette approche, effective pour l'identification de souches du GII et en particulier des souches GII.4, a montré des limites pour les souches GI, moins représentées

dans les bases de données.

L'approche métagénomique reste l'approche de préférence pour l'étude d'échantillons environnementaux où la possibilité d'identifier d'autres virus entérique humains éventuellement présents, présente un intérêt certain. Cependant la préparation de l'échantillon et la mise au point de stratégies d'enrichissement sont des étapes nécessaires pour limiter les séquences de la matrice et des bactéries. Récemment l'analyse métagénomique appliquée à un échantillon de fraises impliquées dans une épidémie de gastroentérite, a conduit à la production de 28,856,294 reads, avec 11 reads identifiables comme des virus de mammifère, et deux reads de NoVs (Bartsch *et al.*, 2018). Dans ce cas précis, l'analyse épidémiologique et le virus identifié chez les consommateurs malades ont permis de valider ces deux reads, mais en l'absence de ces informations, toute conclusion aurait été impossible.

La puce SureSelect développée par Brown et collègues (2016) permet un enrichissement ciblé des NoVs grâce aux amorces multiples ciblant chaque position du génome. Cette approche permet un séquençage sans biais entre les génotypes connus et a été efficace en générant des génomes complets à partir d'échantillons faiblement contaminés (Ct \geq 36).

L'inconvénient de cette méthode réside dans l'échec éventuel de générer des séquences pour un génotype émergent lorsque les amorces existantes sont trop divergentes.

Une autre stratégie de capture pour générer des génomes complets de NoVs mais sans utilisation d'amorces a été proposée afin d'enrichir spécifiquement en ARN de NoVs et de réduire la quantité d'ARN bactérien non polyadénylé. Un procédé de capture basé sur la reconnaissance de la queue poly(A) a été utilisé après l'extraction des acides nucléiques permettant l'augmentation de reads de NoVs à partir d'échantillons de selles et l'obtention de génomes complets (Fonager *et al.*, 2017). Toutefois la polyadélation, jouant un rôle de stabilisation des ARNm, est commune et la grande majorité des ARNm sont polyadénylés, ce qui limite un enrichissement ciblé des NoVs.

L'application et l'optimisation des méthodes d'enrichissement ciblées à NoVs sans être trop restrictif ont été objet de mes recherches et seront expliqués plus en détail dans la section « travaux originaux » de ce manuscrit.

7. APPROCHES COMPLEMENTAIRES

La RT-PCR reste la méthode la plus courante de détection et de quantification des NoVs mais, reposant sur l'amplification d'un fragment court du génome viral, elle ne donne pas d'information sur l'intégrité du virion. En effet, les résultats de l'amplification proviennent de génomes de particules infectieuses, de particules complètes non infectieuses, mais également des particules virales endommagées. Ces inconvénients justifient l'intérêt de développer des approches capables de détecter les particules virales intègres, donc potentiellement infectieuses. Différentes techniques de détection existent et leur utilisation a été évaluée lors de ces travaux de thèse. Certaines ont été utilisées pour répondre aux questions scientifiques de ce manuscrit, comme les mucines gastriques porcines (Chapitre 1) et le virus tulane (Chapitre 2). D'autres approches novatrices ou considérées d'intérêt scientifique sont brièvement discutées ici.

7.1 PHAGE DISPLAY

Cette technique utilise les interactions protéine-protéine en se servant de phages affichant des polypeptides capables de reconnaître les VLP de NoVs et qui contiennent également le génome codant. Cela permet une interaction rapide entre les polypeptides des phages et la cible (*biopanning*). Seuls les phages spécifiquement liés sont élués de la cible par action du pH acide et incubés avec *E. coli* pour être amplifiés. Plusieurs cycles successifs de *biopanning* sont effectués en utilisant les populations résultantes afin d'enrichir la population de phages avec les peptides les plus adaptés à lier les NoVs (Sidhu *et al.*, 2003).

Cette méthode a été utilisée pour détecter le NoV dans des échantillons de selles cliniques en utilisant la technique ELISA afin de confirmer la spécificité de liaison (Rogers *et al.*, 2013). La sélection de peptide est souvent limitée basée sur quelques souches de NoVs et donc sont difficilement applicables en test de routine.

Une étude a optimisé un système d'immunochromatographie sur membrane basé sur des particules de phages couplées à des anticorps anti-NoVs GI.1 marqués et en utilisant une approche immunoenzymatique en sandwich ELISA (Figure 30). Cette approche a permis d'acquérir une sensibilité analytique pouvant atteindre une limite de détection de 10^7 VLP/mL (Hagström *et al.*, 2015).

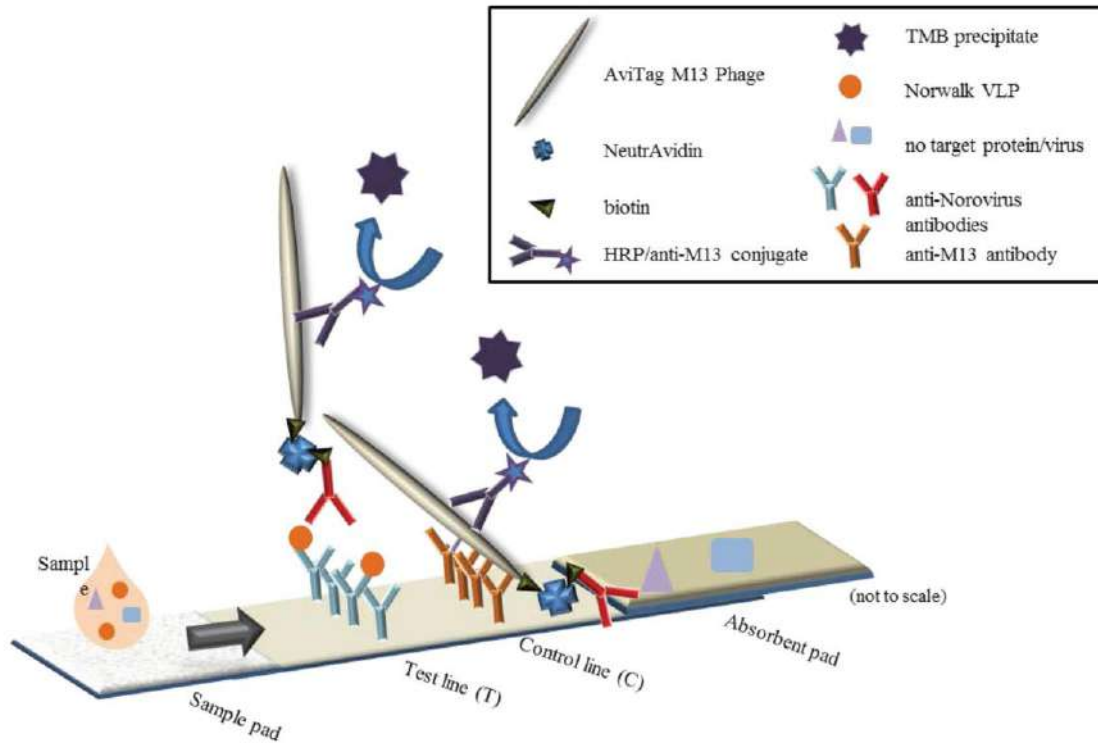


Figure 30 Détection des VLP de NoVs GI.1 à l'aide de nanoparticules de phage marquées dans un essai à flux latéral (Hagström et al., 2015)

7.2 PMA

Le PMA (propidium monoazide) est un colorant intercalant de l'ADN/ARN qui entraîne une fluorescence marquée après liaison aux acides nucléiques et exposition à une lumière visible intense (Figure 31). Lors de la photolyse, le colorant réagit de manière covalente avec l'ADN/ARN, entraînant ainsi une modification permanente du génome, ce qui empêche une amplification par PCR. De plus, le colorant est imperméable à la membrane cellulaire et peut donc être utilisé de manière sélective pour lier uniquement le génome libre dans l'échantillon (non protégé par une capsid) permettant une amplification ciblée de cellules pathogènes viables lors de test de quantification en temps réel (Karim *et al.*, 2015).

L'utilisation de ce colorant pour différencier les particules potentiellement infectieuses de NoVs et rotavirus a donné des résultats controversés lors d'utilisation sur des échantillons complexes, comme du *chorizo*, des palourdes (Quijada *et al.*, 2016), des huîtres ou encore des eaux usées (Randazzo *et al.*, 2018). Les amorces utilisées pendant la PCR, la région du génome ciblée, ainsi que la méthode d'inactivation virale utilisée (conduisant à différents dommages de la capsid) pourront influencer les résultats (Karim *et al.*, 2015).

Le PMAxx, un colorant structurellement similaire au PMA mais avec des propriétés chimiques différentes, a été comparé au PMA, avec des résultats légèrement meilleurs mais avec un certain nombre de faux positifs (Oristo *et al.*, 2018).

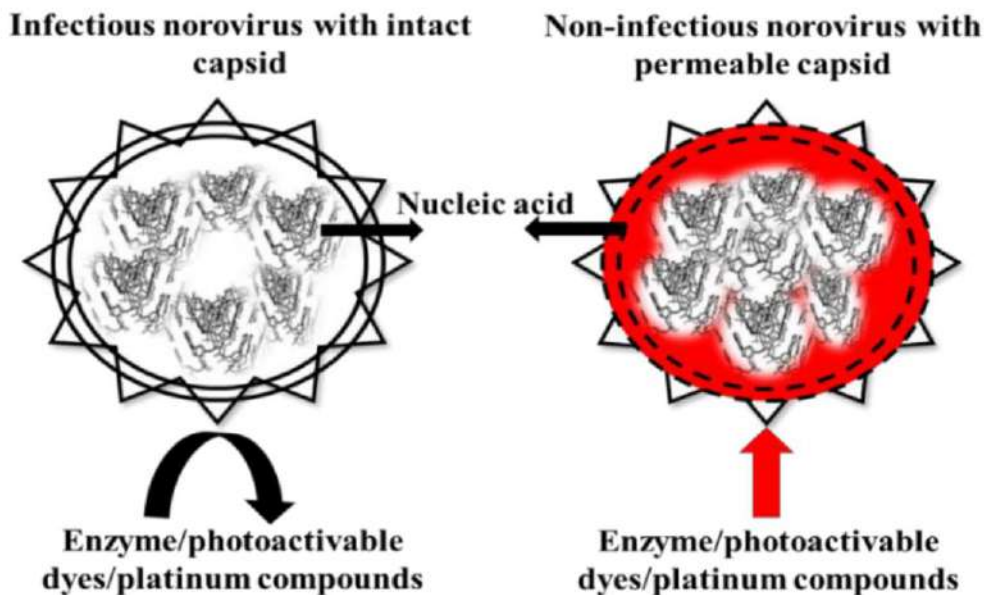


Figure 31 approche de qRT-PCR en utilisant le PMA pour différencier les NoVs infectieux et non infectieux selon la perméabilité de la capsid virale (Adapté Gyawali et al., 2019)

7.3 LES APTAMERES

Les aptamères sont de courtes séquences d'oligonucléotides avec des structures tridimensionnelles dotées d'une grande capacité de liaison résultant de la formation de différentes liaisons chimiques non covalentes (Figure 32) (Beier *et al.*, 2014). Des aptamères spécifiques sont sélectionnés *in vitro* en fonction de leur affinité pour une molécule cible à l'aide d'une méthode d'enrichissement moléculaire exponentiel appelée SELEX (de l'anglais *Systematic Evolution of Ligands by Exponential Enrichment*). Par exemple cette méthode a été utilisée pour sélectionner des aptamères dirigés contre la protéine structurale de la capsid (VP1) de la souche GII.4 (Beier *et al.*, 2014), ou contre le domaine plus extérieur de la protéine virale (le domaine P) (Moore *et al.*, 2015). En outre, un panel de souches NoVs GI et GII ont été détectés dans des échantillons de selles et d'aliments en utilisant un aptamère M6-2 (Escudero-Abarca *et al.*, 2014) qui a aussi été utilisé pour étudier le statut d'infectivité des particules virales en utilisant comme modèle la souche GII.4 Sydney (Moore *et al.*, 2016; Gyawali *et al.*, 2019). Certains aptamers présentent une réactivité assez large permettant de détecter plusieurs souches de GII.4 suggérant une utilisation en diagnostic clinique (Moore *et al.*, 2015). Giamberardino *et al.* (2013) ont produit un aptamère contre le virus murin (MNV), substitut cultivable des NoV, présentant une réactivité croisée avec les VLP GII.3 NoVs.

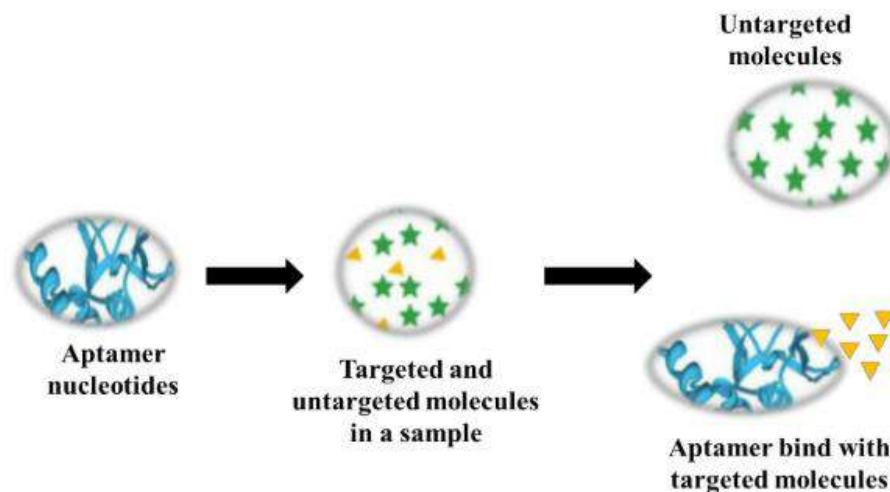


Figure 32 Diagrammes de flux montrant une approche de détection de norovirus basé sur l'utilisation d'aptamères (Adapté de Gyawali et al., 2019)

7.4 TECHNOLOGIE MICROFLUIDIQUE

La technologie microfluidique est la miniaturisation des tests moléculaires. Pendant l'essai, l'échantillon traverse différentes régions du dispositif microfluidique, soit par capillarité, soit par pression à l'aide de micro-pompes. Plusieurs réactions de capture, d'isolement et d'identification du virus peuvent se produire dans différentes sections de dispositifs microfluidiques à l'intérieur d'un système fermé (Figure 33) assurant un contrôle automatisé de toutes les étapes et pouvant réduire les erreurs humaines en augmentant précision, reproductibilité et fiabilité des résultats (Gyawali et al., 2019).

Actuellement, un type de technologie microfluidique appelée « systèmes d'analyse sur papier » a été utilisé pour détecter des faibles quantités de NoVs à partir d'échantillons d'eau en utilisant un smartphone (Figure 34). Le système contient une puce de papier contenant des billes de polystyrène fluorescent liées à des anticorps anti-NoVs. Lorsque des particules virales sont présentes dans l'échantillon, elles se lient aux billes formant des agglomérats fluorescents. Un microscope installé sur la camera du smartphone permet ensuite d'amplifier le signal lumineux et une application analyse le nombre de pixels fluorescents et le convertit en nombre de particules de NoV. La technique est suffisamment sensible pour détecter des concentrations très faibles, de l'ordre de 10 attogrammes (10^{-18} g) de NoVs par mL (Chung et al., 2019).

L'utilisation de la RT-qPCR microfluidique est un autre exemple d'application de cette technologie pour la détection des NoVs dans des échantillons environnementaux et alimentaires. Cette approche a déjà été utilisée pour la quantification simultanée de onze virus humains majeurs, notamment les entérovirus, les virus d'Aichi, les adénovirus, les astrovirus, les sapovirus, les rotavirus, les NoVs, les virus de l'hépatite A et de l'hépatite E à partir d'échantillons d'eau prélevés dans l'environnement (Ishii et al., 2014; Coudray-Meunier et al., 2016).

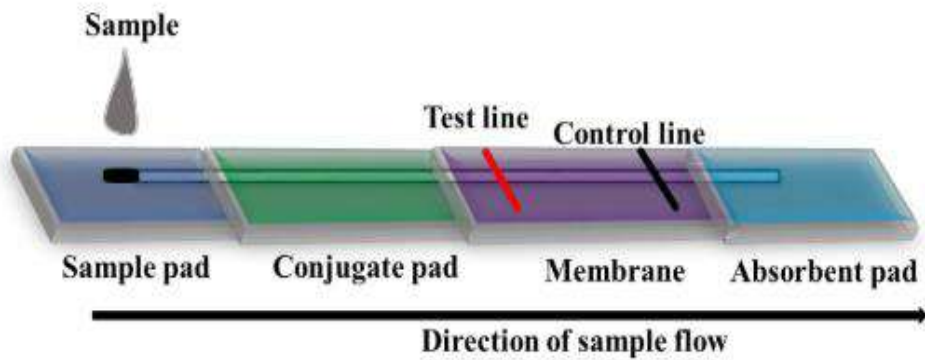


Figure 33 Diagrammes de flux montrant une approche de détection de norovirus en utilisant la technologie microfluidique (Adapté de Gyawali *et al.*, 2019).

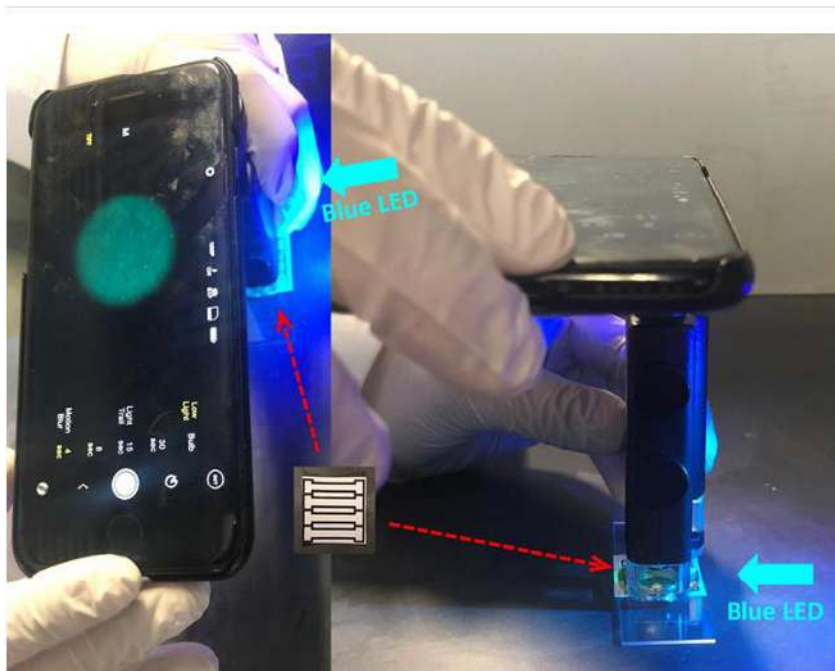


Figure 34 Smartphone et dispositif microfluidique sur papier pour la détection des NoVs (<https://www.medgadget.com/2019/08/smartphone-and-paper-microfluidic-device-for-norovirus-detection.html>)

7.5 LES MUCINES GASTRIQUES PORCINES

Les mucines sont des glycoprotéines produites par des cellules sécrétoires spécialisées présentes dans tout le tractus gastro-intestinal et forment les principaux constituants macromoléculaires de la salive et du mucus. La salive, présente dans la bouche et l'œsophage, lubrifie ces tissus et constitue la première barrière contre les agents pathogènes entériques ingérés (McGuckin *et al.*, 2011; Schomig *et al.*, 2016).

Les mucines sont composées d'environ 80% de glucides, principalement de la N-acétylgalactosamine, de la N-acétylglucosamine, du fucose, du galactose et de l'acide sialique (acide N-acétylneuraminique) et de traces de mannose et de sulfate. Parmi les sucres présents à la surface des mucines, on retrouve les HBGA. Sur la base de la capacité des NoVs à se fixer aux HBGA, Canon et Vinje (2008) ont utilisés des HBGA synthétiques liés à des billes magnétiques afin de récupérer les NoVs dans des échantillons de selle.

Cependant, la production d'HBGA synthétique reste limitée considérant les coûts de production élevés (Moore *et al.*, 2016). En raison des similitudes génétiques existantes entre les humains et les porcs, les mucines gastriques porcines (PGM) ont été utilisées pour leur capacité à lier les NoVs. Les résultats ont montré que les PGM contiennent certains HBGA, tels que les antigènes A, H1 et Lewis b (Tian *et al.*, 2005; Tian *et al.*, 2007). Ces mucines, conjuguées à des billes magnétiques (PGM-MB), ont été donc utilisées pour capturer les particules de NoVs à partir d'échantillons alimentaires ou environnementaux, suggérant ainsi une capsid intacte. Cette approche semble pouvoir différencier les NoVs potentiellement infectieux des non-infectieux et minimiser la présence d'inhibiteurs dans l'échantillon (Cannon and Vinjé, 2008; Tian *et al.*, 2012). Toutefois, les particules de capsid partiellement endommagées mais avec le domaine P2 intacte, peuvent être capables de se lier aux PGM-MB (Afolayan *et al.*, 2016). De plus la répétabilité de la méthode dépend de la présence de composants additionnels capables d'interférer pendant l'analyse, par une qualité et/ou une proportion d'HBGA variable en fonction de l'animal à partir duquel ces mucines ont été extraites, et avoir une réactivité souche-dépendante (Moore *et al.*, 2016).

7.6 VIRUS TULANE COMME MODELE DE NoVs

Après la mise en évidence du virus Tulane (TV) à partir des selles de macaque rhésus, l'analyse phylogénétique du génome a montré que ce virus constituait un nouveau genre, appelé *Recovirus* au sein de la famille des *Caliciviridae*. Par ailleurs ce virus Tulane et les autres souches caractérisées par la suite présentent de nombreuses similitudes avec les NoVs (Farkas *et al.*, 2008; Farkas, 2015). Biologiquement le TV se comporte comme les NoVs Humains, notamment dans la transmission par voie féco-orale, dans l'interaction avec les HBGA et les signes cliniques chez les singes (Farkas *et al.*, 2015). L'intérêt de sa multiplication relativement aisée sur des cellules de rein de singe rhésus (LLC-MK2) a renforcé l'utilisation de ce virus comme modèle des NoVs, par rapport aux autres virus modèles tels que le calicivirus félin, les NoVs murins ou encore les virus Aichi (Cromeans *et al.*, 2014). La pertinence du modèle TV pour étudier le comportement des NoVs lors de la contamination des huîtres a été démontrée au laboratoire (Drouaz *et al.*, 2015). Ainsi après bioaccumulation dans des huîtres, la répartition tissulaire du TV était comparable à celle du virus Norwalk, prototype du NoV GI, avec la majorité des particules virales de TV et du virus de Norwalk détectées dans les tissus digestifs. Un virus contrôle, le mengovirus appartenant à la famille des *Picornaviridae*, était détecté de façon homogène dans les branchies, le manteau, et les tissus digestifs. Le suivi de la persistance de ces virus par détection du génome a montré cependant des différences après plusieurs semaines. Les virus de Norwalk et de Tulane ont présenté des demi-vies assez similaires, alors que le mengovirus a disparu plus rapidement. Une autre étude a permis de démontrer la persistance de particules infectieuses pendant trois semaines dans des huîtres en bassin (Figure 35) (Polo *et al.*, 2018). Ces observations corroborent des observations réalisées par l'équipe lors d'épidémies liées à la consommation d'huîtres pour lesquelles l'évènement contaminant a pu être estimé (Polo *et al.*, 2016; Le Mennec *et al.*, 2017).

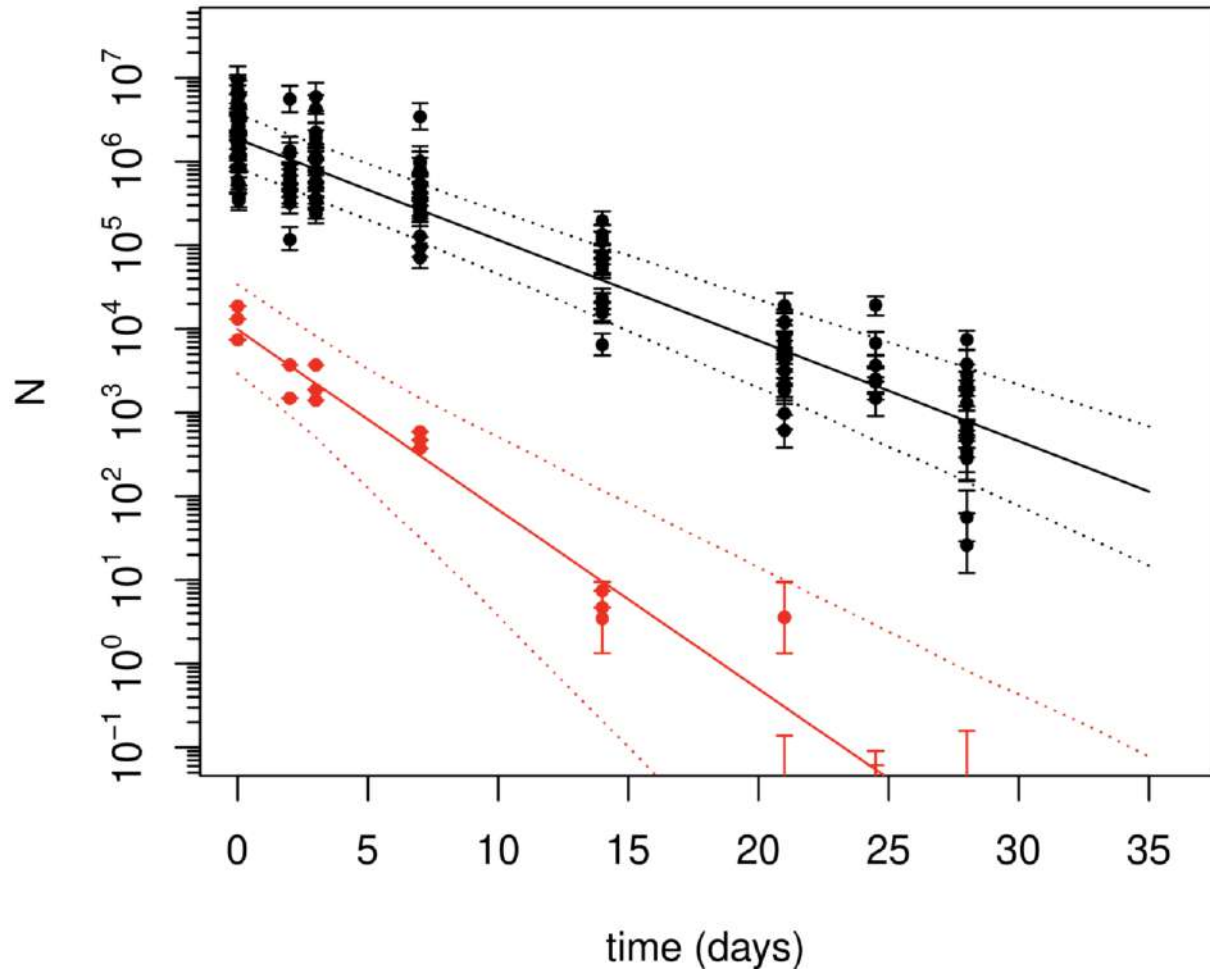


Figure 35 Persistance du virus Tulane (TV) dans les huîtres, déterminée par détection génomique (lignes noires) et test d'infectivité (lignes rouges). Les lignes en trait plein représentent la courbe de décroissance moyenne et les lignes pointillées les intervalles de prédiction à 95%, tels que produits par l'analyse de régression inverse. Axe X: temps (en jours), axe Y: nombre de copies d'ARN ou de particules TCID50 / g de DT. (Polo et al., 2018)

8. REPLICATION DES NOVs SUR CULTURE CELLULAIRE

Pendant longtemps l'impossibilité de cultiver les NoVs en laboratoire a représenté la limite majeure pour une complète compréhension des mécanismes biologiques et pathogéniques de ce virus. Finalement en 2016, environ 50 ans après la découverte des NoVs, un système de réplication virale sur cellules épithéliales intestinale humaine a été développé (Ettayebi *et al.*, 2016). Ces avancées permettront de mieux comprendre les interactions existantes entre NoVs et l'hôte et les mécanismes d'infectiosité virale. Des études pour évaluer la persistance du caractère infectieux des NoVs dans l'environnement pourront également être réalisées avec d'importantes répercussions sur la compréhension de l'épidémiologie de ce virus.

8.1 PHYSIOLOGIE ET ANATOMIE DE L'INTESTIN GRÊLE CHEZ L'HOMME

L'intestin est l'un des organes les plus complexes du point de vue architectural et fonctionnel. D'une longueur de 5 à 7 mètres et d'un diamètre de 3 cm, l'intestin grêle fait suite à l'estomac et est prolongé par le gros intestin. Il se différencie en trois segments, le duodénum (où commence la digestion par action des acides biliaires et des enzymes digestives), le jéjunum et l'iléon qui débouche sur le gros intestin (Wells and Spence, 2014).

L'intestin grêle se compose d'une paroi et un lumen. La paroi est formée de 4 enveloppes :

- la muqueuse : couche la plus interne contenant des nombreuses glandes pour la production du mucus protecteur. Elle est constituée d'un épithélium qui repose sur un fin tissu conjonctif (ou *lamina propria*). Les replis de l'épithélium et du tissu conjonctif forment des cryptes et des villosités qui augmentent la surface intestinale pour une meilleure absorption des nutriments lors de la digestion.
- la sous-muqueuse : couche intermédiaire où passent les vaisseaux et les nerfs
- la musculuse : couche externe constituée des fibres musculaires
- le péritoine : fin enveloppe qui recouvre la paroi externe de l'intestin

La muqueuse intestinale est recouverte d'une couche de cellules (épithélium) comprenant des types cellulaires différents selon leur localisation (Figure 36) :

- sur les villosités : les **entérocytes**, cellules hautes et étroites, présentent un pôle apical pourvu de microvillosités afin d'augmenter la surface de contact et jouent un rôle d'absorption des nutriments et d'eau ; les **cellules caliciformes** (ou cellules en gobelet) productrices du mucus ; et les cellules **entéroendocrines**, qui produisent des hormones ayant une action locale ou sur d'autres organes.
- au fond des cryptes : les **cellules de Paneth** secrètent des enzymes digestives et participent à la défense immunitaire en sécrétant des enzymes antimicrobiens (les lysozymes) et les **cellules souches**, cellules principales indifférenciées, cylindriques, se divisent activement de manière asymétrique, donnant les différentes cellules qui constituent l'intestin et au fil de la différenciation cellulaire remontent les cryptes pour rejoindre la partie apicale des villosités intestinales (Wells and Spence, 2014; Mahe *et al.*, 2015).

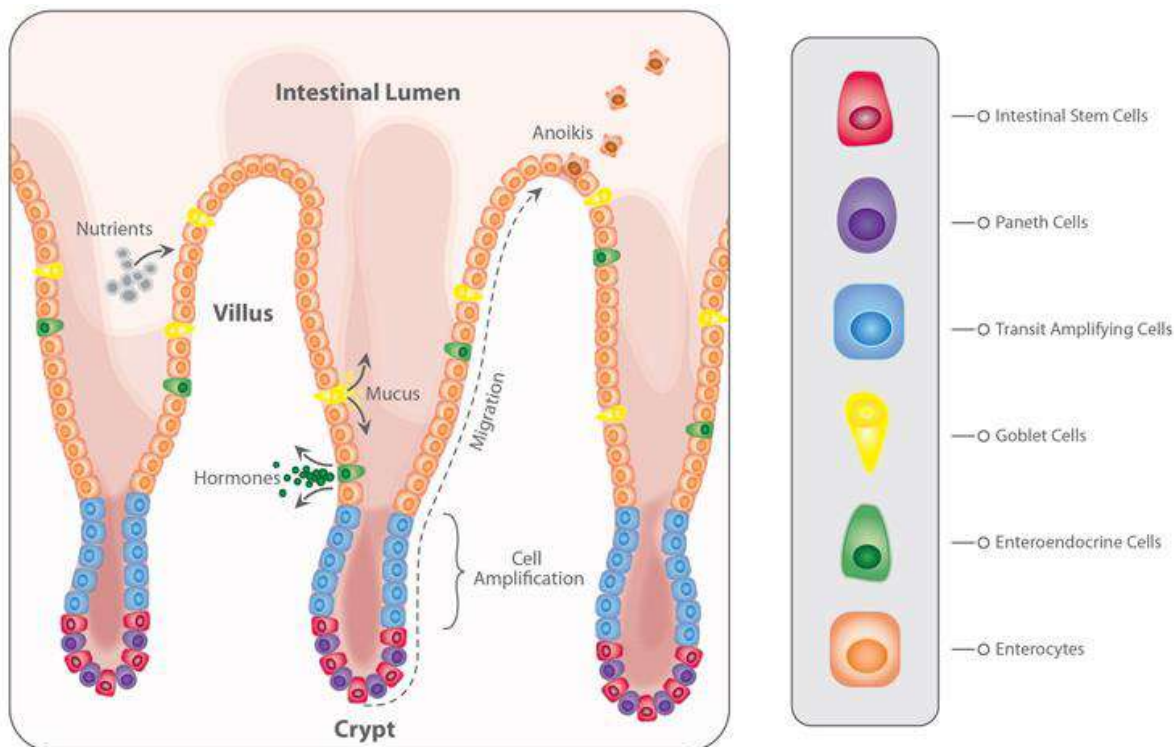


Figure 36 Organisation de l'épithélium intestinal : entérocytes (rose), cellules caliciformes (jaune), cellules enteroendocrines (vert), cellules de Paneth (violet), cellules souches indifférenciées (rouge) et en voie de différenciation (bleu).

(<https://www.stemcell.com/technical-resources/area-of-interest/organoid-research/intestinal-research/overview.html>)

8.2 MISE EN CULTURE DE CELLULES SOUCHES ET LEUR DIFFÉRENCIATION

Pendant longtemps, la culture de cellules épithéliales intestinales humaines n'a pas été concluante pour la multiplication des NoVs car les entérocytes en culture primaire, cultivés directement à partir de tissu humain, subissaient une apoptose trop rapide pour permettre l'évaluation de la réplication virale (Estes *et al.*, 2019). Des lignées de cellules épithéliales intestinales existent, issues de carcinome comme les cellules Caco2, mais ne permettent pas la réplication des NoVs, probablement car elles sont dé-différenciées. Suite à la mise au point d'une culture en trois dimensions (3D) de ces cellules, et inoculation d'une souche de NoVs, un effet cytopathique avait été noté, avec une faible augmentation de la concentration en particules mais la reproductibilité de la technique était faible et aucune suite n'a été donnée (Straub *et al.*, 2011, 2013).

Les avancées scientifiques et le développement du système de culture de cellules souches, ont permis de reconstruire un épithélium intestinal, de souris d'abord et humain ensuite, en laboratoire à partir de cellules souches (Sato and Clevers, 2013; Mahe *et al.*, 2015; Yin *et al.*, 2016; Workman *et al.*, 2017). Ces cellules souches sont isolées à partir de biopsies ou des prélèvements intestinaux issus de donneurs adultes sains, dont on isole les cryptes (Figure 37).

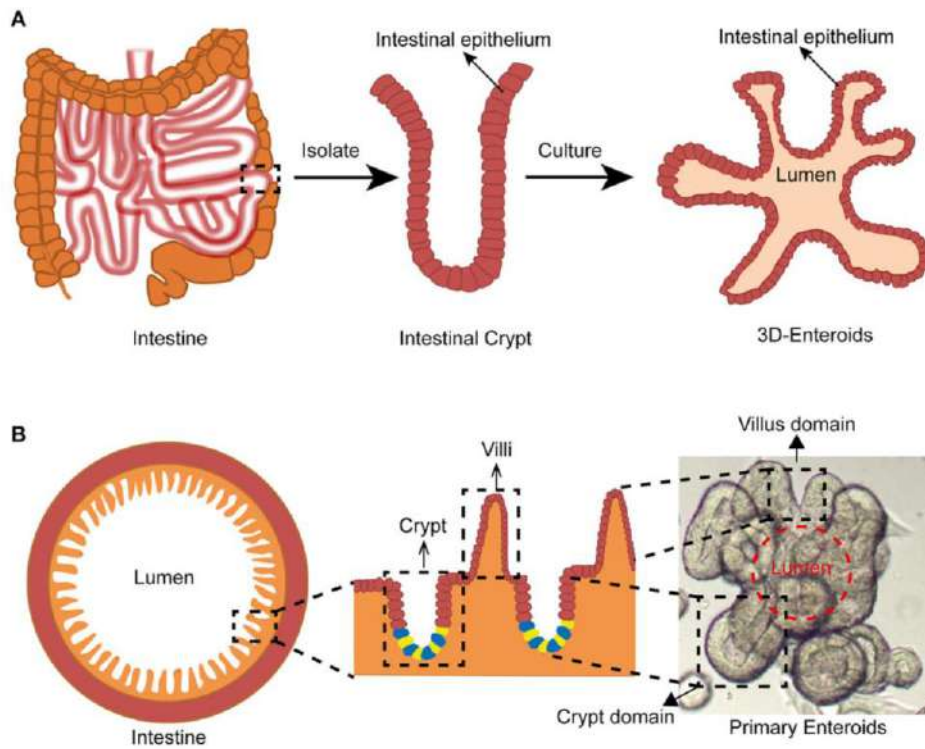


Figure 37 Modèle de culture des entéroïdes.

A) Les cryptes intestinales peuvent être isolées à partir de prélèvements de chirurgie ou de biopsies, suivis d'une culture en entéroïdes 3D. B) Les entéroïdes reconstituent l'architecture d'un intestin sain contenant des villosités et des cryptes (adapté de Yin and Zhou, 2018).

Les cryptes ainsi isolées sont incorporées dans un gel mimant la matrice extracellulaire (Matrigel) et cultivées en 3D dans un milieu enrichi en facteurs de croissance essentiels destinées à favoriser la croissance de cellules souches (Sato and Clevers, 2013; Fujii *et al.*, 2019). Les facteurs nécessaires à cette culture sont produits par des lignées cellulaires stables:

- Wnt : Protéine produite et sécrétée à partir de la lignée cellulaire de souris LWnt3A. Ces cellules sont des fibroblastes qui composent le tissu conjonctif sous-cutané. Le gène Wnt-3a code pour la protéine Wnt, qui aide à réguler la division cellulaire et la spécialisation des organes pendant le développement et la régénération tissulaire (Willert *et al.*, 2003) ;
- R-Spondin : la protéine RSP01 appartient à la famille des R-spondines. Elle induit la prolifération de cellules épithéliales de cryptes intestinales, augmente la guérison épithéliale intestinale et soutient le renouvellement des cellules souches épithéliales intestinales (Warner *et al.*, 2015). De plus, elle potentialise la voie de signalisation activée par Wnt (Mah *et al.*, 2016) ;
- Noggin : Cette glycoprotéine est utilisée comme supplément du milieu de culture pour favoriser le maintien des cellules souches dans un état pluripotent empêchant leur différenciation spontanée. En outre elle favorise la croissance cellulaire (Lugli *et al.*, 2016).

Une seule cellule souche forme une structure symétrique 3D initialement sphéroïdale avec un domaine apical qui fait face à l'intérieur du lumen nouvellement formé et une surface basolatérale en contact avec le Matrigel. Suite à multiplication des cellules, la symétrie est

brisée par la formation de « bourgeons » latérales ressemblante aux cryptes intestinales (Figure 38). Ces structures sont appelées entéroïdes (dérivés de l'intestin grêle) et contiennent tous les types de cellules épithéliales typiques de l'intestin. Ils sont différents des organoïdes qui contiennent en plus du tissu mésenchyme (Mahe *et al.*, 2015).

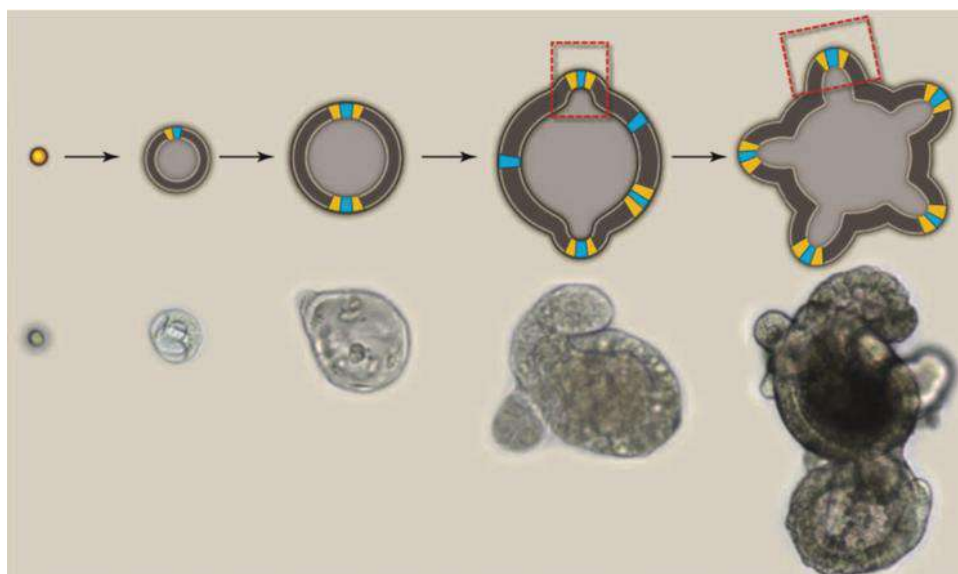


Figure 38 Schématisation des phases de développement des entéroïdes. En évidence, les cellules souches représentées en jaune et les cellules de Paneth en bleu. (Sato and Clevers 2013)

Les entéroïdes ont des tailles variables (plusieurs μm à mm) et doublent de taille généralement en 24h. Après environ cinq jours de culture et le retrait des facteurs de croissance essentiels, les entéroïdes se différencient pour former un épithélium composé d'entérocytes matures, de cellules entéroendocrines et de cellules caliciformes, reconstituant l'architecture d'un épithélium intestinal sain contenant des villosités et des cryptes (Zachos *et al.*, 2016). Pour la culture des NoVs, les entéroïdes en 3D non différenciés sont dissociés à l'aide de la trypsine et sont ensemencés sur des plaques de culture enduites de collagène afin de former un tapis cellulaire en monocouche, dont la face apicale est accessible pour l'infection virale (Figure 39). Le milieu de culture est ensuite appauvri en facteurs de croissance, ce qui favorise alors la différenciation cellulaire des cellules souches dans les différentes cellules adultes qui composent l'épithélium intestinal. Après quatre à cinq jours de différenciation, les cellules sont prêtes à être infectées (Ettayebi *et al.*, 2016; Costantini *et al.*, 2018; Estes *et al.*, 2019).

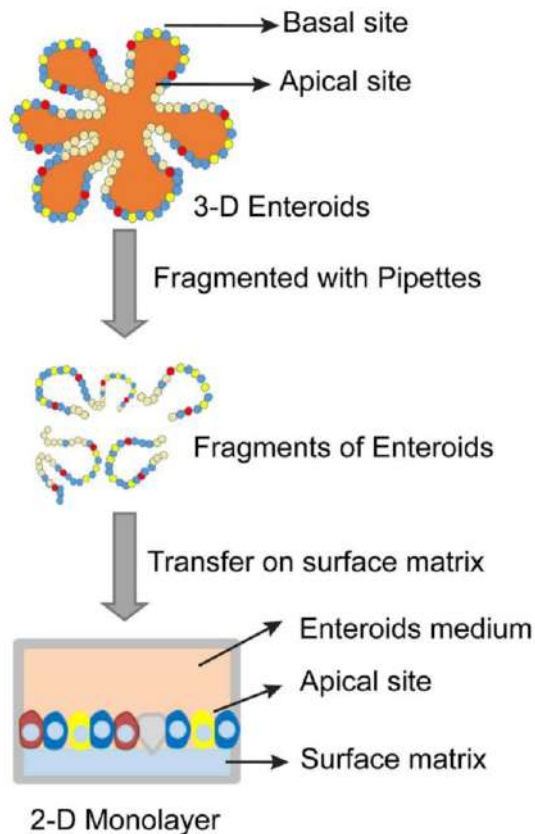


Figure 39 Diagramme schématisée du passage des enteroides de 3D à 2D pour la formation de monocouche (Yin and Zhou, 2018)

8.3 PREPARATION DE L'INOCULUM ET INFECTION

L'inoculum, l'échantillon contenant le virus purifié et préalablement filtré (le plus souvent une solution de selle ou du virus en suspension dans de l'eau), est d'abord quantifiée par RT-PCR en temps-réel. Environ 10^3 particules virales sont nécessaires pour une infection réussie (dose infectieuse 50%, ID_{50}), avec une dose légèrement inférieure pour GII.4 que pour GII.3 (Estes *et al.*, 2019). Ces valeurs d' ID_{50} correspondent aux valeurs calculées lors des essais chez les volontaires humains (Atmar *et al.*, 2014). Toutefois, la quantification par RT-PCR prend en considération toutes les particules virales dans l'inoculum, sans différencier entre infectieuses et non infectieuses. Des adaptations de la concentration estimée de l'inoculum peuvent être nécessaires et sont probablement liés aux conditions des stockages des selles ainsi que de l'âge du donneur. Costantini *et al.* (2018) ont mis en évidence une corrélation entre succès de la replication virale et selles collectées d'individus ≤ 2 ans et stockes à -70°C . Cependant, un article récent démontre que la concentration est plus importante que l'âge du donneur (Chan *et al.*, 2019).

Dans notre laboratoire la dose utilisée est de 10^4 à 10^5 particules virales.

Deux plaques sont préparées et mises en contact avec l'inoculum en parallèle pendant une heure. Après lavages pour éliminer les particules non fixées, une plaque sera congelée 1h après infection et lavage, et représentera le « temps zéro » (T_0) (nombre de particules virales fixées sur les cellules). Dans l'autre plaque, incubée à 37°C , le virus commence à se multiplier environ 12h après infection et atteint un plateau en 24-72h. A partir de ce délai, la plaque est congelée et servira à quantifier le nombre de particules produites.

8.4 REPLICATION DES SOUCHES DE NoVs

Les études conduites sur des volontaires humains avaient donné des indices importants sur la réplication des NoV. Ils avaient nettement mis en évidence la nécessité de la bile et des acides biliaires (par exemple les glycochenodeoxycholic, GCDCA) comme facteurs de croissance dans des lignées cellulaires et ils avaient aussi noté la présence d'antigènes viraux et d'une réponse immunitaire dans les entérocytes de la *lamina propria* de l'intestin grêle (Estes *et al.*, 2019).

De plus, selon le statut génétique de la personne infectée (sécréteur ou non-sécréteur), les HBGA seront exprimés (ou non-exprimés) au niveau des villosités intestinales, affectant la liaison et la réplication de certaines souches de NoVs (Estes *et al.*, 2019; Graziano *et al.*, 2019). Par exemple des NoVs GII.3 peuvent infecter des individus non-sécréteurs (Ruvoën-Clouet *et al.*, 2013; Ettayebi *et al.*, 2016).

Plusieurs souches de NoVs sont cultivables, comme par exemple GI.1, GII.1, GII.3, sept souches différentes de GII.4 et GII.17 (quantifiées par RT-PCR) et GII.6, GII.8, GII.12 et GII.14, déterminé par immunofluorescence (Estes *et al.*, 2019). Une synthèse des souches testées en culture cellulaire est reportée en Annexe II. La plupart nécessite l'addition de bile pour la réplication, mais des différences peuvent exister, comme la souche GII.4 qui se développe sans bile bien que la réplication soit renforcée en présence de bile (Ettayebi *et al.*, 2016).

Malgré de multiples effets connus de la bile sur les fonctions cellulaires (action de détergent, augmentation de la digestion et de l'absorption des graisses, régulation des processus métaboliques et inflammatoires en activant diverses voies de signalisation...), le mécanisme qui affecte la réplication de certaines souches de NoVs n'est pas encore clairement élucidée (Estes *et al.*, 2019). Une augmentation du titre viral n'a été observée que lorsque les entéroïdes, et non le virus, étaient traités avec de la bile, ce qui indique que la bile joue une action sur les cellules et non sur le virus (Ettayebi *et al.*, 2016).

La réplication virale peut être quantifiée par RT-PCR ou par d'autres méthodes telles que la cytométrie en flux (détermination du pourcentage de cellules infectées), par immunofluorescence, par la détection du traitement protéolytique de protéines exprimées à l'aide de Western blots ou par visualisation directe de particules de virus produites par microscopie électronique (Ettayebi *et al.*, 2016). Il semble donc que le succès de la réplication dépende de la concentration virale de départ et de la souche composant l'inoculum (Ettayebi *et al.*, 2016 ; Costantini *et al.*, 2018). Des taux de réplication plus importants ont été reportés pour GII.4 (avec augmentation moyenne du titre de 670 fois) que pour GII.3 (accroissement du titre en moyenne de 48 fois) (Ettayebi *et al.*, 2016).

Dans un échantillon de selle positif en NoVs, une valeur de $Ct \leq 30.1$ mesurée par RT-PCR pourrait indiquer la présence de NoVs infectieux (Chan *et al.*, 2019). Les auteurs proposent cette valeur affirmant au même temps la nécessité d'être vigilants dans l'interprétation d'un valeur seuil fixe en contexte clinique car un protocole de référence capable d'harmoniser la variabilité des tests entre différents laboratoires fait défaut pour la quantification des NoVs. De plus la sensibilité du protocole de réplication actuel nécessite d'autres développements car les variations entre les échantillons et les génotypes du virus mis en culture sont importantes.

8.5 APPLICATION ET DEVELOPPEMENTS FUTURS

La réplication *in vitro* des NoVs va permettre de nombreux développements jusqu'alors impossibles. Par exemple cette approche a permis d'évaluer l'efficacité de différentes techniques d'inactivation. Le contact des virus avec des solutions à base d'alcool (70% d'éthanol ou isopropanol) pendant 5 min n'a pas été suffisant pour inactiver complètement les virus, alors que des solutions à base de chlore (≥ 50 ppm) ont inactivé le virus en 1 min (Costantini *et al.*, 2018). Considérant les traitements thermiques, une exposition du virus à 60°C pendant 15 minutes suffit pour l'inactivation (Ettayebi *et al.*, 2016).

Ces résultats montrent l'importance et les applications possibles de ce système de culture pour évaluer des mesures de désinfection. Toutefois cette approche est encore en cours d'optimisation. Augmenter le nombre des souches cultivables et simplifier le protocole actuellement existant, représenteraient des avancées importantes.

En terme de recherche plus fondamentale, la culture d'enteroïdes pour la réplication des NoVs, permettra de mieux connaître et comprendre la biologie du virus (identification du récepteur, des voies d'internalisation, des mécanismes de réplication virale) et d'avancer sur les approches à utiliser pour induire la neutralisation du virus afin de développer des stratégies pour la prévention ou le contrôle des infections par les NoVs (Ettayebi *et al.* 2016 ; Estes *et al.*, 2019).

OBJECTIFS DES TRAVAUX DE THESE

Cette thèse se situe dans un moment particulier pour l'étude des NoVs, avec d'importantes avancées biologiques et technologiques, telles que la mise au point d'un système de culture d'enteroïdes pour étudier la réplication virale, et le développement de la métagénomique pour identifier les souches présentes dans l'environnement.

Ces avancées ont constitué les deux lignes conductrices de mes travaux, avec une implication plus forte sur la métagénomique. Mes travaux comportent deux parties:

- Travail de développement méthodologique pour optimiser la détection de ce virus, améliorant les taux de récupération du génome virale lors des étape d'élution et concentration en vue de la préparation des librairies pour le séquençage metagenomique. Le Chapitre 1 décrit les études sur la spécificité des mucines gastriques porcines (PGM) à lier les capsides des NoVs. Cette approche avait comme objectif d'enrichir les extraits en NoVs et potentiellement de donner des informations biologiques sur l'état du virus concentré, basée sur l'intégrité de la capside virale. Les articles, publiés ou soumis, présentent les travaux méthodologiques réalisés afin d'optimiser la détection des NoVs par un approche métagénomique. Je me suis pleinement investie dans l'analyse des séquences et ayant eu l'opportunité de réaliser un séjour chez M. Koopmans je me suis formée à la bio-informatique ;
- Étude de la persistance de l'infectiosité de souches de NoVs en utilisant les enteroïdes (modèle développé par M. Desdouits dans notre laboratoire en lien avec l'équipe de M.K. Estes). Ce travaux, encore preliminaires, sont présenté dans le Chapitre 2.

TRAVAUX ORIGINAUX

CHAPITRE 1 : ETUDE DE LA SPECIFICITÉ DES PGM

INTRODUCTION

Les mucines représentent l'élément principal du mucus, gel très hydraté (95% d'eau) qui recouvre la paroi interne des organes creux de l'organisme. Les mucines appartiennent à la famille des glycoprotéines et sont produites par des cellules sécrétrices caliciformes présentes tout au long du tractus gastro-intestinal (Madsen *et al.*, 2016).

Les différentes mucines rencontrées chez les mammifères partagent certaines caractéristiques communes (Dekker *et al.*, 2002) :

- Il s'agit de polymères de grande taille, composés de sous-unités (monomères) de glycoprotéines hautement glycosylées avec un noyau central peptidique entouré d'une série d'oligosaccharides.
- Les chaînes latérales glucidiques représentent classiquement 80-85% du poids de la mucine. Elles sont constituées des sucres complexes tel que le galactose (Gal), fucose (Fuc), N-acétyl-galactosamine (GalNAc), N-acétyl-glucosamine (GlcNAc) et d'acides sialiques en quantités variables selon les types de mucines.

Certains sucres exprimés sur la surface des mucines sont semblables aux sucres qui composent les antigènes tissulaires de groupe sanguin (HBGA), responsables d'une liaison spécifique de la capsid des NoVs (McGuckin *et al.*, 2011; Le Pendu *et al.*, 2014). Pour ces raisons, les mucines gastriques porcines (Porcin Gastric Mucins, PGM) ont été proposées pour étudier l'intégrité du site de reconnaissance de ce ligand dans la capsid virale et en conclure des informations sur le pouvoir infectieux du virus, la résistance à certaines conditions environnementales ou comme méthode de sélection des NoV parmi les autres microorganismes présents dans l'échantillon (Cannon and Vinjé 2008; Tian *et al.*, 2011 ; Tian *et al.*, 2017 ; Dancho *et al.*, 2012). Cependant à la lecture de certains articles il nous a semblé que des détails concernant des contrôles de spécificité et de sensibilité étaient absents. Nous avons voulu compléter ces aspects dans le but d'utiliser cette méthode de sélection des NoVs parmi les autres microorganismes présents dans l'échantillon et d'en déduire des informations biologiques complémentaires, comme des informations sur l'intégrité de la capsid virales et des sites de reconnaissance responsables de la liaison du virus aux récepteurs cellulaires, étapes primordiales pour la multiplication cellulaire.

Afin de vérifier leur spécificité, nous avons comparé les PGM et les mucines bovines extraites des glandes sous-maxillaires (Bovin Submaxillary Mucin, BSM). Les deux typologies des mucines sont caractérisées par leur présence à la surface de la membrane cellulaire et pour former un gel muqueux. Toutefois, elles diffèrent dans la quantité d'acides sialiques (1% et 9-17% pour les PGM et BSM, respectivement) et dans la structure des sucres qui les composent (Figure 40), ce qui peut avoir un effet sur la charge moléculaire impactant la liaison des NoVs aux mucines.

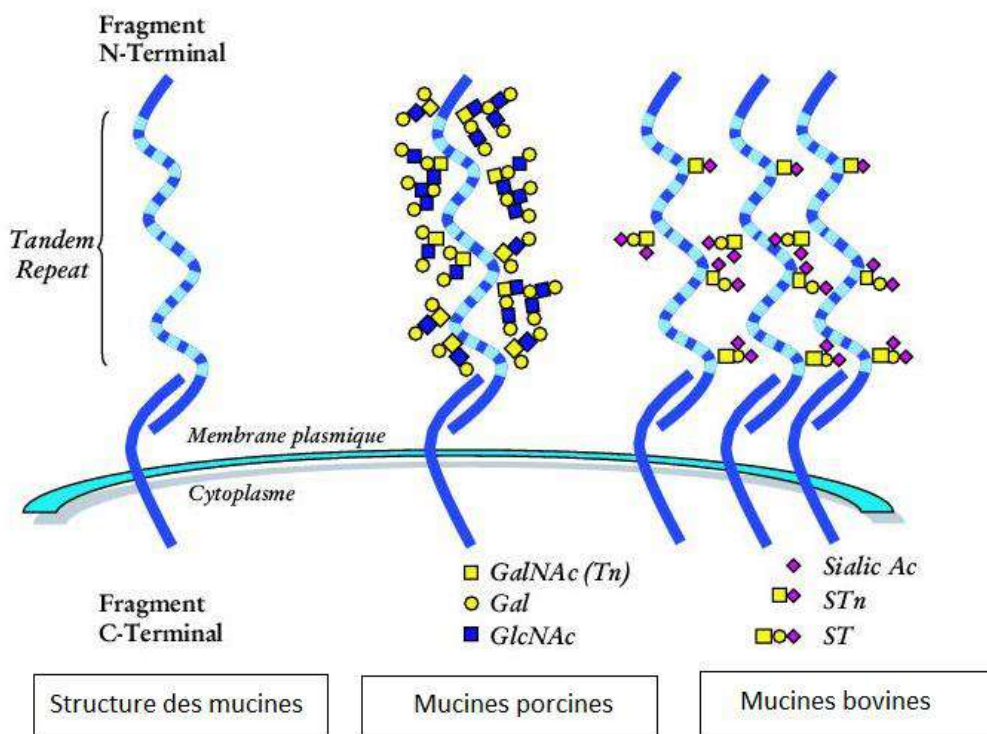


Figure 40 Structure et composition des mucines porcines et bovines.

Les structures de base des chaînes de O-glycosylations sont représentées. Les mucines bovines portent des chaînes de O-glycosylations courtes de type Tn (N-acétyl-galactosamine seul) ou « core 1 » ($\text{Gal}\beta 1,3\text{GalNAc}$ ou antigène T), largement sialylées. Les mucines porcines portent plutôt des O-glycosylations de type « core 2 » ($\text{Gal}\beta 1,3(\text{GlcNAc}\beta 1,6)\text{GalNAc}$) qui peuvent en plus être décorées par des motifs fucosylés de type HBGA (antigène H, antigène A ...) non représentés sur le schéma (Modifié de Limacher and Acres, 2007)

Objectifs de l'étude :

- vérifier la spécificité des PGM dans la reconnaissance des VLP GI.1 et GII.4 ;
- évaluer la sensibilité de liaison en utilisant différentes concentrations de PGM ;
- inclure des contrôles négatifs tels que les BSM et les BSM désialylés ;
- évaluer la présence d'un effet souche-dépendant.

MATERIEL ET METHODES :

Pour répondre à ces questions un test ELISA semi-quantitatif et une approche 'In-Situ-Capture-PCR' (ISC-PCR) ont été développés. La mise au point du test ELISA a commencé par différents essais d'optimisation de la méthode (sélection du tampon de « coating », durée et température d'incubation de la plaque pour favoriser la fixation des mucines, dosage protéique, gamme de dilution des VLPs et des anticorps). Le test ELISA a été réalisé en utilisant des VLPs GI.1 et GII.4 et leurs anticorps spécifiques.

Nous avons également utilisé le protocole ISC-PCR pour évaluer le comportement de souches de NoV envers ces mucines. Dix-huit échantillons de selles positives en NoV (6 souches de GI et 12 de GII) ont été testés basé sur le protocole publié par Zhou *et al.* (2017) (Annexe III présentant le détail des souches utilisées ; Annexe IV avec le chemin des protocoles ELISA et ISC-PCR).

- Dé-sialylation des BSM

Une solution à 1 mg/mL de BSM (Sigma-Aldrich) a été incubée avec H_2SO_4 (concentration

finale de 0,2 M) pendant 1h à 80°C dans un bain à sec. Puis la solution est neutralisée avec 1,10 volumes de NaOH 0,2 M et dialysée avec du PBS en utilisant des concentrateurs Vivaspin Turbo 15 capable de retenir les protéines >30 KDa (Sartorius) (travail réalisé par A. Breiman).

- Dilution et préparations

Tampon coating (TC) : Carbonate de Sodium 0.1 M et Bicarbonate de sodium 0.1 M, pH 9,5

Préparation des mucines : des dilutions en série de 10 en 10 ont été préparées avant chaque essai en utilisant du TC. Pour les PGM les concentrations obtenues étaient entre 100 mg/mL et 0.1 µg/mL. Pour les BSM et les dBSM, elles étaient comprises entre 10 mg/mL et 0.1 µg/mL.

Purification des selles : une solution de selle à 10% en PBS a été préparée, mélangée au Vortex (30 sec) et FastPrep (20 sec) avant traitement aux ultrasons (3 fois pendant 1 min intercalé d'une minute sur la glace) suivi d'une centrifugation pendant 10 min à 1500 x g. Le surnageant a été récupéré et re-centrifugé avec les mêmes conditions. Le surnageant a été filtré en série (de 5 à 0.22 µM), répartis dans différents tubes et stocké à -20°C.

- test ELISA

Les dilutions de mucines ont été distribuées dans les puits de plaques Nunc Maxisorp (ThermoFisher Scientific, Roskilde, Denmark) (100 µL/puit), ainsi que des VLPs (1 µg/mL) pour le contrôle positif (T+) et du TC pour le contrôle négatif (T-). Trois puits ont été préparés pour chaque condition testée avec une incubation pendant la nuit à 37°C en chambre humide. Le lendemain, seuls les puits du T+ ont été lavés 3 fois avec le tampon de lavage (PBS-tween 0.05). Tous les puits ont été saturés avec une solution de Bovine Serum Albumine (BSA) à 3% en PBS, pendant 40 min à 37°C. Après trois lavages par du PBS-Tween 0,005% (TPBS) les VLPs (10µg/mL) ont été ajoutées dans les puits et l'ensemble a été incubé pendant 1h à 37°C. Après trois lavages une nouvelle incubation a été réalisée avec l'anticorps (Ac) I^{aire} pendant 1h à 37°C, puis l'Ac II^{aire} de nouveau pendant 1h à 37°C, avec trois lavages entre chaque étape. La révélation a été effectuée par addition de tétraméthylbenzidine, (TMB) 100 µL dans chaque puit et incubation entre 12 et 15 min à l'obscurité avant arrêt de la réaction par addition de H₃PO₄ 50 µL/puit. La densité optique (DO) a été lue à 450nm avec un spectrophotomètre (Safire ; Tecan).

Ce test permet d'avoir une approche semi-quantitative en calculant un ratio (R) qui correspond à la DO de l'échantillon divisée par la DO des témoins négatifs constitués du tampon de « coating » sans les VLPs mais avec les anticorps afin de représenter le bruit de fond. Une valeur seuil a été fixée sur la base des expériences précédentes (Maalouf *et al.*, 2010). Si R est ≥ 2 l'échantillon est considéré positif, significatif d'une liaison spécifique.

- ISC-PCR

La préparation des plaques a été effectuée comme décrit ci-dessus avec deux puits par dilutions de PGM (identique à celles décrites ci-dessus) et incluant un T- (TC). Après

saturation par du PBS-BSA 3% pendant 40min à 37°C et trois lavages, les extraits de selles (100 µL/puit) ont été ajoutés et incubés pendant 1h à 37°C. Les duplicats ont été ensuite extraits séparément et quantifiés par rRT-PCR. Une dilution des selles à 10% en PBS et extraite directement a été utilisée comme témoin.

RESULTATS PRELIMINAIRES

Lors de la mise au point du test ELISA, deux températures (4°C et 37°C) et deux durées d'incubation (1h et toute la nuit) ont été comparées. Une température de 37°C pendant toute la nuit a été retenue comme la meilleure condition pour favoriser le « coating » des PGM et a été sélectionnée pour le reste de l'étude (Figure 41).

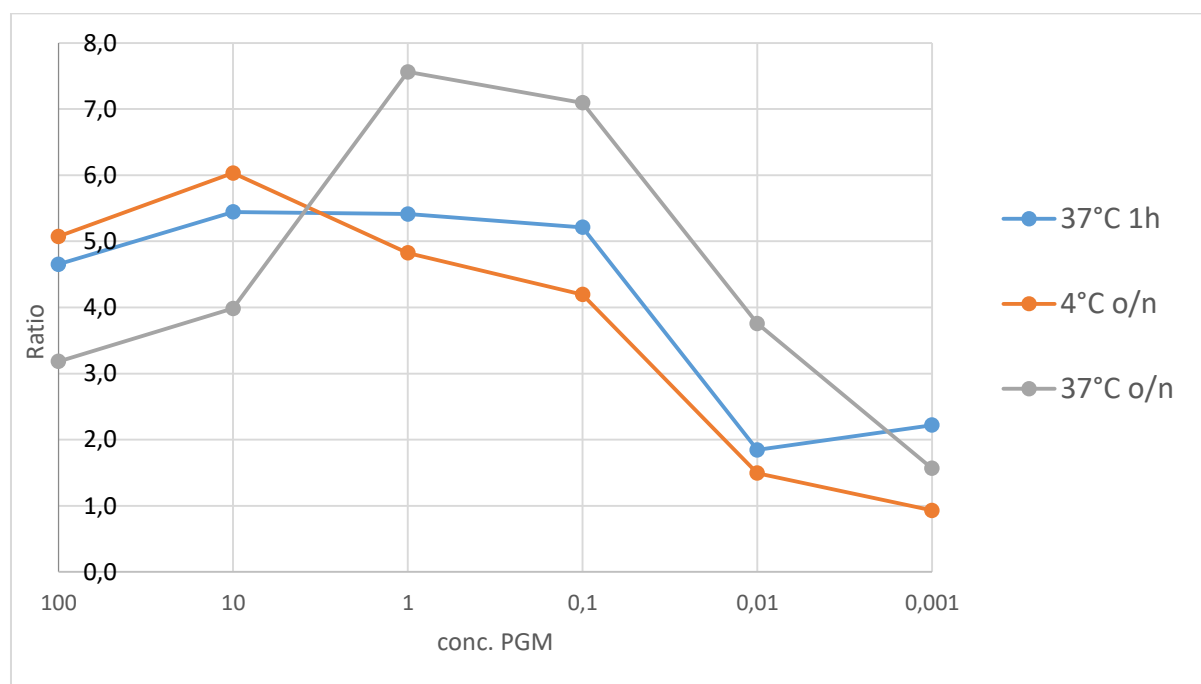


Figure 41 Choix des conditions à adopter pour favoriser le « coating » des mucines. Les PGM ont été testés à 37°C pendant 1h (ligne bleu), 37° toute la nuit (o/n) (gris) et 4°C o/n.

Un test ELISA a été mené afin d'évaluer la composition des ligands glycosylés exprimés par les PGM et les BSM avant et après désialylation (Figure 42). Nous disposons d'anticorps anti-Tn qui reconnaît le GalNAc seulement quand il est bien exposé (non masqué par des autres molécules comme les acides sialiques) ainsi que d'un anticorps qui reconnaît le GalNAc sialylé (STn). Comme attendu, les mucines bovines sont fortement positives pour le STn et le signal disparaît après traitement à l'acide sulfurique, alors que la quantité de Tn augmente. Par ailleurs, les PGM sont négatives pour le STn, mais sont environ 3,5 fois plus riches en antigènes fucosylés (A et H type 2) que les BSM. Nous avons également constaté que le marquage par UEA-I était diminué par le traitement de dé-sialylation ce qui suggère que l'acide sulfurique altère aussi les antigènes fucosylés, quoique dans une moindre mesure.

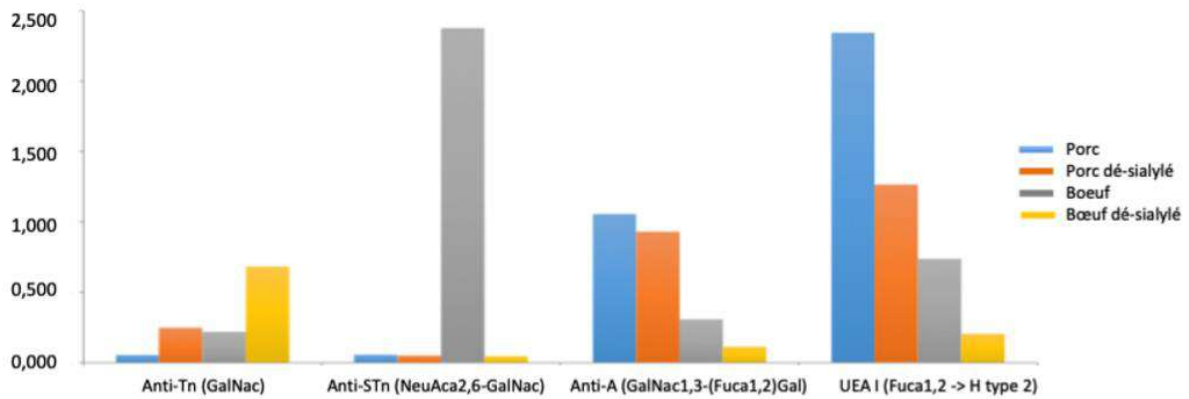


Figure 42 Test ELISA qui montre les ligands glycosylés exprimés par les mucines de porc (bleu) et de boeuf (gris) ainsi que l'efficacité de la désialylation (mucines de porc désialylés en orange, de boeuf en gris) (Adrien, INSERM)

Après validation des conditions de « coating » des plaques ELISA et de la désialylation correcte des BSM, des tests ELISA ont été réalisés pour vérifier la spécificité de liaison entre PGM, BSM, dBSM et VLPs de NoVs GI.1 et GI.4. Pour chaque plaque, des contrôles positifs (puits « coatés » avec les VLP, sans mucines), et négatif (puits avec tampon BSA) ont permis de valider l'analyse et de calculer le ratio entre la DO de l'échantillon et du contrôle négatif. Un effet dose-réponse en cloche a été mis en évidence pour la fixation des VLPs des deux souches NoVs sur les PGM, avec un maximum de liaison à une concentration de 0.1 µg/mL (Figures 43 et 44). Une réponse non spécifique a été observée pour les mucines bovines, BSM et dBSM, avec une seule expérience positive pour les VLP GI.1, sans effet considérable de la désialylation.

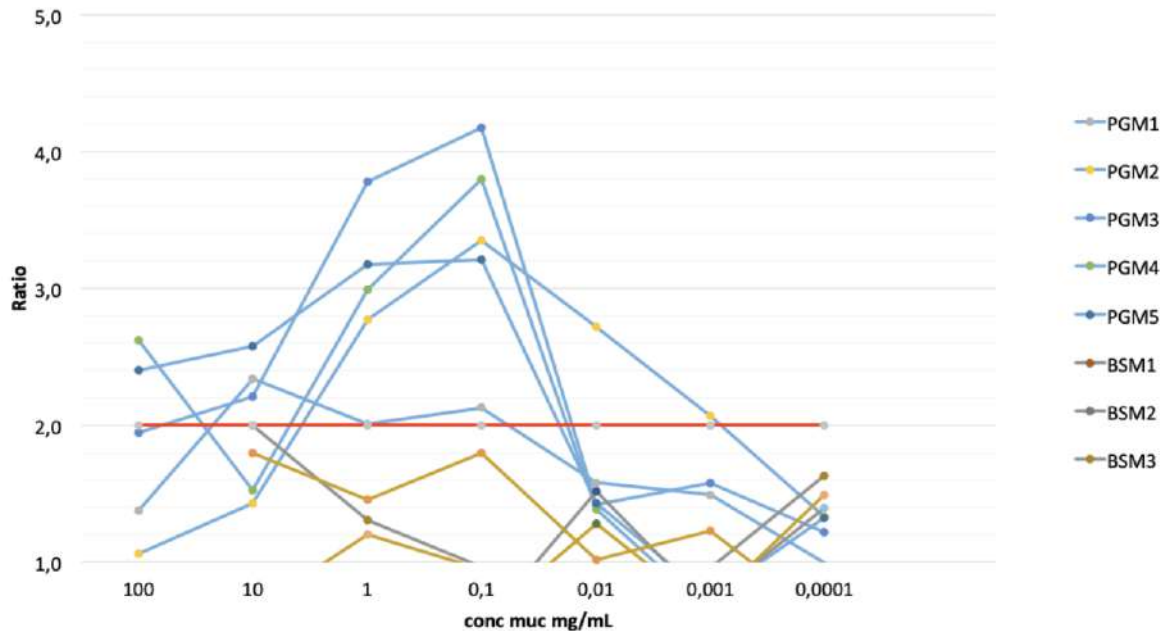


Figure 43 Courbe effet dose-dépendante de la fixation des VLP GII.4 sur les PGM. La limite de positivité est fixée à 2 (ligne rouge). Les BSM (lignes grises) et les dBSM (marron) sont inférieures à la limite de positivité. Les PGM (lignes bleues), dépassent cette limite, avec la fixation maximale est observée pour une concentration en mucine de 0,1mg/mL. Les différentes courbes correspondent à différentes expériences indépendantes.

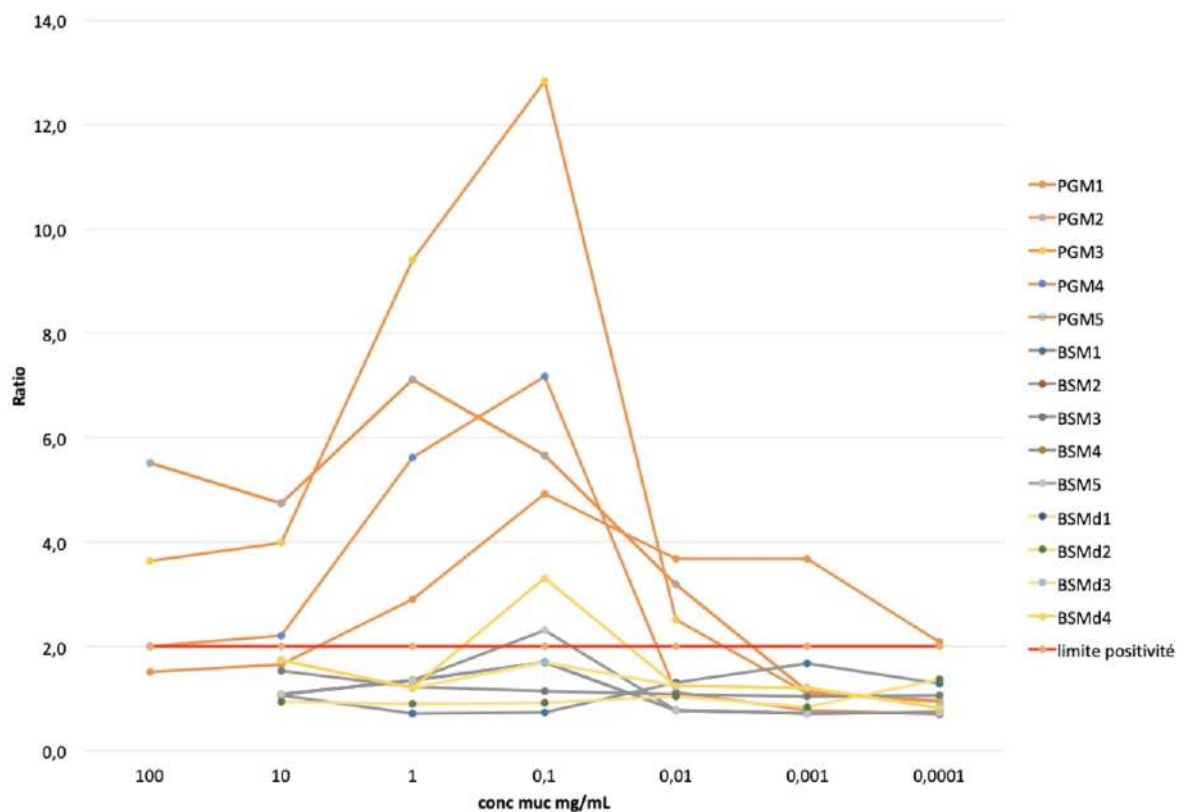


Figure 44 Courbe effet dose-dépendante des liaisons entre PGM et VLP GI.1. La limite de positivité est fixée à 2 (ligne rouge). Les BSM (lignes grises) et dBSM (jaune) sont inférieures à la limite de positivité. Pour les PGM (orange), le maximum de fixation est obtenu, comme avec les capsides GII, pour une concentration de 0,1 mg/ml. Les différentes courbes correspondent à différentes expériences indépendantes.

Après avoir démontré la présence d'une interaction spécifique entre PGM et VLPs, nous avons voulu appliquer une approche ISC-PCR pour étudier l'affinité de liaison sur différentes souches de NoVs à partir d'échantillons de selles purifiées. Les résultats préliminaires ont montré un écart moyen (ΔCt) de 9 Ct entre le témoin et après capture avec PGM, sans effet dose-réponse (Tableau 8). Malgré divers essais d'optimisation du protocole en variant le nombre des lavages et le temps d'incubation, aucun résultat satisfaisant n'a été obtenu à l'heure actuelle.

Tableau 8 Exemple d'une manip faite en utilisant une souche GI.1 (S538).

PGM mg/mL	Moyenne	Ecart-type	ΔCt
100	37,2	0,6	8,0
10	36,7	0,4	7,6
1	37,9	1,1	8,7
0,1	37,3	0,6	8,1
0,01	37,5	0,8	8,4
0,001	38,1	0,8	9,0
0,0001	38,6	0,1	9,5
TC	No Ct	No Ct	No Ct
S538 direct	29,14	0,2	
S538 1/10	31,98	0,1	

La moyenne et l'écart-type sont calculés sur les Ct des trois répliquas pour chaque concentration de PGM. Le ΔCt est la différence entre le contrôle (selle extrait sans utilisation des PGM et la valeur de Ct obtenue avec PGM)

DISCUSSION

Nos résultats montrent une spécificité de liaison entre PGM et les VLP de NoVs GI.1 et GII.4 avec une relation dose-réponse dite *non-monotone*. L'analyse graphique des données montre une courbe en forme de cloche avec une liaison évidente à faible dose mais pas à forte dose. La même concentration optimale a été observée pour les deux souches testées. Des fixations non spécifiques ont été observées pour les BSM et les BSM désialylées (dBSM). En utilisant des selles de NoVs GI.1 et GII.4, une faible fixation a été observée une seule fois mais sans effet dose-réponse évidente. De façon plus surprenante cette même approche en utilisant les PGM et des selles de GI.1 et GII.4 n'a pas donnée de résultats concluants. Ce protocole ISC-PCR nécessite vraisemblablement une optimisation en variant l'impact de la durée et de la température d'incubation, en considérant le nombre des lavages et en ajoutant des contrôles avec les VLP en parallèle. Un aspect également important à considérer est la concentration en particules virales utilisées ($1,73 \times 10^5$ et $2,68 \times 10^6$ pour GI.1 et GII.4 respectivement) dans les selles, inférieure à celles utilisées avec les VLP (10^{11} VLP/mL). L'utilisation d'une série de dilutions de selles pourrait être également intéressante pour la bonne réussite de l'ISC-PCR. Inversement des dilutions de VLP pourraient permettre d'évaluer l'impact de la quantité de capsides mise en contact avec les PGM pour la sensibilité de ce test. La récente découverte des vésicules contenant plusieurs particules virales lors de l'excrétion du virus dans les selles (Santiana *et al.*, 2018) suggère que ces vésicules pourraient empêcher l'interaction entre les capsides virales et les PGM. Des lavages des selles avec des détergents tels que Tween, Trizol ou Pyrophosphate pourraient être nécessaires. Cependant la méthode utilisée ici est proche de celle utilisée pour mettre en évidence du NoVs infectieux sur entéroïdes et devrait être suffisante. Un surnageant de

NoVs provenant de la culture cellulaire pourrait être comparé à un surnageant de selles (lavées et non lavées).

Enfin, pour assurer que le lavage ne soit pas trop agressif sur les capsides, des VLP soumises au même protocole de lavage pourraient être utilisées comme contrôle.

CHAPITRE 2 : ETUDE DE LA PERSISTANCE DES NOVs DANS L'EAU DE MER

INTRODUCTION

Les NoVs humains excrétés à forte concentration par les individus infectés, résistent dans l'environnement, et ainsi contaminent les eaux de surface et les zones côtières. Une conséquence majeure de la contamination de ces eaux est la contamination possible d'aliments lors de leur production et ainsi la survenue d'épidémies alimentaires.

Comme indiqué précédemment, à l'heure actuelle, seule la rRT-PCR est utilisée pour détecter les virus dans les aliments et donc sans information sur l'infectiosité du virus détecté. La multiplication *in vitro* des NoVs étant une technique récente, un virus de substitution a souvent été utilisé pour évaluer la persistance du pouvoir infectieux en se basant sur des similitudes permettant l'extrapolation des résultats. Parmi les *Caliciviridae*, le virus Tulane, un virus simien, est celui qui semble biologiquement le plus proche des NoVs, notamment dans la transmission par voie fécale-orale, dans l'interaction avec les HBGA et les signes cliniques chez les singes, tout en étant facilement cultivable *in vitro* (Cromeans *et al.*, 2014; Arthur and Gibson, 2015; Farkas, 2015). Dans les huîtres, des travaux réalisés antérieurement au laboratoire ont montré que la distribution dans les différents tissus et la persistance du virus Tulane étaient similaires au virus de Norwalk (GI.1) comme démontré par rRT-PCR (Drouaz *et al.*, 2015). La poursuite de ces travaux a montré une persistance de virus Tulane infectieux pendant environ trois semaines au sein des tissus de l'huître (Polo *et al.*, 2018). L'équipe de M. Estes (BCM, Houston) nous ayant transmis les cellules souches intestinales humaines permettant la réplication des NoVs, nous avons mis en place cette technique au sein de l'équipe (M. Desdouits) et l'avons appliqué à l'étude de la persistance des NoV dans l'environnement littoral.

Objectifs de cette étude :

- Estimer la persistance de l'infectiosité des NoVs dans l'eau de mer en utilisant le modèle de culture cellulaire des enteroïdes (Ettayebi *et al.*, 2016) ;
- Comparer la stabilité des NoVs à celle du TV infectieux.

MATERIEL ET METHODES

Deux expériences indépendantes, chacune ayant duré un mois, ont été réalisées. Pour chaque expérience, des flacons stériles contenant 100 mL d'eau de mer naturelle passée sur filtre à sable (provenant de la station Ifremer de Bouin) ont été contaminés séparément avec du TV et NoV GII.3 pour la première expérience, et avec du TV, NoVs GII.3 et GII.4 pour la deuxième. Après homogénéisation, le mélange a été distribué dans 10 tubes en verre de 10mL pour chaque virus. Les tubes ont été incubés à 12°C sous agitation à l'obscurité. Pour l'analyse de la persistance du génome et de l'infectiosité, un tube a été prélevé à intervalles réguliers pour chaque virus. Des tubes d'eau de mer sans virus ont également été préparés et traités en parallèle, en tant que témoins négatifs pour s'assurer de l'absence de contamination au cours de l'expérience.

Pour le suivi par RT-PCR, 100 µL de l'échantillon ont directement été mélangé au tampon de lyse du kit NucliSENS (BioMérieux, France) puis les diverses étapes du kit ont été suivies.

Pour le test d'infectiosité, le protocole mis au point pour la récupération de TV infectieux depuis les tissus d'huîtres (Polo *et al.* 2018) a été adapté. Brièvement, 9.9 mL d'eau de mer ont été filtrés

en série (de 5 μm jusqu'à 0.22 μm) et concentrés par filtration tangentielle en utilisant des tubes Amicon® retenant les molécules de plus de 100 kDa. Le retentât a été transféré sur des colonnes de dessalage Sephadex (PD-10) préalablement conditionnées avec du milieu d'infection. A partir de cet éluât de 1,5 mL, 100 μL sont prélevés pour vérifier l'efficacité des étapes de concentration et purification par rRT-PCR et le reste est utilisé pour les tests d'infection : TCID₅₀ sur cellules LLC-mk2 pour le TV, répllication sur entéroïdes pour les NoVs. Les protocoles détaillés pour la préparation des plaques TCID₅₀ et des entéroïdes en monocouche sont décrits dans l'ANNEXE V et VI. Après l'incubation, la multiplication des NoVs a été calculée par la comparaison de la quantité de génome viral après 1 h et 72 h, suivant le protocole de Ettayebi *et al.* (2016). Pour le TV, l'effet cytopathique dans les plaques de culture a été évalué 5 jours après infection et la concentration virale a été calculée selon le modèle mathématique proposé par Reed et Muench (1938).

RESULTATS PRELIMINAIRES

Suite au protocole de récupération et purification des virus, aucun effet de toxicité cellulaire n'a été observé. Ce protocole a été aussi efficace pour concentrer les NoVs que le TV et aucune perte de particules virales n'a été constatée. Pour chaque point de mesure, un pourcentage de récupération (inoculum/eau de mer) a été calculé afin d'évaluer l'efficacité du processus de purification.

Des petites variations peuvent cependant être notées selon les virus. Ainsi pour les deux souches de NoVs, la proportion de virus récupéré après purification montre peu de variabilité tout au long de l'étude, alors que pour TV la variabilité a été beaucoup plus importante avec une efficacité moindre de l'étape de purification en fin d'expérimentation (Figure 45). Le rendement moins homogène de TV et la tendance à une diminution de la récupération virale au fil du temps suggèrent une fragilité de TV plus importante par rapport aux NoVs. Cependant des résultats complémentaires sont nécessaires pour confirmer ces premières observations.

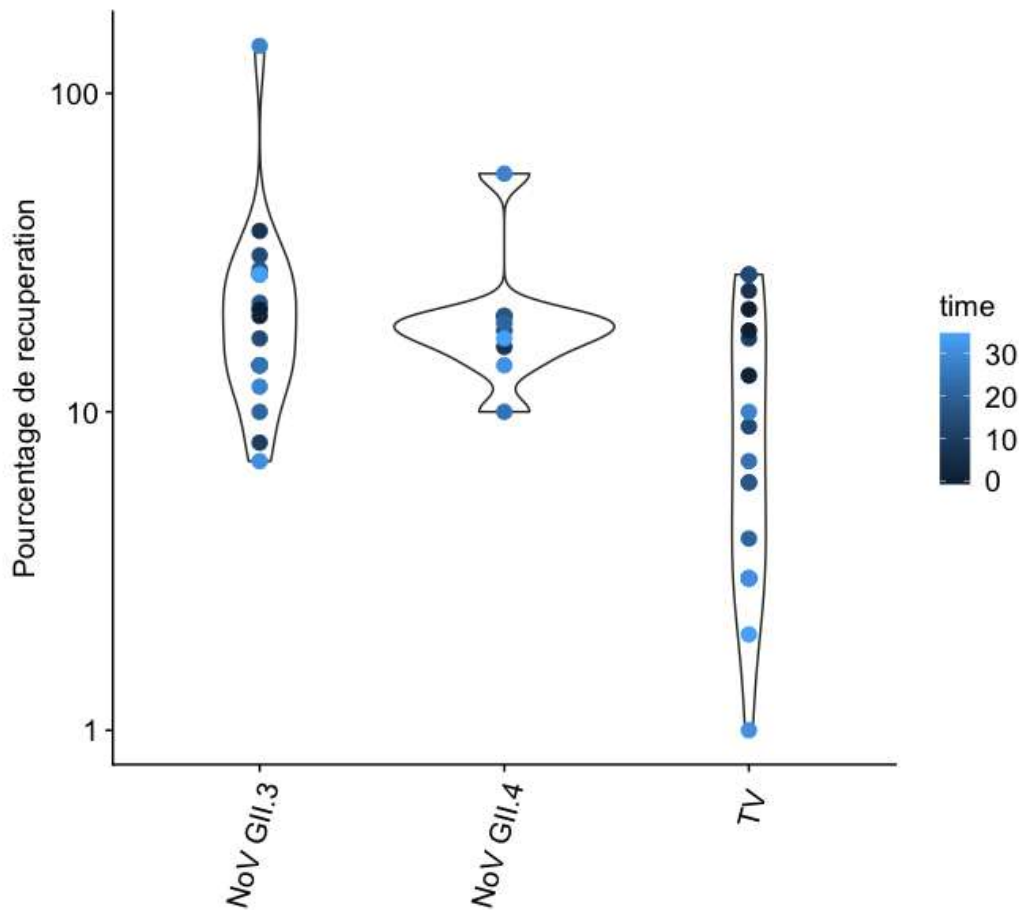


Figure 45 Diagramme en violon de la récupération d'ARN de NoVs GII.3, GII.4 et TV suite au protocole de purification des échantillons d'eau de mer. Les points représentent chaque prélèvement du début de l'étude (temps 0 jour) au dernier prélèvement (T34 jours) dans une échelle de couleurs du bleu foncé au bleu clair.

La quantification du génome viral dans l'eau de mer (Edm) par qRT-PCR a permis de détecter la présence de NoVs et TV tout au long de l'étude (Figure 46). La concentration de TV, un log plus élevée au début de l'expérience (T0) montre une diminution rapide dès le premier prélèvement et 1log de perte est observé après une semaine (T7). La forte décroissance de TV montre une fragilité d'ARN par rapport aux NoVs, plus stables. Pour les souches GII.3 et GII.4, la diminution en terme de copie de génome est moins importante (<1log) et la pente de décroissance est similaire pour les deux souches. Les variations observées par exemple pour la souche GII.4 peuvent être dues à l'imprécision de la rRT-PCR pour la quantification ou à la présence d'agrégats viraux prélevés aléatoirement lors de l'échantillonnage.

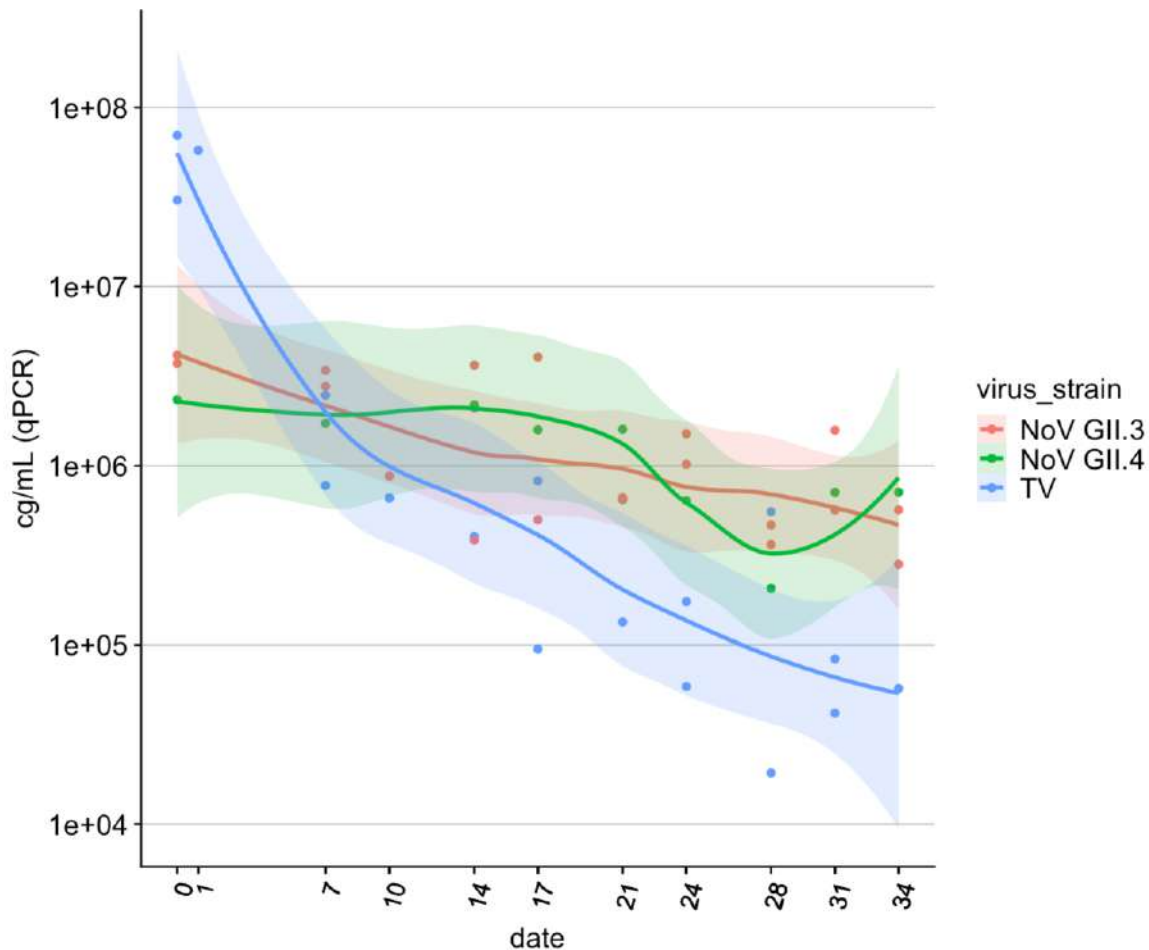


Figure 46 Concentration de l'ARN viral de TV et des souches GII.3 et GII.4 de NoVs quantifiées par rRT-PCR et exprimées en copies génome par mL (cg/mL) dans l'eau de mer. Le halo représente l'intervalle de confiance.

Outre le suivi du génome viral, nous avons pour la première fois pu évaluer directement la stabilité des NoVs GII.3 et GII.4 infectieux par leur répllication sur entéroïdes humains. Il est important de se rappeler que l'approche de culture cellulaire sur enteroïdes est qualitative et non quantitative. Le génome viral est quantifié 1h et 72h après infection des cellules, en triplicat, et la moyenne des différences (ΔCt) entre ces deux mesures permet de calculer un ratio d'amplification ($2^{\Delta Ct}$). Un ratio > 3 indique la présence d'un virus infectieux, et a été observé pour les deux virus tout au long de la deuxième expérience, et pour le GII.3 jusqu'à 24 jours lors de la première expérience (Figure 47). De plus, le ratio d'amplification montre une décroissance pour les deux souches lors des deux expériences, qui pourrait refléter la dégradation du virus et sa moindre capacité à se répliquer. Sur la plupart des échantillons, deux mesures ont pu être obtenues en utilisant des inoculum viraux purifiés depuis l'eau de mer, frais ou après congélation à 80°C, et montrent des résultats similaires. Parfois, la mesure à partir d'inoculum frais n'a pas été possible, ce qui peut expliquer une partie des variations observées pour la souche GII.4 notamment. Ces aléas et la probable variabilité naturelle des virus soulignent la nécessité de répéter ces expériences.

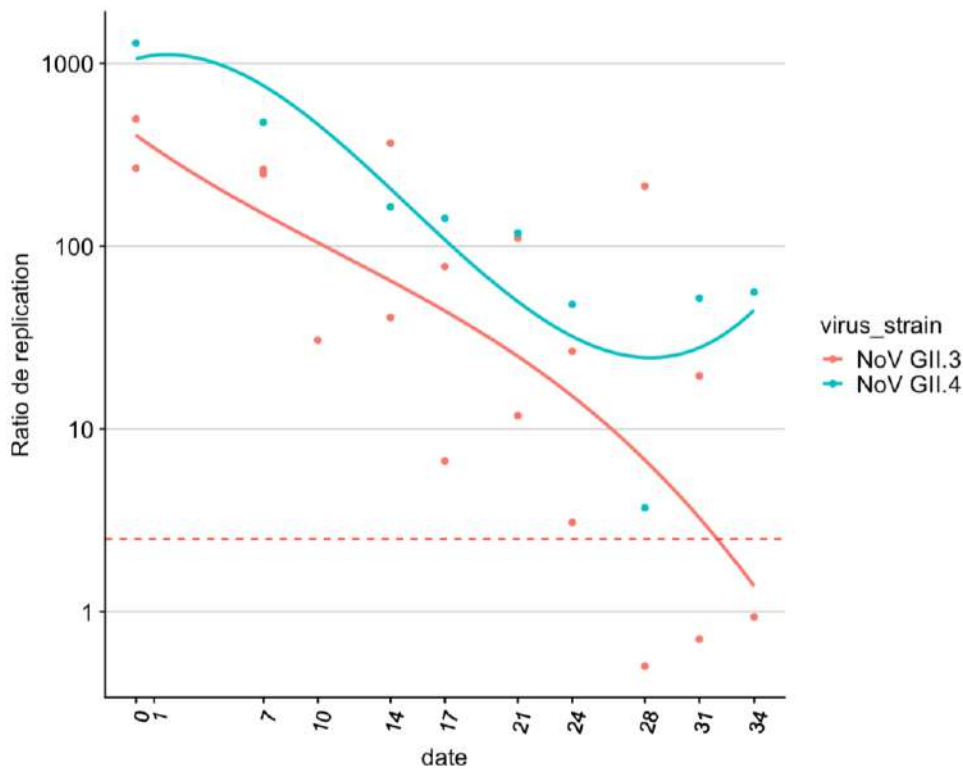


Figure 47 Ratio de répliation des souches NoVs GII.3 et GII.4 au cours du temps. Chaque point représente un échantillon prélevé lors de la première ou de la deuxième expérience, et la ligne pleine représente le modèle de régression linéaire. La ligne pointillée indique la valeur seuil d'infectiosité (fixé à 3).

Pour le TV, la méthode TCID50 permet de quantifier le titre viral infectieux, exprimé en TCID50/mL (Figure 48). La décroissance est assez rapide avec un titre infectieux atteignant le seuil de détection en deux à trois semaines.

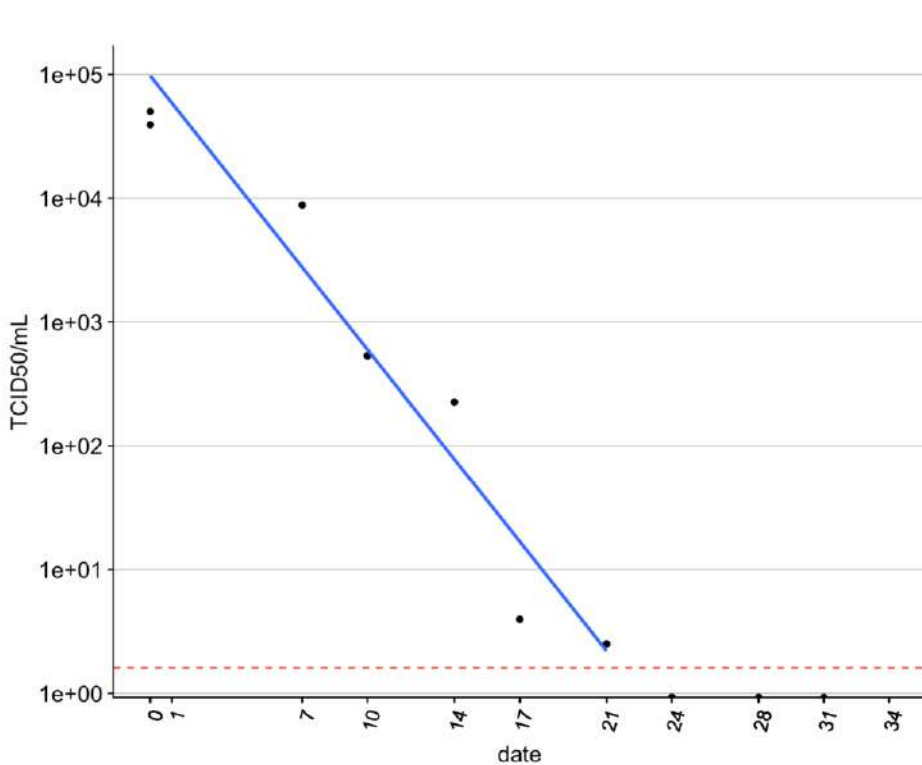


Figure 48 Décroissance du titre infectieux de TV indiquée par TCID50/mL. Chaque point représente un échantillon prélevé lors de la première ou de la deuxième expérience, et la ligne pleine représente la droite de régression linéaire. La ligne pointillée indique la valeur seuil de détection de l'infectiosité (fixé à 5 TCID₅₀/mL).

DISCUSSION

Cette étude constitue une première approche pour évaluer la persistance du pouvoir infectieux des NoVs dans l'eau de mer. Cette approche est novatrice et aucune donnée n'a jamais été publiée à l'heure actuelle. Basé sur nos expériences précédentes nous avons utilisé en parallèle le TV, modèle pertinent biologiquement, déjà validé dans les huîtres, et facile d'utilisation. La diminution du titre infectieux du TV est comparable aux données précédentes obtenues dans les tissus digestifs des huîtres (Polo et al. 2018). Les deux souches de NoVs sont restées infectieuses plus longtemps ce qui confirme la grande résistance de ces virus dans l'environnement. D'autres expériences sont nécessaires pour confirmer ces résultats préliminaires, et d'autres souches de NoVs doivent être évaluées. Le point limitant pour réaliser ces expériences est la disponibilité d'échantillons de selle contenant du NoVs infectieux. L'interprétation par rapport au TV doit considérer que les deux systèmes cellulaires sont différents, et également que le TV utilisé pour inoculer l'eau de mer est issu de culture cellulaire alors que les NoVs sont issus de selles humaines, ayant potentiellement un impact sur la stabilité du virus. Cette fragilité expliquerait la récupération virale moins efficace pour TV suite au protocole de purification et la perte rapide observée lors de la 1^{ère} semaine d'expérience et qui continue toute le long de l'étude, phénomène que nous avons déjà observé lors des travaux précédents (Polo *et al.*, 2018).

ARTICLE I

1.1 INTRODUCTION A L'ETUDE

Les virus entériques présents dans les eaux usées sont excrétés par les individus infectés. Les traitements des eaux usées n'éliminant pas la totalité des particules virales, ceux-ci se retrouvent dans l'environnement où ils peuvent alors contribuer à l'infection de nouveaux individus via la consommation d'eaux ou d'aliments contaminés. La variabilité des virus entériques est importante, plus de 100 espèces différentes ont été identifiées dans les eaux usées (Fernandez-Cassi *et al.*, 2018 ; Hoque *et al.*, 2019 ; Gerba *et al.*, 2018). L'étude des souches virales dans les eaux usées est de particulière importance car elle reflète les souches en circulation dans la population humaine. Ces données peuvent avoir un rôle dans le développement de nouvelles stratégies de préventions de la propagation virale. Cependant, l'approche de détection et quantification par biologie moléculaire nécessitant l'utilisation d'amorces sélectives restreint l'identification aux seuls virus déjà connus et soupçonné être présents dans l'échantillon. Le séquençage haut débit permet une analyse plus large identifiant tous les virus présents dans un échantillon sans *a priori* (Forbes *et al.*, 2017). La fragmentation des génomes d'abord et l'assemblage des fragments (reads) en contigs ensuite, sont des étapes clés du séquençage haut débit. Leur succès est inévitablement lié à la concentration virale initiale dans l'échantillon et au nombre de reads correspondantes produites lors du séquençage. L'obtention de longues séquences est nécessaire pour garantir la spécificité de l'identification considérant le nombre très important de séquences produites par cette technique.

Par exemple pour NoVs l'identification des zones spécifiques du génome (ORF1 et ORF2) est primordiale pour le typage des souches détectées.

De plus, seul un nombre de reads important peut garantir une bonne couverture des séquences assemblées, et permettra de différencier une souche variante d'une erreur ponctuelle lors du séquençage.

Toutefois l'abondance de microorganismes présents dans les eaux usées ou dans les échantillons fécaux est très importante incluant une grande variété de virus, bactéries, parasites, bactériophages pas seulement humains mais aussi des plantes et des insectes. Lors du séquençage haut débit, ceci constituera le *bruit de fond*, communément appelé le *background* utilisant un nombre parfois très important de reads et diminuant la possibilité de séquencer les microorganismes d'intérêt. Dans la métagénomique environnementale ciblée à la détection des virus humains une approche spécifique est donc indispensable afin de se débarrasser de la majorité des autres microorganismes pendant les étapes d'élution et d'extraction des acides nucléiques. Les caractéristiques physico-chimiques des particules virales peuvent alors être utilisées pour sélectionner les virus d'intérêt, en utilisant par exemple des caractéristiques connues de la capsid virale, sa taille ou le point isoélectrique.

La présence de sites de liaison spécifiques peut aussi être utilisée comme méthode de sélection et d'enrichissement. Les NoV ainsi que des autres virus entériques humains tels que les rotavirus interagissent avec la famille des glycoprotéines de type mucine par des sites de liaison spécifiques présents sur la capsid (Le Pendu *et al.*, 2014 ; Hu *et al.*, 2018). La capacité des mucines gastriques porcines (PGM) à lier la capsid des NoV a ainsi été précédemment étudiée avec un approche immuno-enzymatique (ELISA) et de PCR *in situ* (Tian *et al.*, 2008 ; Li *et al.*, 2015 ; Dancho *et al.*, 2012).

L'application des méthodes de séquençage haut débit peut aujourd'hui apporter des informations complémentaires à la détection par PCR.

1.2 OBJECTIF DE L'ETUDE

L'objectif de cette étude était d'évaluer trois différentes approches d'élution et d'enrichissement des virus entériques à partir d'échantillons d'eaux usées et de selles pour optimiser l'analyse par métagénomique. Une attention particulière a été donnée à l'identification des souches de NoVs, souvent détectées dans les eaux usées et première cause de gastroentérite aigüe non bactérienne chez l'Homme (Sano *et al.*, 2016 ; Schaeffer *et al.*, 2018 ; de Graaf *et al.*, 2017). Afin d'enrichir nos bibliothèques en séquences de NoVs nous avons sélectionné trois protocoles basés sur les propriétés connues de ces virus. La première méthode est basée sur l'utilisation d'une puce d'enrichissement (SureSelecte) développée spécifiquement pour cibler les NoV après extraction des acides nucléiques. La deuxième méthode est uniquement basée sur la résistance au pH de la capsid virale, en utilisant des variations de pH (acide puis basique) favorisant la précipitation et la récupération des particules virales. La troisième méthode est basée sur la présence de sites de liaison aux glycanes sur la capsid des NoVs. En raison de la forte homologie des mucines entre l'Homme et le porc, les PGM ont été utilisées pour sélectionner les NoVs pendant l'étape d'élution.



Characterization of Norovirus and Other Human Enteric Viruses in Sewage and Stool Samples Through Next-Generation Sequencing

Sofia Strubbia¹ · My V. T. Phan² · Julien Schaeffer¹ · Marion Koopmans² · Matthew Cotten^{2,3,4,5} · Françoise S. Le Guyader¹

Received: 5 July 2019 / Accepted: 17 August 2019 / Published online: 24 August 2019
© The Author(s) 2019

Abstract

This study aimed to optimize a method to identify human enteric viruses in sewage and stool samples using random primed next-generation sequencing. We tested three methods, two employed virus enrichment based on the binding properties of the viral capsid using pig-mucin capture or by selecting viral RNA prior to library preparation through a capture using the SureSelect target enrichment. The third method was based on a non-specific biophysical precipitation with polyethylene glycol. Full genomes of a number of common human enteric viruses including norovirus, rotavirus, husavirus, enterovirus and astrovirus were obtained. In stool samples full norovirus genome were detected as well as partial enterovirus genome. A variety of norovirus sequences was detected in sewage samples, with genogroup II being more prevalent. Interestingly, the pig-mucin capture enhanced not only the recovery of norovirus and rotavirus but also recovery of astrovirus, sapovirus and husavirus. Documenting sewage virome using these methods provides information for molecular epidemiology and may be useful in developing strategies to prevent further spread of viruses.

Keywords Human enteric viruses · Norovirus · Sewage · Metagenomic · Virome

Introduction

Increasing human population leads to increased demand for agricultural products and water, wastewater re-use will be necessary, which will pose a risk for virus pollution of the environment and subsequent affects viral transmission (Sano et al. 2016). More than 100 species of enteric viruses have been identified in human feces and in sewage such as astroviruses, caliciviruses, enteroviruses, enteric adenoviruses, and rotaviruses (Fernandez-Cassi et al. 2018; Hoque et al. 2019;

Metcalf et al. 1995; Gerba et al. 2018). Documenting viral prevalence and diversity in sewage may be a useful method for monitoring viruses circulating in the community (Smith et al. 2016). However, their detection by PCR approaches requires previously identified sequences for primer design and will not detect more distant viruses.

Metagenomics refer to the description of all nucleic acids sequences present in a sample (Forbes et al. 2017). Most of human enteric viruses have small RNA genomes making their detection difficult. Indeed, their relative abundance if compared to bacteria, phage, parasites present in the environment challenge their detection and identification (Cotten and Koopmans 2016; Nieuwenhuijse and Koopmans, 2017; Adriaenssens et al. 2018). Thus, viral metagenomics is generally performed by removing as much host and bacteria as possible followed by nuclease treatment to remove free nucleic acids (Kim et al. 2017; Nieuwenhuijse et al. 2017). Detection of viruses can be performed after de novo assembly of short-read data into longer sequences (contigs) followed by a variety of computational methods for detecting known and novel viral sequences (Cotten et al. 2014; Cotten et al. 2016; Oakeson et al. 2017). However, each sample type is unique and methods need to be adapted to account for

✉ Françoise S. Le Guyader
sleguyad@ifremer.fr

¹ Ifremer, Laboratoire de Microbiologie, LSEM-SG2M, BP 21105, 44311 Nantes Cedex 3, France

² Department of Viroscience, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands

³ Present Address: London School of Hygiene and Tropical Medicine, London, UK

⁴ Present Address: Uganda Virus Research Institute, Entebbe, Uganda

⁵ Present Address: MRC-Centre for Virus Research, Glasgow, UK

origin and to address the metagenomic objective. A concrete example was provided by the analysis of samples collected during the Tara oceans expedition (Alberti et al. 2017). The strategy applied allowed generation of data from a variety of organisms, including viruses and plankton from oceanic samples collected worldwide. Their approach showed the importance of using separate processing steps to analyze the different compartments of a volume of (sea)water. To identify human enteric viruses in complex samples such as stool or sewage samples, a method able to select these small particles resistant to acidic conditions, to eliminate bacteria and to decrease as much as possible background genomes such as phage genomes will be useful.

The aim of the current study was to evaluate methods for human enteric virus detection using metagenomics with a focus on norovirus. Noroviruses have great genomic diversity and they are divided into seven genogroups and many genotypes based upon genomic sequence phylogeny (de Graaf et al. 2017). Three of these genogroups (GI, GII and GIV) infect humans, and constitute the principal agent of acute gastroenteritis worldwide. Importantly for environmental research norovirus are excreted at high concentrations by infected individuals and they are highly persistent (Atmar et al. 2018). Their concentrations in sewage, that may vary among countries, are usually high and their presence in waters constitutes a major public health issue nowadays (Sano et al. 2016; Schaeffer et al. 2018). To specifically enrich our metagenomics libraries in norovirus sequences, we selected three protocols that take advantage of known properties of these viruses. The first method uses SureSelect target enrichment (Agilent) with probes matching human norovirus sequences. The second method comprises pH variations, based on norovirus capsids isoelectric point and on their resistance to both high and low pH. The third method is based on norovirus binding to glycans structures that are present in human but also pig mucins.

Materials and Methods

Samples

Five human stool samples (sample 570, 287, 5, 581, 582) positive for norovirus collected between January 2008 and October 2016 were used as a 10% suspension in water. Four samples were analyzed following method B and one sample following method B and C (stool sample 570) (Fig. 1). Seven raw sewage samples were collected from a gathering point upstream of any process between January 2014 and February 2017 from different sewage treatment plants located in the southern part of Brittany (France). Three sewage samples (sample 1777, 1797, 1854) were analyzed using method A

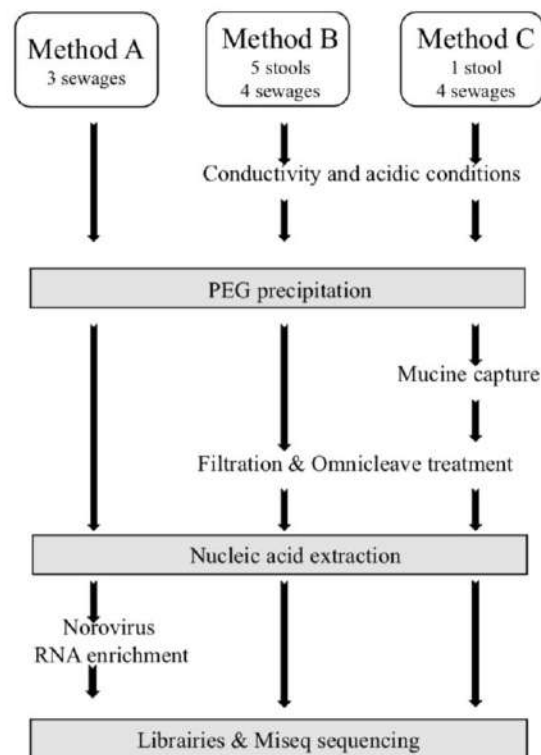


Fig. 1 Schematic of sample processing and sequencing used in this study. The three main frames on the top contain the method names (A, B and C), the number and the type of sample treated with each protocol. Key steps common to all protocols are represented in gray rectangle

and the remaining four samples (sample 1887, 1919, 1920, 1927) were analyzed using method B and C.

Method A: Polyethylene Glycol Precipitation and Norovirus SureSelect Target

Sewage samples were concentrated using a polyethylene glycol precipitation method (PEG) as previously described (Lewis et al. 1988; Schaeffer et al. 2018). Briefly, 40 mL of sample were mixed with 10 mL of a 50% PEG 6000 solution (Sigma-Aldrich, St Quentin France) (Fig. 1). After gentle agitation overnight at 4 °C, the mixture was centrifuged for 1.5 h at 13,500 × g and the pellet was suspended in 3 mL of 0.05 M glycine buffer (pH 9). After nucleic acid extraction as described below, norovirus RNA sequence were enriched using the SureSelect target enrichment specific array during the library preparation (details in the library preparation paragraph) (Brown et al., 2016).

Method B: Acidic Treatment and PEG Precipitation

To optimize human enteric viruses recovery, viruses were acid adsorbed to organic or inorganic particles present in the samples by increasing the conductivity of stool suspension or sewage samples to 2000 μS by addition of 5 M sodium chloride at pH 3 (Mullendore et al. 2001), and then concentrated using PEG precipitation as described above (Fig. 1). After centrifugation, the pellet was suspended in 3 mL of 0.05 M glycine buffer (pH 9), and filtered through 5, 1.2, 0.45 and 0.22 μm filters (Minisart NML 17594, NML17593, PES16533, PES16532). Then, the filtrate was incubated with 2000 Units of OmniCleave Endonuclease (Epicentre, Madison, USA) for one hour at 37 °C to eliminate free nucleic acids, followed by nucleic acid extraction.

Method C: Porcine Gastric Mucin (PGM) Capture

Samples were prepared following method B and after incubation with the OmniCleave Endonuclease, porcine gastric mucin (PGM) capture was applied (Fig. 1). Type III PGM (7.5 mg/mL, Sigma, France) was conjugated to MagnaBind™ carboxyl-derivatized beads (Pierce Biotechnology, Rockford, IL, USA) according to the manufacturer's protocol. Subsequently, 100 μL of PGM-bead suspension was mixed for 1 hour at room temperature with 3 mL glycine buffer PEG sewage concentrate (Tian et al. 2008). After capturing the beads using a magnetic rack, the supernatant was discarded and the beads were released and suspended in 1 mL of sterile water. This volume was further diluted in the lysis buffer for nucleic acid extraction as described below.

Nucleic Acid Extraction

Nucleic acids were extracted from sample concentrates prepared using the three methods by adding 10 mL of the chaotropic agent guanidine thiocyanate reagent lysis buffer (bioMerieux, Lyon, France) and incubation for 10 min at room temperature. Then, for samples prepared using Method C, the supernatants were transferred to new tubes after bead capture using the magnetic rack. Then, 140 μL of paramagnetic silica bead suspension was added (NucliSens kit, bioMerieux) to all the tubes and further incubated for 10 min at room temperature. The beads were captured using the magnetic rack and the volume was reduced to 2 mL for further extraction and purification steps as recommended (Schaeffer et al. 2018). A final step of RNA cleaning and concentration step was performed using a Zymo-spin column (RNA Clean & Concentrator, Zymo Research, Irvine, USA). The final step was a DNase treatment for 30 min

at 37 °C with 5 Units of Turbo DNase, (Ambion, ThermoFisher Scientific, France) (method B and C).

Norovirus Quantification

A microfluidic-based digital one-step RT-PCR, that allow quantification without external calibration curves was performed using primers and probes targeting the ORF1-2 region (Polo et al. 2016). Positive and negative controls were included in each series, and quantification was calculated using the Poisson distribution (QuantStudio™ 3D Analysis Suite™ Cloud Software, version 3.0.3; ThermoFisher). The final result was expressed as RNA copies/ μL .

Library and Sequencing

For method A, cDNA were synthesized using Superscripts II and random primers according to the manufacturer protocol (Life Technologies). The second-strand cDNA synthesis was performed using 5 U of Klenow (Invitrogen) polymerase in a final volume of 30 μL followed by SureSelect enrichment (Agilent) with the RNA bait design previously designed (Brown et al. 2016). NGS libraries were prepared using a SureSelect^{XT} Illumina paired-end sequencing library protocol (Agilent).

For methods B and C, after cDNA synthesis using SuperScript II and random primers according to the manufacturer's protocol, the second-strand DNA was synthesized according to the manufacturer's protocol (New England Biolabs). NGS libraries were prepared using the NEB Next Ultra DNA Library Prep Kit for Illumina (New England Biolabs) according to the manufacturer's instructions.

Sequencing was performed on the Illumina MiSeq platform, with Phage *PhiX174* added to samples to standardize the runs.

Sequence Analysis

Illumina adapters were removed from the raw reads and resulting reads were trimmed using QUASR (Watson et al. 2013) from the 3' end to reach a median Phred score ≥ 35 , which means a base call accuracy between 99.9% and 99.99%. Reads shorter than 80% of the original read length were discarded.

General De Novo Assembly

De novo assembly was performed with quality-controlled reads using SPAdes v.3.10.1 (Bankevich et al. 2012). A variety of assembly conditions were examined but in general assembly with no error correction or read normalization yielded the largest initial contig set. Virus specific contigs were identified using Usearch (Edgar 2010), against a set of

39 virus family or subfamily specific protein databases and virus family specific contigs were further assembled into larger contigs using mapping of the contigs against the closest identified full genome. A penultimate consensus genome was generated from the contigs. A final check of the genome was performed by mapping all quality-controlled reads to the penultimate consensus genome and a final majority nucleotide consensus genome was generated. All expected reading frames were examined, any disruption was checked and resolved by consulting the original reads across the query site.

Virus Family Specific De Novo Assembly

All virus family-specific reads were then harvested by mapping to a comprehensive set of all sequences > 500 nt available for that family in GenBank using Bowtie2 (Langmead et al. 2012). The resulting virus family-specific reads were de novo assembled using SPAdes v.3.10.1 (Bankevich et al. 2012). Contigs shorter than 500nt were removed from subsequent analysis and contigs with coverage below 10 (determined by Bowtie2 mapping) were carefully examined to avoid assembling contigs with varying coverage. Further assembly and genome checking were performed as described above.

Results

One objective of this work was to explore three biochemical and nucleic acid enrichment methods in sample preparation for viral NGS (Fig. 1). Overall, large contigs matched to one of these eight viral families (*Astroviridae*, *Caliciviridae*, *Nodaviridae*, *Leviviridae*, *Microviridae*, *Picornaviridae*, *Picobirnaviridae* and *Reoviridae*) (Fig. 2). These families were detected in multiple samples and we focused on these families for the remaining analyses. No sequence with homology to the Hepatovirus genus or *Hepeviridae* family were detected.

Norovirus Concentrations in Sewage Samples

Quantification performed on four samples displayed comparable concentration (around 10^3 RNA copies (c)/ μ L) for norovirus (Table 1). Norovirus GI was detected in all samples, with concentrations 10 to 100 times lower compare to norovirus GII concentrations. After PGM capture (method C) norovirus concentrations was similar for three samples, but was lower for sample 1920, both for genogroup I and II (Table 1).

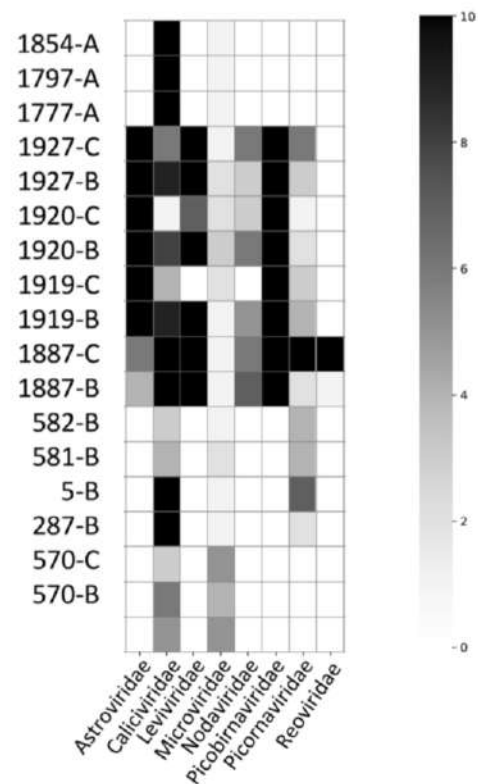


Fig. 2 Heat map of larger viral contig yield. Quality-controlled short-read data were de novo assembled into large contigs and contigs were classified at the protein level by comparison with virus family-specific reference sets (see Methods). The numbers of contigs in each sample with >60% protein identity and 500 nt minimum contig length were plotted in a heat map. The colorbar to the right indicates number of contigs detected per sample in each virus family

Utility of Target Enrichment (Method A) to Study Norovirus Diversity

Targeting norovirus sequences using the norovirus enrichment capture allowed the recovery of long fragments with good coverage as the lowest count was 8.225 (Table 2). Almost complete genomes were obtained for seven GII strains and for two GI strains. Half of a GIV.1 norovirus sequence was also obtained. Another advantage of this approach is the sensitivity as the full-genome sequence for GI strain was obtained from sewage samples 1777 and 1854, having both a quite low concentration of 3 RNAc/ μ L.

Table 1 Norovirus concentrations in wastewater samples used for this study and corresponding calicivirus reads

Sample	Method	Norovirus*		Calicivirus reads
		GI	GII	
E1777	A	3	497	NA
E1797	A	64	39	NA
E1854	A	3	37	NA
E1887	B	7	219	2507
	C	2	223	13,548
E1919	B	19	361	3242
	C	6	346	10,635
E1920	B	20	466	10,406
	C	3	81	3962
E1927	B	26	140	1750
	C	11	637	2777

NA data not collected

*Concentrations obtained using digital RT-PCR are expressed as RNA copies/ μ L for genogroup I (GI) and II (GII)

**Number of reads mapped on contigs > 500nt using Bowtie2

Table 2 Norovirus strains identified in sewage samples using method A

Sample	Contig length	Genotype*	Coverage**
E1777	7278	GIIP21-GII.3	19.586
	7484	GIIP7-GII.6	99.576
	6828	GIIPe-GII.4syd-v2	23.144
	7497	GIIPe-GII.4syd	21.732
	7648	GIP3-GI.3	16.817
E1797	7343	GIIP17-GII.7	85.503
	7431	GIIP7-GII.6	47.253
	3467	GIP9-GI.9	12.576
	7679	GIPb-GI.6	71.396
E1854	1582	GIIP16-GII.13	12.72
	6509	GIIP17-GII.17	8.225
	1221	GIIPg	8.663
	7641	GIPb-GI.6	15.532
	3687	GIV	12.327

*Sequences were identified using the online norovirus genotyping tool v2.0 (Kroneman et al. 2011)

**Coverage was calculated using Bowtie2 (Langmead et al. 2012)

Impact of PEG Precipitation (Method B) on Virus Sequence Yield

Viruses were first concentrated using high-molecular weight PEG precipitation for both stool and sewage samples. PEG has been used to concentrate enteric virus in oysters and water samples (Lewis et al. 1988) and when applied here, it was efficient in recovering long norovirus contigs, with

Table 3 Norovirus and other human virus detected in stool samples using method B

Sample	Virus	Contig length	Genotype
S570	Norovirus	7609	GIIPe-GII.4 syd
S287	Norovirus	7542	GIIP21-GII.3
	Norovirus	1104	GIIP17-GII.17
S5	Enterovirus	653	coxsackievirus A9
	Norovirus	7390	GIIP4NewOrleans-GII.4syd
	Norovirus	3399	GII.P21-GII.3
	Enterovirus	1387	coxsackievirus A9
S581	Norovirus	7528	GIIP17-GII.17
	Enterovirus	1421	coxsackievirus A9
S582	Norovirus	7536	GIIP17-GII.17
	Enterovirus	1363	coxsackievirus A9

five complete norovirus genomes obtained. An additional advantage of the PEG approach when combined with random primed deep sequencing was the detection of other enteric viruses present in the samples such as coxsackievirus in stool samples (Table 3), sapovirus, astrovirus and rotavirus in sewage samples (Table 4).

Impact of Mucin Capture (Method C) on Virus Sequence Yield

We aimed to investigate if a mucin capture step could provide enrichment of norovirus and other enteric virus materials. Calicivirus read numbers increased after mucin capture except for sample 1920 for which norovirus GI and GII concentrations were also lower suggesting a failure in one purification step or a selection of some norovirus strains by the PGM (Table 1). No reproducible impact of the mucin capture on contig lengths was observed. Examining the lengths of norovirus contigs as a measure of success, four norovirus contigs were longer with mucine capture, while six norovirus contigs showed reduced length with mucin capture (compared to contig yield without PGM capture, Method B) (Table 4). An unexpected observation was the impact of this treatment on a number of other viruses as shown by the viral family-specific read yields for eight virus families. For *Astroviridae*, *Reoviridae*, *Nodaviridae* and *Picornaviridae* families, the yield of specific reads (as a percentage of total reads for that sample) was increased, as reported in upper panels of Fig. 3 comparing orange markers (method B) to blue markers (method C). This was confirmed by the identification of a complete genome for a human rotavirus A genotype G9-[P8], and for astrovirus (full genome for one strain and four additional strains identified) compared to method B (Table 4). In contrast, the yields of *Leviviridae* and *Picornaviridae* families remained similar or slightly reduced. A large amount of the phage *PhiX174* (a member

Table 4 Human enteric viruses detected using method B and C

Sample	Method B			Method C			
	Virus	Contig	Genotype	Virus	Contig	Genotype*	
S570	Norovirus	7609	GIIPe-GII.4	Norovirus	6976	GIIPe-GII.4	
E1887	Sapovirus	7414	GI.2	Sapovirus	7491	GI.2	
	Norovirus	3241	GIIP22-GII.17	Norovirus	2366	GII.P22	
	Norovirus	2961	GII	Norovirus	2129	GII	
	Norovirus	2996	GII.P7-GII.17	Norovirus	3126	GII.P17-GII.17	
				Norovirus	3024	GII.P17-GII.6	
		Rotavirus	836		Rotavirus	3283	G9[P8]
	Astrovirus	3860	Type 2	Astrovirus	6519	Type 2	
E1919				Husavirus	494	Sp.16915_89	
	Norovirus	4598	GII.P16-GII.2	Norovirus	1028	GIIP16	
	Norovirus	2692	GII	Norovirus	3004	GII.P16-GII.4 syd	
	Sapovirus	1209	GII.3				
	Astrovirus	6810	Type 1	Astrovirus	6412	Type 1	
E1920	Norovirus	2239	GII	Norovirus	510	GII.2	
	Norovirus	332	GI–GI.3				
	Astrovirus	6782	Type 1	Astrovirus	1878	Type 1	
E1927	Norovirus	944	GII.2	Norovirus	631	GII.Postdam	
				Norovirus	616	GII	
				Norovirus	631	GI–GI.3	
				Sapovirus	571	GI.1 Seoul	
		Astrovirus	1878	Type 1	Astrovirus	6588	Type 5
				Astrovirus	3948	Type 1	
			Aichi virus	2692	Type 1		

*Strain identification was assigned using the online Norovirus genotyping tool v2.0 (Kroneman et al. 2011)

of the *Microviridae* family) was added as a carrier after the PGM capture, thus we cannot make conclusion about the impact of PGM on the *Microviridae* family.

Discussion

This work aimed to generate enteric virus sequence data from human stool and sewage samples, with a special interest in norovirus. There are several challenges when applying metagenomics deep sequencing to describe the total virome of sewage samples (Bibby et al. 2013; Adriaenssens et al. 2018). One challenge is the high levels of dilution of human enteric viruses and the presence of a large variety of other microorganisms, plant, algae, chemical and organic compounds. Indeed, in non-outbreak settings, human enteric viruses are present at lower concentrations compared to bacteria or phages. In a previous study, we demonstrated that at least 1–4% of the population need to be infected to detect hepatitis E virus in raw sewage from a small sewage treatment plant, showing the potential to rapidly detect outbreak in the population (Miura et al. 2016). Another challenge of sewage virome characterization is to obtain a representative sample of

the viral load and diversity from the wastewater. Concentration of large volumes of water are currently used to decrease the sampling variability and increase the sensitivity threshold of the assays (Lewis et al. 1988; Schaeffer et al. 2018; Fernandez-Cassi et al. 2018). However, this may lead to increase concentration of inhibitors of the enzymes used for molecular assays, as human sewage may contain detergents, medicine, food additives, food waste and other chemicals and thus some purification steps are needed (Hata et al. 2017). Composite samples obtained over 24 h, as used for this study, may be a convenient way to overcome some of these problems as it can be representative of what is entering in the sewage treatment plant without the need of large volumes. Different methods previously developed for sewage, water or shellfish analysis use the PEG precipitation that help to concentrate biomolecules by altering their hydration from solution and viruses (Lee et al. 1981; Lewis et al. 1988; Metcalf et al. 1995; Jiang et al. 1992). This easy-to-use method was proposed for the first assay able to detect norovirus in stool (Jiang et al. 1992) and in preliminary test we verified that it helped to recover the complete norovirus genome from stool samples (data not shown). As human enteric viruses tend to aggregate or to bind to different types of particles,

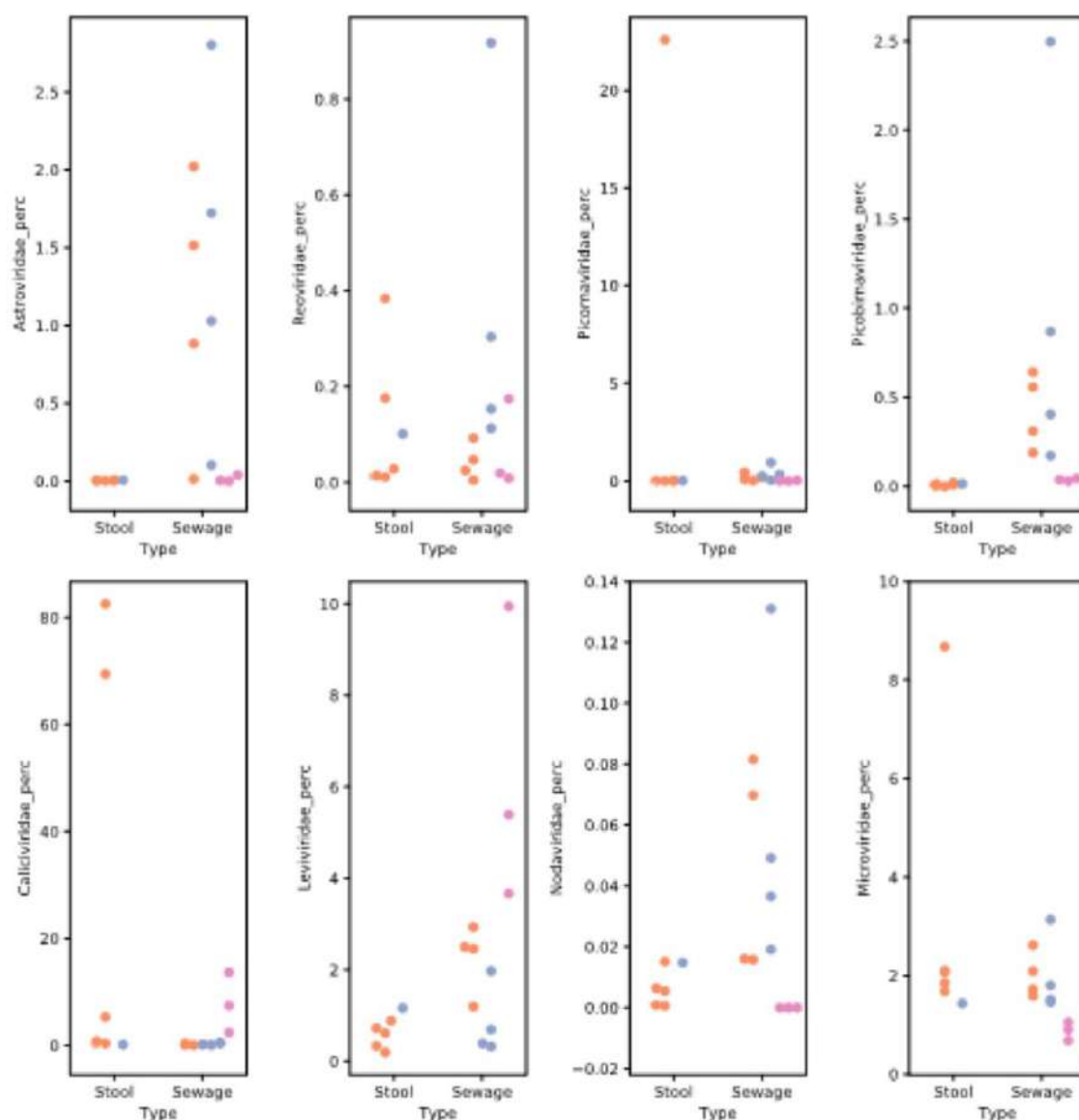


Fig. 3 Scatter plot of reads identified for each of 8 virus families (percentage of total reads). Quality-controlled reads for each of the 18 sample/conditions were mapped to virus family reference sets (see Methods) for *Astroviridae*, *Caliciviridae*, *Nodaviridae*, *Leviviridae*, *Microviridae*, *Picornaviridae*, *Picobirnaviridae* and *Reoviridae*. The

number of reads identified in each family was plotted as percentage of the total quality-controlled reads for that read set. Data sets were separated into sample types (stool or sewage) and colored according sample treatment (method A: pink, method B: orange and method C: blue)

including bacteria, we applied an elution step based on pH variations under controlled conductivity conditions (Miura et al. 2013; Samandoulgou et al. 2015; Mullendore et al. 2001; da Silva et al. 2011). This added step may explain the higher efficacy to obtain long sequence of

human enteric viruses compared to a published study that failed to identified norovirus strains (Hjelmsø et al. 2017).

Capsid structures of a number of human enteric viruses including norovirus and rotavirus have been found to interact with the mucin family of glycoproteins due to the presence

of binding sites for the complex carbohydrates on mucin (Le Pendu et al. 2014). This binding affinity was proposed to select norovirus particles from water samples using human histo-blood group antigen or PGM that present similar structure as it may increase the specificity by capturing non-damaged capsid and the sensitivity of detection by facilitating inhibitor removal (Zhou et al. 2017; Tian et al. 2008). In this study, PGM capture increased norovirus concentrations in the nucleic acid extracts, however, no difference was observed in terms of contig lengths or diversity of identified sequences. Combining PGM and HBGA type-B antigen could have helped to improve this approach (Tian et al., 2017). One beneficial impact of the PGM capture was the detection of full genome of a rotavirus genogroup A genotype G9-[P8], confirming the binding affinity of rotavirus strains to glycans (Hu et al. 2018). When compared to method B (PEG without mucin capture), rotavirus was identified but full genome was not obtained, suggesting that PGM combined with random primed deep sequencing would be superior to detect full genomes of norovirus and rotavirus. It was more surprising to find a beneficial impact of PGM capture on astrovirus, with an increased diversity of strains identified. The PGM capture led also to the characterization of a few sequence of husavirus strain. Some husaviruses have been described as a novel virus family within the order Picornavirales, and are common in human fecal material (Oude Munnink et al. 2015). However, very little is known about astrovirus or husavirus PGM interactions, which may be due to non-specific interactions such as sialic acid recognition or other glycan affinity to viral capsid structure (Shanker et al. 2017; Hu et al. 2018). Adding purification steps enhance the risk to lose viruses as observed for one sample and also to introduce some bias of selection. Furthermore, when dealing with complex samples with a mixture of low amounts of viral contaminations such as sewage, this may be an issue which needs to be further investigated. Three biological replicates were found to be a useful approach to minimize potential biases and to give more confidence in analyzing aquatic viromes (Kim et al. 2017).

One objective of this work was to evaluate norovirus diversity present in stool and sewage samples. Some works based on a metabarcoding approach targeting the ORF2 portion of norovirus genome were successfully applied to sewage samples, but this approach limits the identification of strains that are not amplified by published primers (Oshiki et al. 2018; Fumian et al. 2019). In our work, we aimed to identify long fragments to be able to capture the diversity of strains circulating in the local population. The SureSelect target enrichment (method A) gave the highest number of full genome for several norovirus strains despite similar norovirus concentrations when compared with sewage samples used (Brown et al. 2016). This method was the only one able to yield norovirus GI strain identifications, important criteria

for environmental sample analysis considering the importance of these strains transmission through the environment (Le Guyader et al. 2012; Verhoef et al. 2015). Although the SureSelect target enrichment array is designed to capture *Caliciviridae* sequences, these targeted sequences were still less than 1% of the total reads after capture, suggesting non-specific binding. Since the binding and wash conditions are proprietary to the manufacturer and the actual bait concentrations are not provided, specificity improvements are limited. This method also allowed to identify a norovirus GIV.1 strain. This genogroup is sporadically detected in sewage samples, and may be under-appreciated as a cause of gastroenteritis, presumably because only a small number of sequences are available in GenBank causing difficulties in primer design (Sima et al. 2011; La Rosa et al. 2012). The two other methods also allowed to characterize some norovirus sequence, all of them being already reported in the NoroNet network (van Beek et al. 2018). Beside the ubiquitous GII.4 strains that have been reported worldwide, it was interesting to detect the GII.P17-GII.17 strain that caused several gastroenteritis outbreaks in multiple countries during this sampling period (Matsushima et al. 2015; Koo et al. 2017). In this regard, a complete characterization of detected viruses is important to identify new strains and thus to help risk manager to take measure to prevent further distribution (Cocolin et al. 2018).

A limitation of this study lies in the small sample size tested. Ideally, it would be more informative to have a larger number of samples being tested for the same approach to evaluate the method efficiency across sample variability. Sequence contents of raw sewage reflect the composition of the microbiome of local population and the diversity of all pathogenic or non-pathogenic bacteria or viruses circulating in the community at a larger extent than individual samples (Newton et al. 2015; Sano et al. 2016). To access this sequence information, we need to develop reproducible, simple, fast and easy to apply methods. The viral agnostic metagenomics approach is still an expensive approach when considering the library preparation or sequencing runs, but all the steps of the methods described here can easily be performed in environmental laboratory already performing PCR analysis as it just needs a centrifuge and basic equipment.

Acknowledgements We acknowledge the help of the SAUR and SUEZ technicians for sampling. We are most grateful to the Genomics and Bioinformatics Core Facility of Nantes (GenoBird, Biogenouest) for the technical support.

Funding This work was funded by EU-H2020 Grant No. 643476 COMPARE. My V.T. Phan is supported by Marie Skłodowska-Curie Individual Fellowship, funded by European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 799417).

Data Availability The short-reads data for this study has been deposited in the European Nucleotide Archive (ENA) <http://www.ebi.ac.uk/ena/>

data/view/PRJEB31600 and to Genbank with the following (temporary) accession number: from MK789654 to MK789656 for NoV GI, from MK907785 to MK907802 for NoV GII, MK726262 for NoV GIV.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Adriaenssens, E. M., Farkas, K., Harrison, C. H., Jones, D. L., Allison, H. E., & McCarthy, A. J. (2018). Viromic analysis of wastewater input to a river catchment reveals a diverse assemblage of RNA viruses. *mSystems*, *3*, e00015–00018.
- Alberti, A., et al. (2017). Viral to metazoan marine plankton nucleotide sequences from the Tara Oceans expedition. *Scientific Data*, *4*, 170093.
- Atmar, R. L., Ramani, S., & Estes, M. K. (2018). Human noroviruses: Recent advances in a 50-year history. *Current Opinion in Infectious Disease*, *31*, 422–432.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., et al. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*, 455–477.
- Bibby, K., & Peccia, J. (2013). Identification of viral pathogen diversity in sewage sludge by metagenome analysis. *Environmental Science and Technology*, *47*, 1945–1951.
- Brown, J. R., Roy, S., Ruis, C., Romero, E. Y., Shah, D., Willimas, R., et al. (2016). Norovirus whole-genome sequencing by sure select target enrichment: A robust and sensitive method. *Journal of Virology*, *54*, 2530–2537.
- Cocolin, L., Mataragas, M., Bourdichon, F., Doulgeraki, A., Pilet, M.-F., Jagadeesan, B., et al. (2018). Next generation microbiological risk assessment meta-omics: The next need for integration. *International Journal of Food Microbiology*, *287*, 10–17.
- Cotten, M., & Koopmans, M. (2016). Next-generation sequencing and norovirus. *Future Virology*, *11*, 719–722.
- Cotten, M., Munnink, B. O., Canuti, M., Deijis, M., Watson, S. J., Kellam, P., et al. (2014). Full genome virus detection in fecal samples using sensitive nucleic acid preparation, deep sequencing, and a novel iterative sequence classification algorithm. *PLoS ONE*, *9*, e93269.
- da Silva, A., Kavanagh, O. V., Estes, M. K., & Elimelech, M. (2011). Adsorption and aggregation properties of norovirus GI and GII virus like particles demonstrate differing responses to solution chemistry. *Environmental Science and Technology*, *45*, 520–526.
- de Graaf, M., Villabruna, N., & Koopmans, M. P. G. (2017). Capturing norovirus transmission. *Current Opinion in Virology*, *22*, 64–70.
- Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, *26*, 2460–2461.
- Fernandez-Cassi, X., Timoneda, N., Partinez-Puchol, S., Rusinol, M., Rodriguez-Manzano, J., Figuerola, N., et al. (2018). Metagenomics for the study of viruses in urban sewage as a tool for public health surveillance. *Science of the Total Environment*, *618*, 870–880.
- Forbes, J. D., Knox, N. C., Ronholm, J., Pagotto, F., & Reimer, A. (2017). Metagenomics: The next culture-independent game changer. *Frontiers in Microbiology*, *8*, 1069.
- Fumian, T. M., Fioretti, J. M., Lun, J. H., dos Santos, I. A. L., White, P. A., & Miagostovich, M. P. (2019). Detection of norovirus epidemic genotypes in raw sewage using next generation sequencing. *Environment International*, *123*, 282–291.
- Gerba, C., Betancourt, W. Q., Kitajima, M., & Rock, C. M. (2018). Reducing uncertainty in estimating virus reduction by advanced water treatment process. *Water Research*, *133*, 282–288.
- Hata, A., Inaba, M., Katayama, H., & Furumai, H. (2017). Characterization of natural organic substances potentially hindering RT-PCR-based virus detection in large volumes of environmental water. *Environmental Science and Technology*, *51*, 13568–13579.
- Hjelmso, M. H., Hellmer, M., Fernandez-Cassi, X., Timoneda, N., Lukjancenko, O., Seidel, M., et al. (2017). Evaluation of methods for the concentration and extraction of viruses from sewage in the context of metagenomic sequencing. *PLoS ONE*, *12*, e0170199.
- Hoque, S. A., Thongprachum, A., Takanashi, S., Mostafa, S. M., Saito, H., Anwar, K. S., et al. (2019). Alarming situation of spreading enteric viruses through sewage water in Dhaka city: Molecular epidemiological evidences. *Food and Environmental Virology*, *11*, 65–75.
- Hu, L., Sankaran, B., Laucirica, D. R., Estes, M. K., & Prasad, B. V. V. (2018). Glycan recognition in globally dominant human rotaviruses. *Nature Communications*, *9*, 2631.
- Jiang, X., Wang, J., Graham, D. Y., & Estes, M. K. (1992). Detection of Norwalk virus in stool by polymerase chain reaction. *Journal of Clinical Microbiology*, *30*, 2529–2534.
- Kim, Y., van Bonn, W., Aw, T. G., & Rose, J. B. (2017). Aquarium viromes: Viromes of human-managed aquatic systems. *Frontiers in Microbiology*, *8*, 1231.
- Koo, E. S., Kim, M. S., Choi, Y. S., Park, K.-S., & Jeong, Y. S. (2017). Occurrence of novel GII.17 and GII.21 norovirus variants in the coastal environment of South Korea in 2015. *PLoS ONE*, *12*, e0172237.
- Kroneman, A., Vennema, H., Deforche, K., Avoort, H. v. d., Peñaranda, S., Oberste, M. S., et al. (2011). An automated genotyping tool for enteroviruses and noroviruses. *Journal of Clinical Virology*, *51*, 121–125.
- La Rosa, G., Fratini, M., Vennarucci, V. S., Guercio, A., Purpari, G., & Muscillo, M. (2012). GIV noroviruses and other enteric viruses in bivalves: A preliminary study. *New Microbiologica*, *35*, 27–34.
- Langmead, B., & Salzberg, S. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*, 357–359.
- Le Guyader, F. S., Atmar, R. L., & Le Pendu, J. (2012). Transmission of viruses through shellfish: When specific ligands come into play. *Current Opinion in Virology*, *2*, 103–110.
- Le Pendu, J., Nystrom, K., & Ruvoen-Clouet, N. (2014). Host pathogen co-evolution and glycan interactions. *Current Opinion in Virology*, *7*, 88–94.
- Lee, J. C., & Lee, L. L. Y. (1981). Preferential solvent interactions between proteins and polyethylene glycol. *Journal of Biology and Chemistry*, *256*, 625–631.
- Lewis, G. D., & Metcalf, T. G. (1988). Polyethylene glycol precipitation for recovery of pathogenic viruses, including Hepatitis A virus and human rotavirus, from oyster, water and sediment samples. *Applied and Environmental Microbiology*, *54*, 1983–1988.
- Matsushima, Y., Ishikawa, M., Shimizu, T., Komane, A., Kasuo, S., Shinohara, M., et al. (2015). Genetic analyses of GII.17 norovirus strains in diarrheal disease outbreaks from December 2014 to March 2015 in Japan reveal a novel polymerase sequence and amino acid substitution in the capsid region. *Euro Surveillance*, *20*, 21173.
- Metcalf, T., Melnick, J. L., & Estes, M. K. (1995). Environmental Microbiology: From detection of virus in sewage and water by isolation to identification by molecular biology—A trip of over 50 years. *Annual Review in Microbiology*, *49*, 461–487.

- Miura, T., Lhomme, S., Le Saux, J.-C., Le Mehaute, Ph, Guillois, Y., Couturier, E., et al. (2016). Detection of hepatitis E virus in sewage after an outbreak on a French island. *Food and Environmental Virology*, *8*, 194–199.
- Miura, T., Sano, D., Suenaga, A., Yoshimura, T., Fusawa, M., Nakagomi, T., et al. (2013). Histo-blood group antigen-like substances of human enteric bacteria as specific adsorbents for human noroviruses. *Journal of Virology*, *87*, 9441–9451.
- Mullendore, J., Sobsey, M. D., & Shieh, Y. S. C. (2001). Improved method for the recovery of hepatitis A virus from oysters. *Journal of Virological Methods*, *94*, 25–35.
- Newton, R., McLellan, S. L., Dila, D. K., Vincis, J. H., Morrison, H. G., Eren, A. M., et al. (2015). Sewage reflects the microbiomes of human populations. *mBio*, *6*, e02574–02584.
- Nieuwenhuijse, D., & Koopmans, M. P. G. (2017). Metagenomic sequencing for surveillance of food- and waterborne viral diseases. *Frontiers in Microbiology*, *8*, 230.
- Oakeson, K. F., Wagner, J. M., Mendenhall, M., Rohrwasser, A., & Atkinson-Dunn, R. (2017). Bioinformatic analyses of whole-genome sequence data in a Public Health Laboratory. *Emerging and Infectious Disease*, *23*, 1441–1445.
- Oshiki, M., Miura, T., Kazama, S., Segawa, T., Ishii, S., Hatamoto, M., et al. (2018). Microfluidic PCR amplification and MiSeq amplicon sequencing techniques for high-throughput detection and genotyping of human pathogenic RNA viruses in human feces, sewage, and oysters. *Frontiers in Microbiology*, *9*, 830.
- Oude Munnink, B. B., Cotten, M., Deijns, M., Jebbink, M. F., Bakker, M., Mohammad, S., et al. (2015). A novel genus in the order *Picornavirales* detected in human stool. *Journal of General Virology*, *96*, 3440–3443.
- Polo, D., Schaeffer, J., Fournet, N., Le Saux, J.-C., Parnaudeau, S., McLeod, C., et al. (2016). Digital PCR for quantifying norovirus in oysters implicated in outbreaks, France. *Emerging and Infectious Disease*, *22*, 2189–2191.
- Samandoulgou, I., Fliss, I., & Jean, J. (2015). Zeta potential and aggregation of virus-like particle of human norovirus and feline calicivirus under different physicochemical conditions. *Food and Environmental Virology*, *7*, 249–260.
- Sano, D., Amarasiri, M., Hata, A., Watanabe, T., & Katayama, H. (2016). Risk management of viral infectious diseases in wastewater reclamation. *Environment International*, *91*, 220–229.
- Schaeffer, J., Treguier, C., Piquet, J.-C., Cochenec-Laureau, N., Le Saux, J.-C., Garry, P., et al. (2018). Improving the efficacy of sewage treatment decrease norovirus contamination in oysters. *International Journal of Food Microbiology*, *286*, 1–5.
- Shanker, S., Hu, L., Ramani, S., Atmar, R. L., Estes, M. K., & Prasad, B. V. V. (2017). Structural features of glycan recognition among viral pathogens. *Current Opinion in Structural Biology*, *44*, 211–218.
- Sima, L. C., Schaeffer, J., Le Saux, J.-C., Parnaudeau, S., Elimelech, M., & Le Guyader, F. S. (2011). Calicivirus removal in a membrane bioreactor wastewater treatment plant. *Applied and Environmental Microbiology*, *77*, 5170–5177.
- Smith, D. B., Paddy, J. O., & Simmonds, P. (2016). The use of human sewage screening for community surveillance of hepatitis E virus in the UK. *Journal of Medical Virology*, *88*, 915–918.
- Tian, P., Engelbrekton, A., & Mandrell, R. E. (2008). Two-log increase in sensitivity for detection of norovirus in complex samples by concentration with porcine gastric mucin conjugated to magnetic beads. *Applied and Environmental Microbiology*, *74*, 4271–4276.
- Tian, P., Yang, D., Shan, L., Li, Q., Liu, D., & Wang, D. (2017). Estimation of Human norovirus infectivity from environmental water samples by in situ capture RT-qPCR method. *Food and Environmental Virology*, *10*, 29–38.
- van Beek, J., de Graaf, M., Al-Hello, H., Ambert-Balay, K., Botteldoorn, N., Brytting, M., et al. (2018). Molecular surveillance of norovirus, 2005–2016: An epidemiological analysis of data collected from the NoroNet network. *Lancet Infectious Disease*, *18*, 545–553.
- Verhoef, L., Hewitt, J., Barclay, L., Ahmed, S. M., Lake, R., Hall, A. J., et al. (2015). Norovirus genotype profiles associated with foodborne transmission, 1999–2012. *Emerging and Infectious Disease*, *21*, 592–599.
- Watson, S. J., Welkers, M. R. A., Depledge, D. P., Coulter, E., Breuer, J. M., de Jong, M. D., et al. (2013). Viral population analysis and minority-variant detection using short read next-generation sequencing. *Philosophical Transactions of the Royal Society B*, *368*, 20120205.
- Zhou, Z., Tian, Z., Li, Q., Tian, P., Wu, Q., Wang, D., et al. (2017). In situ capture RT-qPCR: A new simple and sensitive method to detect human norovirus in oysters. *Frontiers in Microbiology*, *8*, 554.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1.4 DISCUSSION

La détection par séquençage haut débit des virus entériques à partir d'échantillons d'eaux usées présente diverses difficultés. Tout d'abord les eaux sont souvent peu contaminées (en comparaison d'un échantillon clinique par exemple). Le prélèvement de larges volumes d'eaux est ainsi une stratégie souvent utilisée pour assurer une représentativité correcte de l'échantillonnage, augmenter la diversité des microorganismes et les chances de détecter le virus d'intérêt. Toutefois les techniques de concentrations utilisées à partir de larges volumes entraînent souvent également la concentration d'inhibiteurs des enzymes utilisés en biologie moléculaire. Détergents, additifs alimentaires, médicaments et autres composants chimiques présents dans les eaux usées peuvent interférer avec les réactions enzymatiques et empêcher l'analyse (Hata *et al.*, 2017). Des étapes de purification sont donc nécessaires afin de limiter ce phénomène. L'échantillonnage d'un volume réduit d'eaux usées mais collecté régulièrement pendant 24h est une stratégie tout aussi efficace. Enfin, la présence d'un grand nombre de microorganismes, de plantes, d'algues et de résidus chimiques et organiques compliquent l'analyse metagénomique.

Trois approches ont été utilisées pendant cette étude afin d'enrichir les échantillons en virus entériques en vue de l'analyse metagénomique.

La première approche (Méthode A) a été basée sur l'utilisation d'un système d'enrichissement sur puce (SureSelect), développée par Brown et collègues (2016) en sélectionnant un pool d'amorces spécifiques pour la détection des séquences appartenant à la famille *Caliciviridae*. Cette méthode s'est révélée être extrêmement sensible et a permis l'obtention d'un génome complet de NoVs GI à partir de deux échantillons d'eau usée faiblement contaminés (3 RNAc/ μ L). De plus plusieurs génomes presque complets de NoVs ont été séquencés permettant la caractérisation de 2 souches GI et 7 souches GII. La moitié d'une souche GIV, a également été obtenue. Toutefois les reads identifiés comme NoVs étaient encore très peu représentés sur la totalité des reads obtenues (<1%), suggérant une spécificité faible. Par ailleurs, les conditions de liaison et de lavage étant spécifiées par le fabricant et les séquences et concentrations des amorces n'étant pas publique, les marges d'améliorations sont limitées.

La méthode B a utilisé l'action du polyéthylène glycol (PEG) pour précipiter et concentrer les particules virales en modifiant leur gradient d'hydratation, et en jouant sur le point isoélectrique (PI) des NoVs pour l'élution virale. Les particules virales ont été adsorbées aux particules organiques et inorganiques contenues dans l'échantillon en variant la conductivité de l'échantillon (2000 μ S) et le pH (3) afin d'induire leur précipitation pendant une première étape de concentration avec PEG et centrifugation. Le changement du pH à 9 a ensuite permis l'élution de ces particules virales. Cette approche avait déjà été proposée pour isoler des NoVs à partir d'échantillons de selles (Jiang *et al.*, 1992). Ici nous avons ajouté des étapes de filtration en série (de 5 μ m jusqu'à 0.22) permettant d'éliminer certaines bactéries et de purification afin d'adapter le protocole à l'approche metagénomique.

Cette méthode a été efficace pour la détection de longues séquences de NoVs y compris cinq génomes complets. De plus, des séquences appartenant à d'autres virus ont été identifiées (coxsackievirus, sapovirus, astrovirus et rotavirus).

La méthode C a utilisé une stratégie de capture des particules virales en utilisant les mucines gastriques porcines. Cette étape intervient après les étapes prévues dans la méthode B et inclue la liaison des PGM type III (Sigma, France) préalablement fixées à des billes magnétiques (MagnaBind

TM-carboxyl-derivatized beads). Les PGM ainsi liées aux billes constituent une structure capable de lier les capsides virales présentant des sites de liaisons intactes et favorisant leur concentration grâce à des étapes d'incubation et de lavage sur un support magnétique. L'application de cette stratégie a permis d'augmenter le nombre de reads de NoVs et aussi d'autres familles virales telles que *Astroviridae*, *Reoviridae*, *Nodaviridae* et *Picobirnaviridae* suggérant la présence d'autres interactions. Toutefois, l'action des mucines n'a pas été toujours reproductible et effective en termes de nombre et taille des séquences virales obtenues.

1.5 CONCLUSIONS ET PERSPECTIVES

Un objectif de cette étude était d'évaluer la diversité des souches de NoVs par le séquençage de longs fragments ou des génomes entiers. L'enrichissement par capture en utilisant la puce SureSelect a donné le plus grand nombre de séquences complètes. Mais tout développement de cette méthode est limité car les étapes clés du protocole, ainsi que la composition du pool d'amorces utilisées, sont contrôlés par le producteur. De plus, l'utilisation de sondes peut se révéler être un facteur limitant dans l'identification de nouvelles variantes de NoVs qui ne seraient pas reconnues par les amorces et donc pas amplifiées.

Les difficultés de reproductibilité rencontrées dans l'utilisation des PGM et la présence des liaisons aspécifiques représentent un facteur limitant dans l'application future de cette méthode.

La méthode B, qui utilise une précipitation par le PEG comme stratégie de concentration, a permis l'identification de longues séquences et de génomes entiers de NoVs. De plus d'autres familles virales ont été détectées permettant de profiter de la stratégie de séquençage metagénomique pour l'étude du virome. D'autres avantages de cette technique résident dans la bonne répétabilité des résultats d'analyse, l'utilisation d'un équipement de laboratoire basique et de coûts de préparation de l'échantillon abordables.

De futurs développements sont néanmoins nécessaires pour augmenter la quantité de reads de NoVs par rapport au total de reads obtenus, en optimisant les étapes de concentration et d'éluion virale. Un des points faibles de cette étude, compliquant parfois l'interprétation des résultats a été le nombre limité d'échantillons. L'utilisation d'un plus grand nombre d'échantillon et l'addition de répliquats techniques permettraient de mieux évaluer l'efficacité des protocoles testés malgré la variabilité des échantillons.

ARTICLE II

2.1 INTRODUCTION A L'ETUDE

Les eaux usées sont des échantillons constitués d'excrétas de milliers de personnes, incluant urines, matières fécales et desquamation cutanées, ainsi que les différentes espèces de bactéries commensales ou pathogènes, les phages, les protozoaires et les virus constituant le microbiote (Nieuwenhuijse *et al.*, 2017; Adriaenssens *et al.*, 2018; Fumian *et al.*, 2019). Parmi cette microflore hautement diverse et variée, les particules de NoVs, petit virus à ARNs non enveloppé, considéré comme la principale cause de gastro-entérite humaine dans le monde, sont rejetés dans l'environnement *via* les excréta des individus infectés. Résistants dans l'environnement et n'étant pas totalement éliminés par les traitements dans les stations d'épuration, les NoVs contaminent les eaux de surface et potentiellement l'environnement marin (Sano *et al.*, 2016). Les huîtres élevées dans la zone côtière au travers de leur activité de filtration de l'eau de mer pour leur besoin physiologique, concentrent des particules y compris les NoVs. La présence dans les tissus digestifs de l'huître de sites de liaisons spécifiques, similaires aux antigènes tissulaires de groupe sanguin (HBGA) chez l'Homme, favorisent l'accumulation de certaines souches de NoVs (Le Guyader *et al.*, 2012). Ces carbohydrates, dont l'expression varie selon la saison, favorisent par exemple la bioaccumulation de la souche de Norwalk GI.1 par rapport aux génotypes qui sont plus fréquemment détectés dans les cas cliniques, tels que les souches de NoVs GII.3 ou GII.4 (Maalouf *et al.*, 2011; Yu *et al.*, 2015). Lorsque des huîtres présentent une contamination par plusieurs souches de NoVs, le risque d'évènements de recombinaison entre la jonction ORF1 et ORF2 du génome du virus peut exister lors de la réplication chez un consommateur infecté par ces diverses souches (Le Guyader *et al.*, 2008 ; van Beek *et al.*, 2018). Connaître la diversité des souches de NoV en circulation dans l'environnement est donc important pour estimer les souches présentes, sources potentielles de nouvelles épidémies.

Ces dernières années, le séquençage de nouvelle génération a été utilisé pour étudier les populations microbiennes dans des échantillons cliniques, permettant ainsi la détection de divers microorganismes cultivables et non cultivables (Nooij *et al.*, 2018; Osunmakinde *et al.*, 2018). L'application de cette même technique aux échantillons environnementaux pose des difficultés en raison des faibles concentrations de NoVs, de la persistance d'inhibiteurs potentiels susceptibles d'empêcher certaines réactions enzymatiques et de la grande diversité d'autres micro-organismes qui composent le bruit de fond lors de l'analyse des séquences obtenues (Strubbia *et al.*, 2019; Hata *et al.*, 2017; Adriaenssens *et al.*, 2018; Flaviani *et al.*, 2018; Osunmakinde *et al.*, 2018).

Pour ces échantillons, la préparation utilisée en amont pour concentrer et purifier les acides nucléiques peut avoir un impact sur la concentration du virus dans l'échantillon et sur le nombre de reads produites lors du séquençage (Hjelmsø 2017; Oechslin *et al.*, 2018; Fernandez-Cassi *et al.*, 2018). En effet, différents types d'échantillons peuvent nécessiter des approches adaptées pour optimiser la récupération du virus (Wylezich *et al.*, 2018 ; Bartsch *et al.*, 2018). Afin d'augmenter la sensibilité du séquençage et enrichir en séquences de NoVs, des systèmes de capture et amplification sélectives peuvent être appliqués. Wylie et collègues (2015) ont développé un système d'enrichissement des séquences (VirCapSeq-VERT) ciblant les virus ADN et ARN susceptibles d'infecter les vertébrés, y compris l'Homme. Cette approche pourrait être un outil valide dans le séquençage d'échantillons environnementales faiblement contaminés.

2.2 OBJECTIF DE L'ETUDE

Ce travail présente des développements méthodologiques axés sur l'identification des NoVs. Trois échantillons d'eaux usées ont été prélevés dans une station d'épuration pendant la période hivernale (février – mars 2018). Après quantification par dPCR, chaque échantillon a été utilisé pour contaminer un bac contenant de l'eau de mer et des huitres (B1, B2 et B3).

Différents protocoles d'élution ont été appliqués aux échantillons d'eaux usées et aux huitres ainsi bioaccumulées. Les échantillons d'huitres ont été séquencés selon une double approche, direct (ou agnostique) et ciblée, utilisant le système de capture VirCapSeq-VERT. La technologie Illumina MiSeq a été utilisée pour séquencer les bibliothèques préparées. L'étude ayant nécessité trois runs indépendants, (un pour chaque expérience de bioaccumulation), un échantillon de référence a été utilisé pour contrôler la variabilité du séquençage haut débit.

L'objectif final était d'obtenir de longs fragments pour caractériser les souches de NoVs en identifiant les séquences codant les régions ORF1 et ORF2. Cette approche devrait permettre d'identifier la diversité des NoVs dans les échantillons d'eaux usées et d'évaluer la sélection en NoVs exercée par les huitres.



Metavirome Sequencing to Evaluate Norovirus Diversity in Sewage and Related Bioaccumulated Oysters

Sofia Strubbia¹, Julien Schaeffer¹, Bas B. Oude Munnink², Alban Besnard¹, My V. T. Phan², David F. Nieuwenhuijse², Miranda de Graaf², Claudia M. E. Schapendonk², Candice Wacrenier¹, Matthew Cotten², Marion P. G. Koopmans² and Françoise S. Le Guyader^{1*}

¹ Laboratoire de Microbiologie, LSEM-SG2M-RBE, Ifremer, Nantes, France, ² Department of Viroscience, Erasmus University Medical Center, Rotterdam, Netherlands

OPEN ACCESS

Edited by:

Hirokazu Kimura,
Gunma Paz University, Japan

Reviewed by:

Wen Zhang,
Jiangsu University, China
David H. Kingsley,
Agricultural Research Service (USDA),
United States
Leesa Bruggink,
Victorian Infectious Diseases
Reference Laboratory, Australia

*Correspondence:

Françoise S. Le Guyader
solzick.le.guyader@ifremer.fr;
sleguyad@ifremer.fr

Specialty section:

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

Received: 24 July 2019

Accepted: 03 October 2019

Published: 17 October 2019

Citation:

Strubbia S, Schaeffer J, Oude Munnink BB, Besnard A, Phan MVT, Nieuwenhuijse DF, de Graaf M, Schapendonk CME, Wacrenier C, Cotten M, Koopmans MPG and Le Guyader FS (2019) Metavirome Sequencing to Evaluate Norovirus Diversity in Sewage and Related Bioaccumulated Oysters. *Front. Microbiol.* 10:2394. doi: 10.3389/fmicb.2019.02394

Metagenomic sequencing is a promising method to determine the virus diversity in environmental samples such as sewage or shellfish. However, to identify the short RNA genomes of human enteric viruses among the large diversity of nucleic acids present in such complex matrices, method optimization is still needed. This work presents methodological developments focused on norovirus, a small ssRNA non-enveloped virus known as the major cause of human gastroenteritis worldwide and frequently present in human excreta and sewage. Different elution protocols were applied and Illumina MiSeq technology were used to study norovirus diversity. A double approach, agnostic deep sequencing and a capture-based approach (VirCapSeq-VERT) was used to identify norovirus in environmental samples. Family-specific viral contigs were classified and sorted by SLIM and final norovirus contigs were genotyped using the online Norovirus genotyping tool v2.0. From sewage samples, 14 norovirus genogroup I sequences were identified of which six were complete genomes. For norovirus genogroup II, nine sequences were identified and three of them comprised more than half of the genome. In oyster samples bioaccumulated with these sewage samples, only the use of an enrichment step during library preparation allowed successful identification of nine different sequences of norovirus genogroup I and four for genogroup II (> 500 bp). This study demonstrates the importance of method development to increase virus recovery, and the interest of a capture-based approach to be able to identify viruses present at low concentrations.

Keywords: norovirus, sewage, oysters, metagenomic sequencing, metavirome

INTRODUCTION

Noroviruses (NoV) are small, icosahedral non-enveloped viruses, belonging to the *Caliciviridae* family, and are recognized as the predominant cause of non-bacterial human gastroenteritis worldwide. Their single strand RNA genome is short (around 7.5 k bases) and contains three open reading frames (ORFs), of which ORF1 encodes the non-structural proteins while ORF2 and ORF3 encode the major and minor capsid structural proteins (VP1 and VP2, respectively). Based on their

genetic characteristics, NoV are classified into at least seven genogroups (G), from GI to GVII, which are further divided into over 30 genotypes (de Graaf et al., 2016).

Noroviruses circulate year-round but in disease surveillance distinct seasonal peaks are seen in regions with winter seasonality, reflecting dynamics of the NoV genotypes that are most commonly associated with sporadic cases and outbreaks. NoV particles released in human vomitus and stools can then be detected in sewage and may contaminate surface waters including the marine environment (Sano et al., 2016). However, a wide and increasing range of other viruses has been identified in human stools, partially associated with gastroenteritis, but also non-mammalian viruses (Nieuwenhuijse and Koopmans, 2017). Raw sewage are rich sample types, consisting of the excreta of thousands of people, including urine, feces, and skin desquamation and all associated commensal and pathogenic bacteria, phages, and protozoa, and viruses, including viruses associated with microbiota and diet associated viruses (Nieuwenhuijse and Koopmans, 2017; Adriaenssens et al., 2018; Fumian et al., 2019). Therefore, finding NoV sequences in this complex matrix may be challenging.

After reaching the marine environment, mainly due to accidental discharge or wastewater treatment plants' effluents, NoV are highly resistant. Oysters farmed in coastal areas can accumulate NoV in their digestive tissues (DT) as they filter large amounts of water and thereby come into contact with numerous molecules and particles. We previously reported that some oyster species selectively accumulate NoV due to the presence of specific carbohydrates, similar to the human histo blood group antigens (HBGA and so called HBGA-like antigens) (Le Guyader et al., 2012). These carbohydrates, with a clear seasonal expression, favor for example GI.1 NoV bioaccumulation when compared to genotypes that are more often detected in the human population, like GII.3 or GII.4 (Maalouf et al., 2011; Yu et al., 2015). When oysters contaminated with multiple strains are consumed, there is a risk of disease for consumers, but it may also result in simultaneous single cell infections leading to recombination events (Le Guyader et al., 2008). These recombination events are mainly happening at the junction between ORF1 and ORF2 (van Beek et al., 2018). Therefore, understanding the full NoV diversity in oysters is important.

During the last years, next generation sequencing has increasingly been used to study microbial populations in environmental samples, allowing the detection of several cultured and non-cultured microorganisms (Nooij et al., 2018; Osunmakinde et al., 2018). Nevertheless virus detection in environmental samples is challenging in metagenomics studies due to the high levels of (background) host and microbial DNA, high virus diversity in sewage, and the high rate of, potentially novel, unclassifiable sequences (Bibby and Peccia, 2013; Hjelmso et al., 2017; Nooij et al., 2018). In addition, low concentrations of NoV, the persistence of potential inhibitors that may prevent certain enzymatic reactions and the large diversity of other microorganisms present in the sample complicates the detection of NoV using metagenomic sequencing (Hata et al., 2017; Adriaenssens et al., 2018; Flaviani et al., 2018; Osunmakinde et al., 2018; Strubbia et al., 2019). Therefore, sequencing outcome is

dependent on sample preparation, meaning that the relationship between sample concentration and the number of NoV reads may vary according to the method applied to concentrate and purify the nucleic acids (Hjelmso et al., 2017; Fernandez-Cassi et al., 2018; Oechslin et al., 2018). A metabarcoding approach, which is amplicon based and thus requires specific primers, was proposed to sequence NoV from environmental samples allowing deep-sequencing even in samples with low virus concentrations and independently from the host background (Kazama et al., 2017; Oechslin et al., 2018). However, this approach and the design of the primers is limited to our current knowledge about NoV diversity and therefore will potentially miss the detection of new NoV strains that are genetically distinct from known diversity. Due to this limitation, metagenomic sequencing would be more suitable for the identification of new NoV strains (Fumian et al., 2019).

Here, we describe the performance of different concentration methods on the recovery of NoV from sewage and oyster samples. These methods allowed us to describe the NoV diversity in sewage samples and in oysters exposed to these sewages, to identify long sequences (>500 bp up to full genome) needed for strain identification, and to compare the NoV diversity in both sample types.

MATERIALS AND METHODS

Samples Collection

Composite raw sewage sampled over 24-h were collected from a sewage treatment plant in a large city in western France (303 800 inhabitants). Three samples, which constituted of at least 5-L, were collected within 2 months (February and March 2018) and transported at 4°C to the laboratory, analyzed for the presence of NoV and kept frozen at -20°C in aliquots of 1 L.

Oysters (*Crassostrea gigas*) were directly purchased from the same producer few days before each experiment, analyzed for NoV contamination and kept at 4°C until use for bioaccumulation experiments.

Bioaccumulation

Three bioaccumulation experiments (B1, B2, B3) were performed, one in February and two in March to limit oyster physiological variability and seawater composition. Aquariums were filled with 22-L of natural seawater seeded with adjusted volume of sewage to reach a comparable NoV concentration for all three experiences (around 10^7 RNA copies/L). Then, 140 oysters were immersed for 24 h at $12 \pm 1^\circ\text{C}$, under oxygenation.

Sewage Preparation

Wastewater samples were prepared according to two different methods.

- For the polyethylene glycol (PEG) method, 40 mL of sample was adjusted to pH 4 by adding HCl, and adjusted to conductivity reading of 2000 μS by adding NaCl 5 M. After 5 min, 10 mL of 50% PEG 6000 solution (Sigma-Aldrich, St-Quentin France) was added. The mixture was incubated under gentle agitation overnight at 4°C and centrifuged at $13,500 \times g$ for 90 min. The

pellet was resuspended in 2 mL of glycine buffer 0.05 M (pH 9) and mixed with 2 mL of Chloroform-Butanol (vol/vol), vortexed at high speed for 30 s and centrifuged for 5 min at 11,000 × g. The aqueous phase was recovered and used for nucleic acid extraction.

- For the Pyro-PEG method, 4 mL of 10 mM sodium pyrophosphate decahydrate was added to 40 mL of sample and incubated for 40 min at room temperature under gentle agitation (Bisseux et al., 2018). Then, the mixture was sonicated for 1 min at maximum power in a cup-horn adaptor (Bandelin, HD 2200), followed by 1 min recovery on ice, which was repeated three times. After centrifugation for 20 min at 8,000 × g, supernatants were recovered, the pH adjusted to 4 and the conductivity was adjusted of 2000 μs by adding NaCl 5 M. After 5 min, 10 mL of 50% PEG 6000 solution (Sigma-Aldrich, St-Quentin France) was added and rocked for 1 h at 4°C. After centrifugation for 90 min at 13,500 × g, the pellet was resuspended in 2 mL of glycine buffer pH 9.

For both methods, the resuspended pellets were filtrated using a cascade of 5, 1.2, and 0.45 μm filter pores (Minisart NML 17594, NML17593, PES16533, and PES16532). The recovered filtrates were incubated for 1 h at 37°C with 2000 Units of OmniCleave EndonucleaseTM (Lucigen Corporation) and 100 μL of MgCl₂ (100 mM).

Oyster Preparation

Following bioaccumulation, oysters were immediately collected and shucked using a sterile knife. Flesh weight was recorded to calculate the allometric coefficient (flesh weight divided by DT weight), used to describe physiology of the animals (Polo et al., 2018). This coefficient was similar (14 to 10) for the three batches of oysters used for the experiments, presuming that they have similar filtering and physiology activity and are thus comparable for the three different experiments. The digestive tissues (DT) were dissected, chopped finely to homogenize, pooled, and distributed into 2 g aliquots, immediately frozen at -80°C.

Viruses were recovered using two methods. For both methods, 2 g of DT were incubated with 2 ml of proteinase K solution (30 U/mg, Sigma-Aldrich, France) for 15 min at 37°C and additional 15 min at 60°C. The mixtures were sonicated for 1 min at maximum power, followed by 1 min on ice, repeated three times. Supernatants were collected after centrifugation for 5 min at 3000 × g and kept at 4°C (method PK). For method PK-PEG, DT were treated using same conditions but then the supernatant was mixed with two volumes of PEG-NaCl 1.2 M and incubated under gentle agitation for 1 h at 4°C, before centrifugation at 11,000 × g for 20 min. Pellets were resuspended in 1 mL of glycine buffer (0.05 M) pH 9. Samples from both methods (PK and PK-PEG), were then incubated for 1 h at 37°C with 2000 Units of OmniCleave Endonuclease (Lucigen Corporation) and 100 μL of MgCl₂ (100 mM).

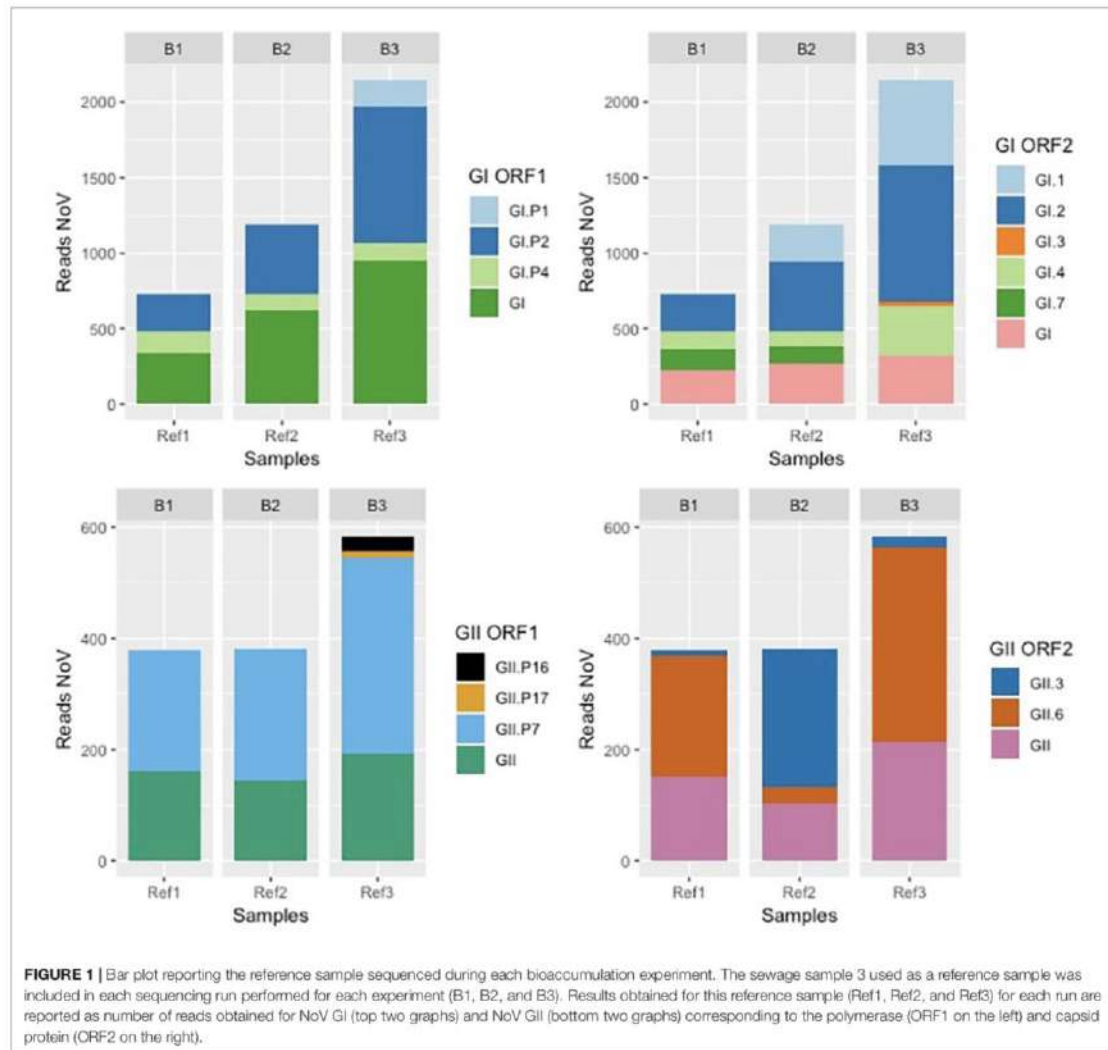
RNA Extraction and Purification

Nucleic acids extraction was performed using lysis buffer (bioMérieux, France) and the NucliSens kit (bioMérieux) followed by DNase treatment for 30 min at 37°C with 25 U TURBOTM DNase (Ambion, Thermo Fisher Scientific, France).

TABLE 1 | Data obtained from the three bioaccumulation experiments.

Experiment	Sample	Method	B1			B2			B3					
			NoV conc.*	Reads per library	NoV reads	Ratio**	NoV conc.*	Reads per library	NoV reads	Ratio**	NoV conc.*	Reads per library	NoV reads	Ratio**
			GI	GII		GI	GII		GI	GII		GI	GII	
Sewage	PEG		1	30	4,379,484	9138	2,09 × 10 ⁻³	<LD	28	4,606,300	138	3,00 × 10 ⁻⁵	83	60
	Pyro-PEG		268	135	4,248,714	696	1,64 × 10 ⁻⁴	575	348	4,135,170	2092	5,06 × 10 ⁻⁴	675	290
Oyster	PK		200	200	4,379,484	21	4,80 × 10 ⁻⁶	150	150	5,325,428	0	0	900	500
	PK-PEG		1100	1250	4,990,302	0	0	0	1150	5,665,130	303	5,35 × 10 ⁻⁵	450	1000
Reference sample**					4,443,016	0	0	1350	1150	4,816,936	0	0	450	1000
					4,146,710	3789	9,14 × 10 ⁻⁴	0	0	3,633,280	6930	1,91 × 10 ⁻³	0	0
Total reads per run [†]					4,323,005	2596	5,94 × 10 ⁻⁴	0	0	4,377,564	2673	6,11 × 10 ⁻⁴	0	0
					21,773,704					23,261,369				

*Norovirus concentrations measured by digital PCR are expressed as RNA copies per mL for sewage samples and RNA copies/g DT for oyster samples for NoV GI and NoV GII, <DL: under the limit of detection. **Ratio of NoV reads obtained to the total number of reads (Tot reads). ***Reference sample correspond to the sewage sample (ww3) used for the B3 experiment, concentrated using the Pyro-PEG method, and included in each run of metagenomic as a control sample. This sample was sequenced in duplicate during the third run. [†]The total reads per run is the sum of the reads obtained for all libraries of the column (except the gray case performed on a separate run). The gray lines correspond to the runs performed after RNA capture using the VircapSeq-VERT.



An additional RNA purification was performed using the RNA Clean & ConcentratorTM-5 kit (Zymo Research, Irvine, CA, United States). Each sample was extracted three times and extracts were pooled to obtain a final volume of 300 μ L. RNA extracts were stored at -80°C in aliquots of 30 μ L.

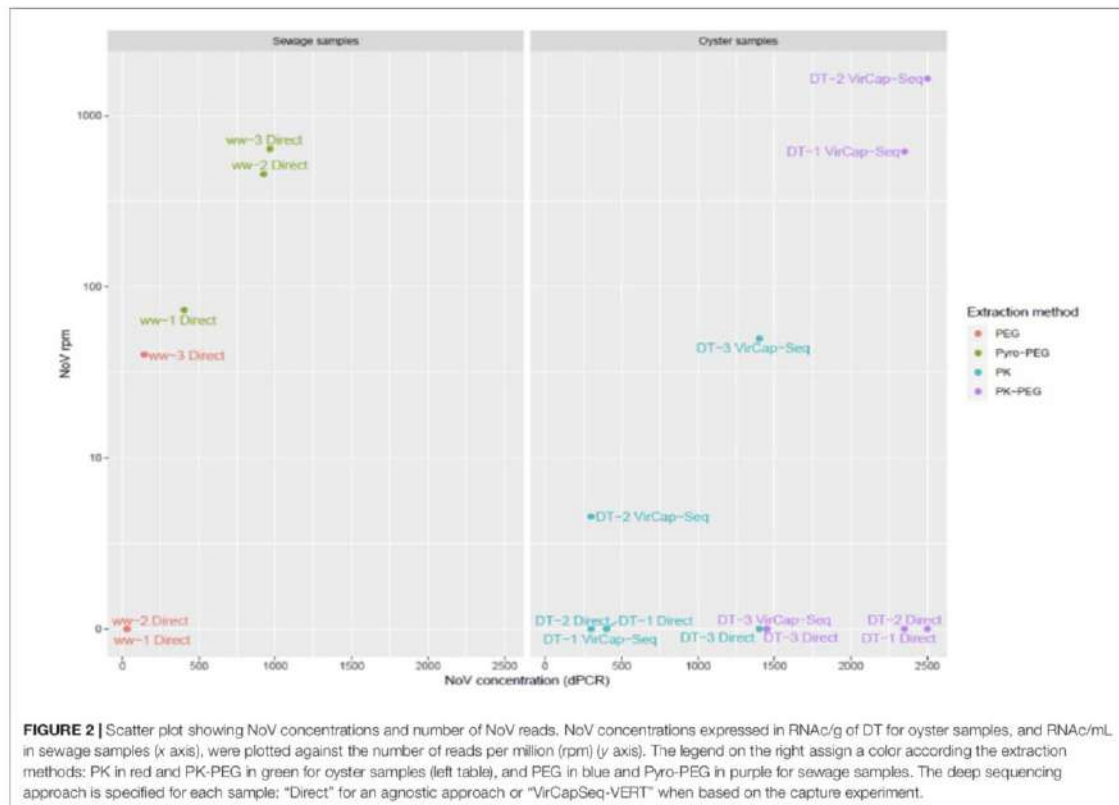
NoV Quantification in Sewage and DT

Noroviruses quantification was performed with a one-step digital RT-PCR using primers and probes targeting the conserved region at the beginning of ORF2 (Maalouf et al., 2011; Polo et al., 2016). Positive and negative controls were included in each series to validate the distribution of positive and negative wells. Quantification was done using the Poisson distribution

(QuantStudioTM 3D Analysis SuiteTM Cloud Software, version 3.0.3; Thermo Fisher Scientific, France) and the final result was expressed as RNA copies/ μ l (Polo et al., 2016).

Library Preparation for Agnostic Metagenomic Sequencing

RNA extracts were transcribed into cDNA using Superscript II (Invitrogen, France) and non-ribosomal hexamers (Endoh et al., 2005). Libraries were prepared using the NEB Next Ultra DNA Library Prep Kit (New England BioLabs, France) according to the manufacturer's instructions. Sequencing was performed using Illumina MiSeq technologies to generate 2×150 bp reads. All samples were sequenced agnostically (referred to



as direct sequencing in the text). Each run corresponded to one bioaccumulation experiment (B1, B2 or B3) including the sewage sample (ww1, ww2 or ww3) prepared with the two methods and the DT (DT-1, DT-2, and DT-3) prepared with the two methods. Each run also included the reference sample (called Ref1, Ref2, and Ref3 according to the run) that correspond to the ww3 sample (extracted with the Pyro-PEG method) split after library preparation and sequenced in each run to compare the performance of the different sequencing runs.

VirCapSeq-VERT Capture

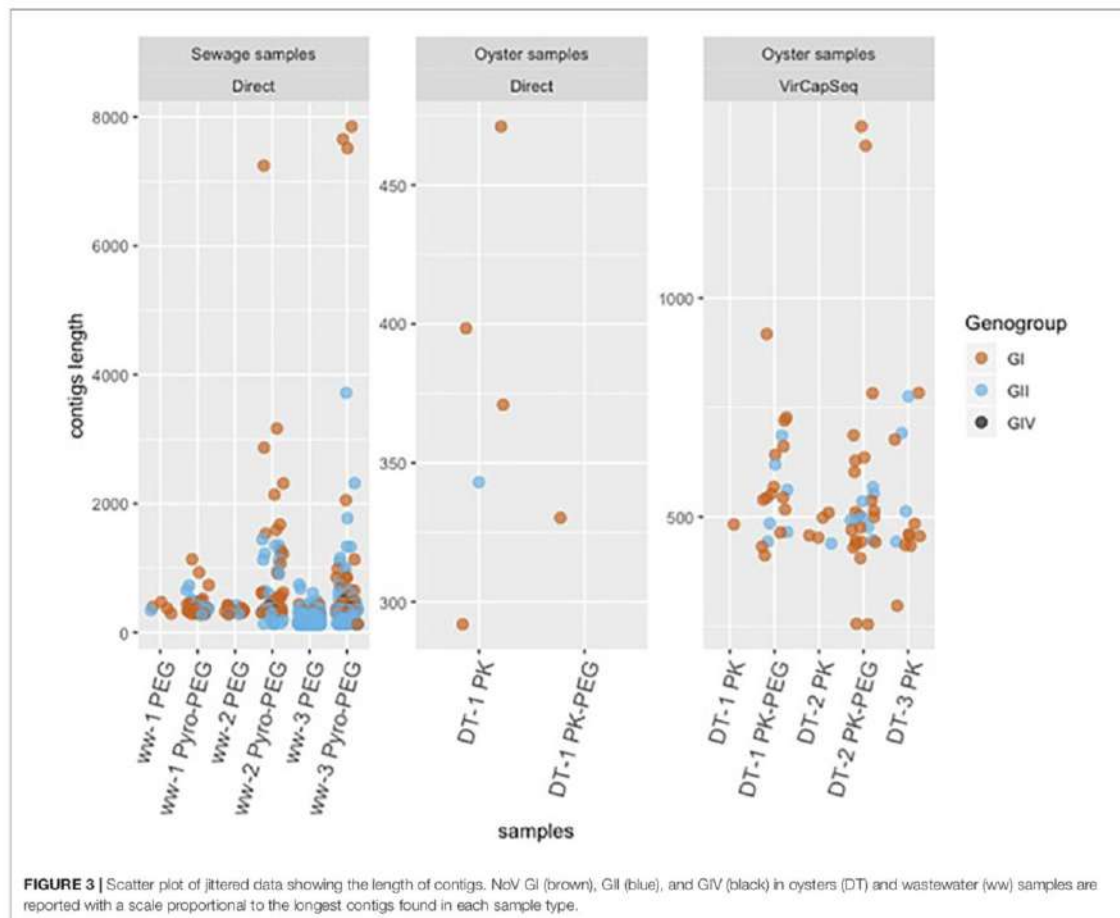
Oysters' DT were also analyzed with a targeted deep sequencing using VirCapSeq-VERT for viral enrichment (Wylie et al., 2015). In short, reverse transcription was performed using random hexamers and SuperScript III (Thermo Fisher Scientific) after which dsDNA synthesis was performed using Klenow (New England Biolabs, France). Libraries were prepared using the KAPA HyperPlus Kit (Roche, France) according to the manufacturer's instructions with slight modifications. The shearing time was reduced to 3 min and adapters were diluted 1:10. After the adapter ligation an addition AMPure bead step was performed. The libraries were quantified and pooled equimolarly after which the capture experiment was performed

(Wylie et al., 2015). Sequencing was performed on an Illumina MiSeq using the MiSeq Reagent KIT v3 (Illumina) to generate 2×300 bp reads.

Sequence Analysis

Illumina adapters were removed from the raw short reads and resulting reads were trimmed from 3' end to reach an average the Phred score ≥ 35 using QUASR (Watson et al., 2013). Resulting reads were *de novo* assembled using SPAdes v3.12.0 and family-specific viral contiguous sequences (call contigs) were classified and sorted by SLIM (Bankevich et al., 2012; Cotten et al., 2014). Then, SLIM outputs were filtered to select contigs with an identity score $\geq 85\%$ and a minimum length of 500 bp to avoid too short and non-informative fragments. Partial overlapping contigs were further assembled into genomes using Geneious® v. 11.1.5 and all SNPs (single nucleotide polymorphisms) between contigs were resolved by counting motifs in the raw read sets. For NoV, final contigs were genotyped using the online norovirus genotyping tool v2.0 (Kroneman et al., 2013). In order to estimate the read coverage, raw reads were mapped to the resulting contigs using Bowtie2 (v2.3.0) (Langmead and Salzberg, 2012).

Maximum likelihood trees were inferred with PhyML v3.0, using the general time reversible (GTR) nucleotide substitution



model (Guindon et al., 2010). VP1 sequences representing different GI and GII genotypes were included in the phylogenetic trees. The GI and GII trees were inferred using an alignment of approximately 800 bp from the middle and end of VP1, respectively. Different regions were selected to be able to include the maximum amount of Illumina sequences.

RESULTS

This study aimed to optimize NoV recovery from sewage and oyster DT samples, to analyze subsequent reads obtained and then to finally compare strains detected in sewage samples and oyster samples contaminated with these sewages.

NoV Concentrations in Sewage and Oyster Samples

For sewage samples, the Pyro-PEG method increased the recovery of NoV when compared to the PEG method, both in

terms of concentrations (RNA copies per μl measured by dRT-PCR) and read numbers after NGS (Table 1). For the sewage sample (ww2), used for the second bioaccumulation experiment (B2), the NoV GI concentration was under the detection limit when treated with PEG, but increased up to more than 500 copies/mL with the pyro-PEG method. When comparing geometric means of the concentrations calculated for the three bioaccumulation experiments the number of NoV GI particles increased over 2 logs while the number of NoV GII particles increased 1 log when using the Pyro-PEG compared to only PEG.

For oyster DT samples, NoV concentrations were always higher using the PK-PEG method, and the increase differed by genogroup (from 0.5 to 9 fold for NoV GI and from 2 to 7.6 fold for NoV GII). These concentrations were representative of oyster contaminations observed in highly polluted area.

Reference Samples Sequencing

Each bioaccumulation experiment (B1, B2, and B3) was sequenced in separate sequence runs and the library of the sewage

sample used for the third bioaccumulation (B3) was included in each run so it can be used as reference sample. The first and second run gave comparable numbers of total reads, while the third run produced two times more reads (Table 1). However, a comparable ratio of NoV reads were obtained from the reference sample in the three different sequence runs: in the first run $5,94 \times 10^{-4}$ NoV reads per million sequence reads (rpm) were obtained and in the second and third run $6,11 \times 10^{-4}$ NoV rpm were obtained. The diversity of NoV genotypes identified in the reference sample in the three different runs is comparable, with a higher number of NoV reads for each genotype? in the third run (Figure 1). Additionally, the third run of the reference sample (Ref3) yielded reads of NoV GI.3 and GII.P16 that were not detected in runs 1 and 2 (Figure 1). This confirms the importance to get sufficient sequence reads to be able to detect all strains that might be present.

Method Comparison for NoV Sequencing From Sewage Samples

To evaluate the depth of NoV sequencing, we tested two different extractions protocols. We investigated if there was a relationship between the NoV concentration (sum of NoV GI and GII) and the number of NoV reads after NGS (Figure 2). Sewage samples treated with the Pyro-PEG method showed higher numbers of NoV reads in agreement with higher NoV concentrations obtained using this method. A small number of reads of NoV GIV were obtained from samples treated with Pyro-PEG method but not from those treated with the PEG method.

Another aim of this method comparison was to increase NoV contig length in order to increase the identification and classification of NoV strains present in our samples. For sewage samples, only the Pyro-PEG method resulted in contigs with a length over 1000 nt for both genogroups and even up to 7000 nt (full genomes) for NoV GI (Figure 3). Using the Pyro-PEG methods, six full NoV GI genomes were recovered from sewage samples (Table 2). For NoV GII, the longest sequence detected was a GII.P7-GII.6 (5050 nt). For some other strains (nearly) full genomes were not obtained but the overlapping region of ORF1 and ORF2 region was sequenced allowing for dual genotyping and classification (Table 3). Some sequences such as GII.6, and GII.4 were detected in all three sewage samples that is not surprising as samples were collected from the same sewage treatment plant over a short period of time. Quite a large

diversity of GI strains was also identified in ww-2 and ww-3, confirming the interest of sewage samples to identify some strains that may not be seen in clinical cases.

Method Comparison for NoV Sequencing From Oyster Samples

For DT samples, only 21 NoV reads were obtained without enrichment and all of them were from the sample DT-1 treated with PK (Table 1). The preparation of the sequence libraries were repeated using a specific enrichment for vertebrate viruses (VirCapSeq-VERT). This sequence run resulted in 31,377,552 sequence reads and NoV contigs were recovered from all three DT (DT-1, DT-2, and DT-3). A positive trend between the number of reads and NoV concentrations was observed after using VirCapSeq-VERT. The PK-PEG method provided a higher amount of NoV reads for DT-1 and DT-2, with 3789 and 6930 NoV reads, respectively, but not for DT-3. Using the PK method, 303 and 458 NoV reads were obtained from DT-2 and DT-3, but none from DT-1.

Direct sequencing from oysters' DT did not provide enough NoV reads to allow strain identification. Using VirCapSeq-VERT, the PK-PEG method provided the longest contig length and in total 57 contigs of >500 nt length were obtained, compared to 19 using the PK method. The longest consensus sequence obtained from oysters' samples was 1393 nt and allowed the identification of a GI.P7-GI.7 strain. The diversity of sequences detected is quite large both for NoV GI and GII (Figures 4A,B).

Comparing NoV Diversity in Sewage and in Bioaccumulated Oysters

One objective of this work was to compare the diversity of NoV in sewage samples and in oysters contaminated with these sewage samples. Due to the failure to obtain NoV reads after direct sequencing, this led us to use VirCapSeq-VERT sequencing. Different methods may induce differences in strains recovery or identifications nevertheless our results allowed us to make some observations (Figure 5). Furthermore, identical genotypes were detected in sewage and DT for GI (GI.p1-GI.1, GI.p4-GI.4, and GI.p7-GI.p7 strains, Figure 4A) and for GII (GII.6 and GII.P16-GII.4, Figure 4B). Phylogenetic trees inferred for partial VP1 GI GII sequences (Figures 4A,B) showed that these sequences were identical. Unfortunately, the third experiment provides only

TABLE 2 | Complete NoV genomes identified.

Sample (run)	Sample code	ORF 1	ORF 2	Length	BLAST score*	Ref. sequence [†]	Identity [‡]
Ref sample (1)	G19_001	GI.P1	GI.1	7134	81.54	KF306212.1	87%
Ref sample (2)	G19_004	GI.P2	GI.2	7619	92.27	KF306212.1	99%
ww-3 (3)	G19_005	GI.P4	GI.4	7581	91.10	LN854563.1	96%
ww-3 (3)	G19_006	GI.P2	GI.2	7433	94.56	KF306212.1	99%
Ref sample (2)	G19_002	GI.P1	GI.1	7173	84.99	NC_001959.2	85%
ww-3 (3)	G19_003	GI.P1	GI.1	7363	84.72	NC_001959.2	86%

Blast score (*) was assigned using the NoV typing tool v2.0 (Kroneman et al., 2011), and reference sequence and percentage of identity were obtained using BLAST (NCBI) ([†]).

TABLE 3 | Noroviruses GII sequences for which it was possible to identify both ORFs (ORF1 and ORF2).

Sample (run)	Sample code	ORF 1	ORF 2	Length	BLAST score*	Ref. sequence#	Identity#
ww-3 (3)	G19_007	GII.P7	GII.6	3720	85.63	MH218642.1	98%
Ref sample (3)	G19_008	GII.P7	GII.6	5050	87.11	MH218642.1	98%
Ref sample (1)	G19_009	GII.P7	GII.6	4292	86.55	MH218642.1	98%
Ref sample (2)	G19_011	GII.P17	GII.17	1374	81.57	KU561249.1	98%
TD (1)	G19_012	GII.P16	GII.4 Syd 2012	686	89.85	KY210980.1	99%

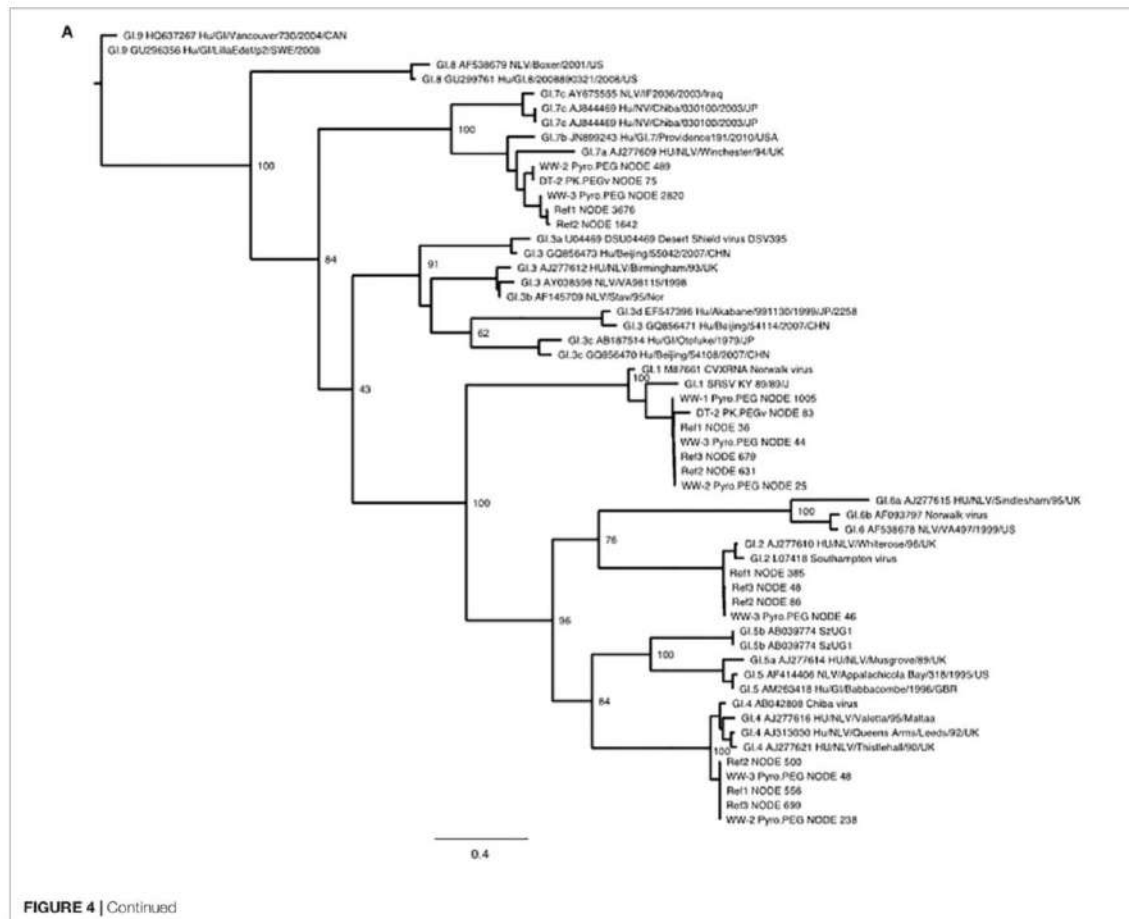
Reference sequence and percentage identity were obtained using BLAST (NCBI). Blast score (*) was assigned using the NoV typing tool v2.0 (Kroneman et al., 2011), and reference sequence and percentage of identity were obtained using BLAST (NCBI) (#).

few sequences from DT (one GI that could not be typed and one GII.17 strain).

Diversity of Other Viruses in Environmental Samples

Among the eukaryotic viruses identified, a number of assembled contigs were shown to belong to other human enteric virus

families including members of the *Reoviridae*, *Picornaviridae*, *Calciviridae*, and *Astroviridae*. For sewage samples, the Pyro-PEG method provided a higher number and diversity of viral reads for all samples (Figure 6A). This is especially true for sewage sample 1 (ww-1) for which no read corresponding to human enteric viruses was identified using the PEG method while using Pyro-PEG allowed the identification of several different viruses. Members of the *Astroviridae* family were most



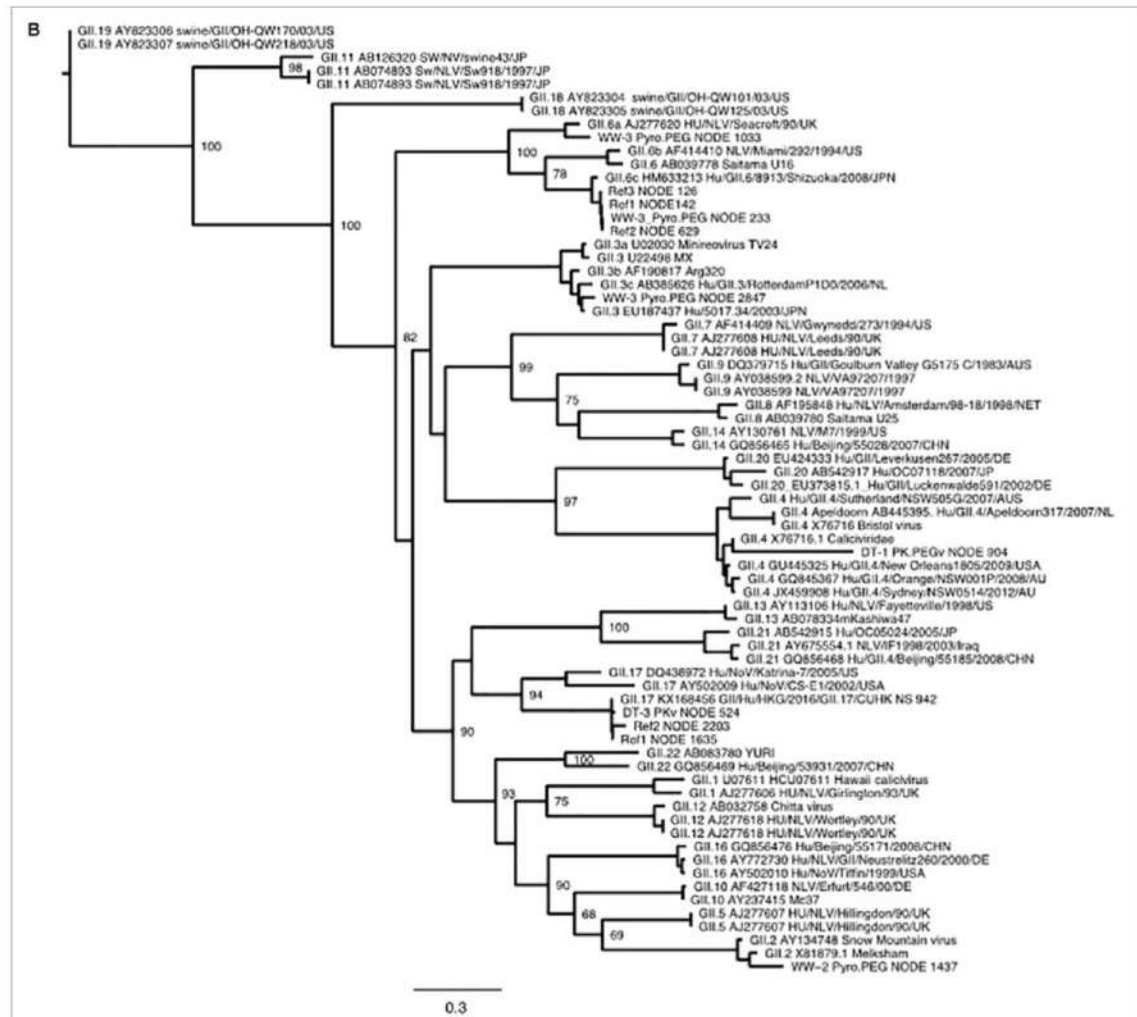
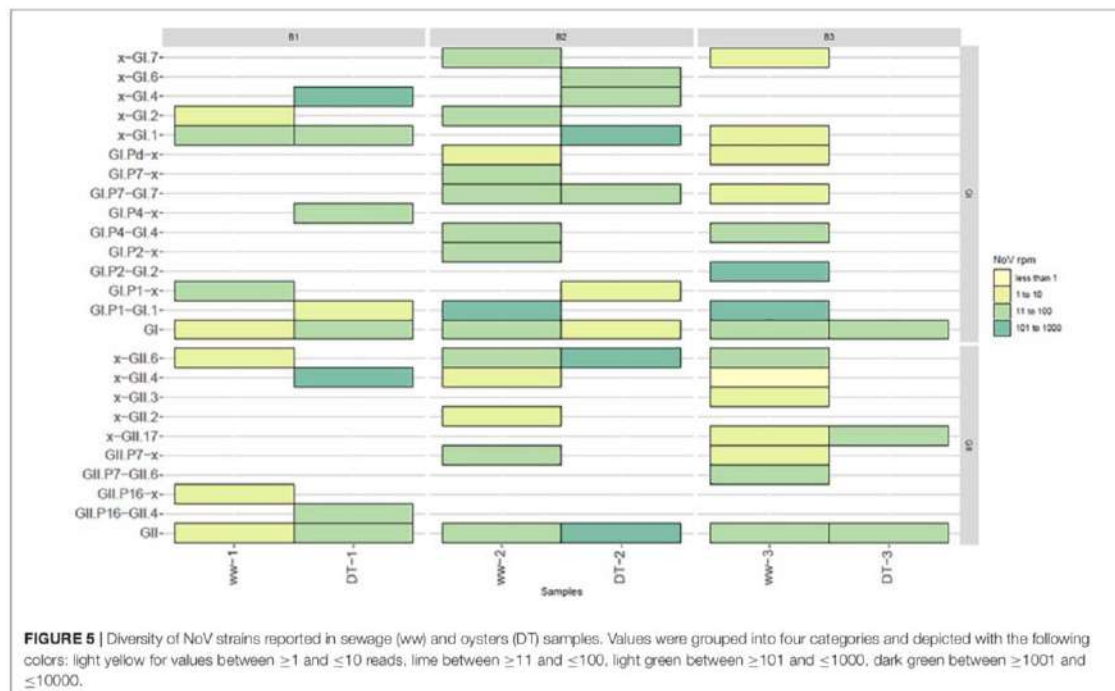


FIGURE 4 | Phylogenetic trees inferred from partial GI (A) and GII (B) VP1 sequences using PhyML. All identified sequences in sewage sample (ww-1, ww-2, and ww-3), the reference sample (Ref1, Ref2, and Ref 3) and the oyster Digestive tissues (DT-1, DT-2, and DT-3). The method used to prepare the nucleic acids are reported, except for the Ref sample as the library was prepared only after the Pyro-PEG method.

abundantly present in sewage samples using both methods and four complete genomes were identified (one type 1, two type 2, and one type 3). Only in one occasion (ww-3), reads of the *Reoviridae* family were 10 time more represented after the PEG method compared to the pyro-PEG method, but contigs did not allowed strain identification. Concerning viruses belonging to the *Caliciviridae* family, a consistent improvement in sequence read recovery and contig length was reported with Pyro-PEG for all three sewage samples (between 1,000 and 10,000 times more), in agreement with the NoV concentrations. Three sapovirus complete genomes were obtained (GI.2, GI.3, and GII.1). For DT

samples (DT-1, DT-2, and DT-3), very few sequences were identified after direct sequencing and only some after the PK-PEG method. This difference between the two protocols persisted after applying the VirCapSeq-VERT capture. The number of viral reads obtained was much higher for all viral families in samples treated with PK-PEG method, except for DT-3 for which very few reads were obtained (Figure 6B). Furthermore, the PK-PEG method led to the identification of viruses belonging to the *Picornaviridae*, *Picobirnaviridae*, *Caliciviridae*, and *Astroviridae* families. Comparing viral families detected both in sewage and DT, only sequence reads derived from the *Reoviridae* family were



reported in ww- 2 and 3 (but with read numbers < 100), while they could not be detected in the corresponding DT extracts, suggesting a low accumulation of rotavirus by oysters.

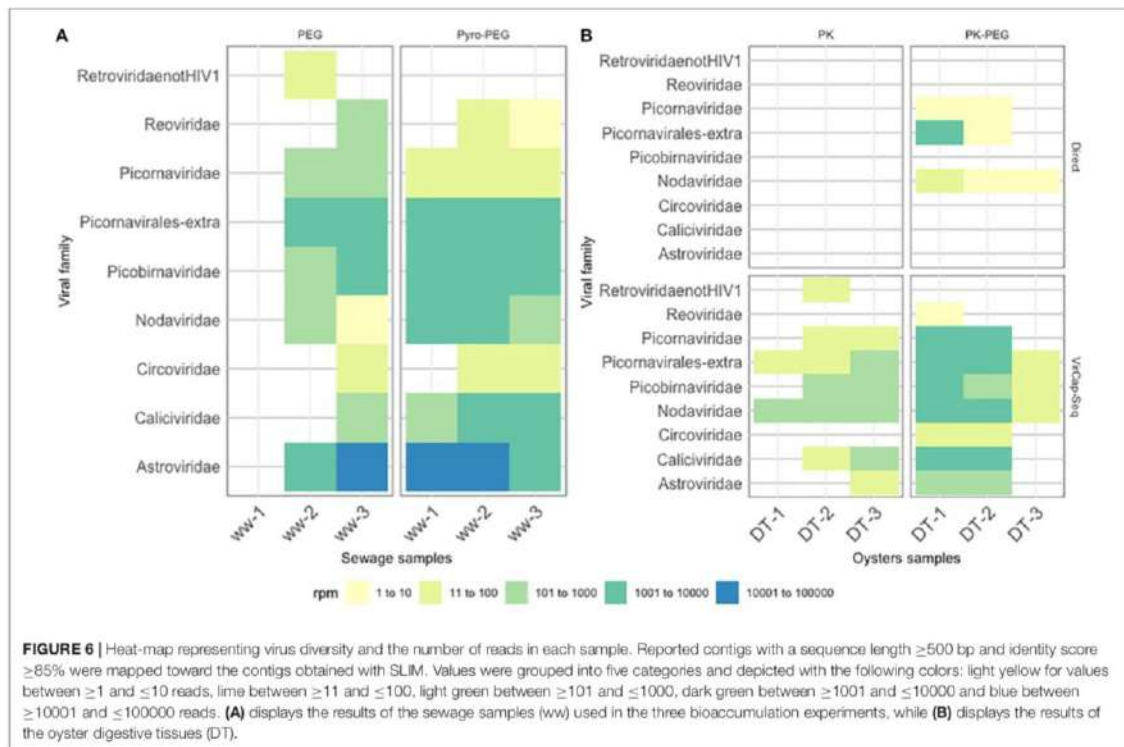
DISCUSSION

There are only a limited number of protocols published for metagenomic NoV sequencing directly from sewage or oyster samples. Here, we evaluated different methods to improve the characterization of human enteric viruses, with a special focus on NoV, in sewage samples and oysters contaminated with these sewage samples using next generation sequencing. Our final goal was to obtain full NoV genomes to be able to study NoV diversity in environmental samples.

Based on experience in sewage analysis using PEG, we optimized the elution/concentration step to increase virus recovery (da Silva et al., 2007). One protocol is based on acidification to enhance binding of the viral capsid to particles according to their isoelectric point before PEG precipitation, a method found efficient to detect low levels of NoV in oysters (Shieh et al., 1999). The other method includes first a chemical elution step using sodium pyrophosphate combined with a sonication step, before the PEG concentration (Bisseux et al., 2018). This sonication step favors the disruption of agglomerates, enhances viral particles elution from organic fragments and reduce the amount of bacteria (Ettayebi et al., 2016; Santiana et al., 2018). This Pyro-PEG method was found efficient as it

allowed the recovery of the highest number of NoV reads and was reproducible, as demonstrated by the increased proportion of NoV reads yielded with this approach when compared to the PEG method.

Using the same rationale, the sonication step was also applied to oyster DT after enzymatic elution of viral particles, before nuclease treatment or before PEG precipitation (Atmar et al., 1995). The impact of some added purification steps may not increase NoV concentrations, but by the elimination of background nucleic acids, they may increase the ratio of NoV reads and the quality of contigs obtained after NGS. In our hands, the PEG approach was already found valuable for stool or sewage samples by allowing longer NoV contig recovery and thus better genotyping (Strubbia et al., 2019). PEG precipitation has also been found useful for the recovery of spiked murine norovirus and human adenovirus, and for NoV metabarcoding, while a combination of centrifugation, filtration and chloroform treatment was found efficient to analyze the virome of environmental samples (Bibby and Peccia, 2013; Hjelmsø et al., 2017; Kazama et al., 2017). Indeed different sample types may need adapted approaches to optimize virus recovery since for instance non-specific approaches such as cryo-fragmentation or mechanical mikro-dismembrator on food samples failed to detect NoV in strawberries (Wylezich et al., 2018). The finding of two NoV reads out of 28,856,294 obtained sequence reads from berries implicated in an outbreak and positive for NoV according to the NoV detection ISO method, showed that improvement is needed (Bartsch et al., 2018).



Since the study is based on the comparison of three individual sequence runs, we used one sewage extract as reference sample to control for variability in deep sequencing. We found this approach valuable as it allowed us to verify that all runs provide similar results. As different aliquots of the library were used, we cannot expect to identify exactly the same sequences, but the global sequence distributions were comparable. This approach also showed that more reads allowed the identification of more strains, an important point when dealing with environmental samples with low contamination levels. Indeed, we found a correlation between the NoV concentrations and the number of reads in sewage samples. This was also verified with the oyster DT extracts as NoV concentrations obtained were too low for an agnostic approach, even if the PEG step after the PK enzymatic elution of the virus increased NoV recovery. For these samples, only the use of VirCapSeq-VERT assay yielded enough NoV reads to allow sequence assembly and strain identification. This suggests that our protocols were able to recover NoV particles but more optimizations are needed to improve the recovery of NoV sequences and probably also for other human enteric viruses. The need to enrich for viral sequences was also needed in some clinical samples confirming the difficulty for viral sequencing (Wylie et al., 2018).

One of the goals of this study was to describe NoV diversity in sewage samples and in artificially contaminated oyster samples. Several NoV strains were identified both in sewage and oyster

samples (such as GI.1, GI.4, and GI.7), while some GI.1 was only detected in DT. Few years ago, we demonstrated that oysters are not passive filters but they can actively select some strains, possible explanation for some greater implication of NoV GI strains in oyster-related outbreaks compared to other ways of transmission (Le Guyader et al., 2012; Verhoef et al., 2015; Yu et al., 2015). To fully understand the role played by ligands it is important to have methods describing the viral diversity in oysters, including the presence of other human enteric viruses. Regarding NoV, it was surprising to find only few sequences of NoV GI.4 in sewage as it is known to be the dominant cause of NoV outbreaks and thus abundant in the environment especially during winter season (de Graaf et al., 2016). This is unlikely to be due to a bias induced by the protocol used, as the PEG method has already been found efficient to characterize GI.4 NoV but further investigation has to be done (Kazama et al., 2017).

To date, no studies have been conducted using metagenomics analysis on contaminated oysters in relation to NoV contamination. Although the present work is based on artificially contaminated oysters, it represents a first step on developing methods to detect NoV from oysters' DT using deep sequencing technologies. Further developments are still necessary to enrich for NoV particles during sample treatment and virus recovery as well as to reduce the host background in order to improve deep sequencing analysis and to obtain longer NoV sequences, necessary for accurate NoV classification. Despite low NoV

concentrations in oysters, they have been implicated in some food borne outbreaks (Polo et al., 2016). Thus, developing sensitive methods for oyster analysis is important to clarify their role in NoV transmission. This work also brings further evidences of the role played by oysters on virus selection and thus will contribute to the understanding of the molecular epidemiology of norovirus. Such approach can be later extended to other food that represent a risk for consumers (Forbes et al., 2017; Cocolin et al., 2018).

DATA AVAILABILITY STATEMENT

The short-reads data for this study has been deposited in the European Nucleotide Archive (ENA) <http://www.ebi.ac.uk/ena/dta/view/PRJEB34617> (ERS3781162–3781182). All full genome sequences were deposited in GenBank and are available under the accession numbers MK956173–MK956178 for NoV genogroup I, numbers MK956197–MK956200 for

genogroup II, and numbers MN510436–MN510441 for astrovirus and sapovirus.

AUTHOR CONTRIBUTIONS

All authors contributed to the work either to design the study, to perform the experiments, to analyze the data, to discuss the results, and to write the manuscript.

FUNDING

This work was supported by the European Community's Horizon 2020 Research and Innovation Program under the COMPARE project, grant agreement N° 643476. MP was supported by Marie Skłodowska-Curie Individual Fellowship, funded by European Union's Horizon 2020 research and innovation programme (grant agreement no. 799417).

REFERENCES

- Adriaenssens, E. M., Farkas, K., Harrison, C., Jones, D. L., Allison, H. E., and McCarthy, A. J. (2018). Viromic analysis of wastewater input to a river catchment reveals a diverse assemblage of RNA viruses. *mSystems* 3:e25-18. doi: 10.1128/mSystems.00025-18
- Atmar, R. L., Neill, F. H., Romalde, J. L., Le Guyader, F., Woodley, C. M., Metcalf, T. G., et al. (1995). Detection of norwalk virus and hepatitis a virus in shellfish tissues with the PCR. *Appl. Environ. Microbiol.* 61, 3014–3018.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comp. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Bartsch, C., Höper, D., Made, D., and John, R. (2018). Analysis of frozen strawberries involved in a large norovirus gastroenteritis outbreak using next generation sequencing and digital PCR. *Food Microbiol.* 76, 390–395. doi: 10.1016/j.fm.2018.06.019
- Bibby, K., and Peccia, J. (2013). Identification of viral pathogen diversity in sewage sludge by metagenome analysis. *Environ. Sci. Technol.* 47, 1945–1951. doi: 10.1021/es305181x
- Bisseux, M., Colombet, J., Mirand, A., Roque-Alfonso, A.-M., Abranel, F., Izopet, J., et al. (2018). Monitoring human enteric viruses in wastewater and relevance to infections encountered in the clinical setting: a one-year experiment in central France, 2014 to 2015. *Eur. Surveill.* 23:7. doi: 10.2807/1560-7917.ES.2018.23.7.17-00237
- Cocolin, L., Mataraga, M., Bourdichon, F., Doulgeraki, A., Pilet, M. F., Jagadeesan, B., et al. (2018). Next generation microbiological risk assessment meta-omics: the next need for integration. *Int. J. Food Microbiol.* 287, 10–17. doi: 10.1016/j.ijfoodmicro.2017.11.008
- Cotten, M., Munnink, B. O., Canuti, M., Deijs, M., Watson, S. J., Kellam, P., et al. (2014). Full genome virus detection in fecal samples using sensitive nucleic acid preparation, deep sequencing, and a novel iterative sequence classification algorithm. *PLoS One* 9:e93269. doi: 10.1371/journal.pone.0093269
- da Silva, A., Le Saux, J.-C., Parnaudeau, S., Pommepuy, M., Elimelech, M., and Le Guyader, F. S. (2007). Evaluation of removal of noroviruses during wastewater treatment, using real-time reverse transcription-PCR: different behaviors of genogroups I and II. *Appl. Environ. Microbiol.* 73, 7891–7897. doi: 10.1128/aem.01428-07
- de Graaf, M., van Beek, J., and Koopmans, M. P. G. (2016). Human norovirus transmission and evolution in a changing world. *Nat. Rev. Microbiol.* 14, 421–433. doi: 10.1038/nrmicro.2016.48
- Endoh, D., Mizutani, T., Kirisawa, R., Maki, Y., Saito, H., Kon, Y., et al. (2005). Species-independent detection of RNA virus by representational difference analysis using non-ribosomal hexanucleotides for reverse transcription. *Nucleic Acids Res.* 33, 1–11. doi: 10.1093/nar/gni064
- Ettayebi, K., Crawford, S. E., Murakami, K., Broughman, J. R., Karandikar, U., Tenge, V. R., et al. (2016). Replication of human noroviruses in stem cell-derived human enteroids. *Science* 353, 1387–1393. doi: 10.1126/science.aaf5211
- Fernandez-Cassi, X., Timoneda, N., Partinez-Puchol, S., Rusinol, M., Rodriguez-Manzano, J., Figuerola, N., et al. (2018). Metagenomics for the study of viruses in urban sewage as a tool for public health surveillance. *Sci. Total Environ.* 618, 870–880. doi: 10.1016/j.scitotenv.2017.08.249
- Flaviani, F., Schroeder, D. C., Lebet, K., Balestreri, C., Highfield, A. C., Schroeder, J. L., et al. (2018). Distinct oceanic microbiomes from viruses to protists located near the antarctic circumpolar current. *Front. Microbiol.* 9:1474. doi: 10.3389/fmicb.2018.01474
- Forbes, J. D., Knox, N. C., Ronholm, J., Pagotto, F., and Reimer, A. (2017). Metagenomics: the next culture-independent game changer. *Front. Microbiol.* 8:1069. doi: 10.3389/fmicb.2017.01069
- Fumian, T. M., Fioretti, J. M., Lun, J. H., dos Santos, I. A. L., White, P. A., and Miagostovich, M. P. (2019). Detection of norovirus epidemic genotypes in raw sewage using next generation sequencing. *Environ. Internat.* 123, 282–291. doi: 10.1016/j.envint.2018.11.054
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010
- Hata, A., Inaba, M., Katayama, H., and Furumai, H. (2017). Characterization of natural organic substances potentially hindering RT-PCR-based virus detection in large volumes of environmental water. *Environ. Sci. Technol.* 51, 13568–13579. doi: 10.1021/acs.est.7b000306
- Hjelmsø, M. H., Hellmer, M., Fernandez-Cassi, X., Timoneda, N., Lukjancenko, O., Seidel, M., et al. (2017). Evaluation of methods for the concentration and extraction of viruses from sewage in the context of metagenomic sequencing. *PLoS One* 12:e0170199. doi: 10.1371/journal.pone.0170199
- Kazama, S., Miura, T., Masago, Y., Konta, Y., Tohma, K., Manaka, T., et al. (2017). Environmental surveillance of norovirus genogroups I and II for sensitive detection of epidemic variants. *Appl. Environ. Microbiol.* 83:e03406-16. doi: 10.1128/AEM.03406-16
- Kroneman, A., Vega, E., Vennema, H., Vinje, J., White, P. A., Hansman, G., et al. (2013). Proposal for a unified norovirus nomenclature and genotyping. *Arch. Virol.* 158, 2059–2068. doi: 10.1007/s00705-013-1708-5
- Kroneman, A., Vennema, H., Deforche, K., Avoort, H.v.d., Peñaranda, S., Oberste M. S., et al. (2011). An automated genotyping tool for enteroviruses

- and noroviruses. *J. Clin. Virol.* 51, 121–125. doi: 10.1016/j.jcv.2011.03.006
- Langmead, B., and Salzberg, S. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/NMETH.1923
- Le Guyader, F. S., Atmar, R. L., and Le Pendu, J. (2012). Transmission of viruses through shellfish: when specific ligands come into play. *Curr. Opin. Virol.* 2, 103–110. doi: 10.1016/j.coviro.2011.10.029
- Le Guyader, F. S., Le Saux, J.-C., Ambert-Balay, K., Krol, J., Serais, O., Parnaudeau, S., et al. (2008). Aichi virus, norovirus, astrovirus, enterovirus and rotavirus involved in clinical cases from a french oyster-related gastroenteritis outbreak. *J. Clin. Microbiol.* 46, 4011–4017. doi: 10.1128/JCM.01044-08
- Maalouf, H., Schaeffer, J., Parnaudeau, S., Le Pendu, J., Atmar, R. L., Crawford, S. E., et al. (2011). Strain-dependent norovirus bioaccumulation in oysters. *Appl. Environ. Microbiol.* 77, 3189–3196. doi: 10.1128/AEM.03010-10
- Nieuwenhuijse, D., and Koopmans, M. P. G. (2017). Metagenomic sequencing for surveillance of food- and waterborne viral diseases. *Front. Microbiol.* 8:230. doi: 10.3389/fmicb.2017.00230
- Nooij, S., Schmitz, D., Vennema, H., Kroneman, A., and Koopmans, M. P. G. (2018). Overview of virus metagenomic classification methods and their biological applications. *Front. Microbiol.* 9:749. doi: 10.3389/fmicb.2018.00749
- Oechslin, C. P., Lenz, N., Liechti, N., Ryter, S., Agyeman, P., Bruggmann, R., et al. (2018). Limited correlation of shotgun metagenomics following host depletion and routine diagnostics for viruses and bacteria in low concentrated surrogate and clinical samples. *Front. Cell. Infect. Microbiol.* 8:375. doi: 10.3389/fcimb.2018.00375
- Osunmakinde, C. O., Selvarajan, R., Sibanda, T., Mamba, B. B., and Msagati, T. A. M. (2018). Overview of trends in the application of metagenomic techniques in the analysis of human enteric viral diversity in africa's environmental regimes. *Viruses* 10:429. doi: 10.3390/v100800429
- Polo, D., Schaeffer, J., Fournet, N., Le Saux, J.-C., Parnaudeau, S., McLeod, C., et al. (2016). Digital PCR for quantifying norovirus in oysters implicated in outbreaks. France. *Emerg. Infect. Dis.* 22, 2189–2191. doi: 10.3201/eid2212.160841
- Polo, D., Schaeffer, J., Teunis, P., Buchet, V., and Le Guyader, F. S. (2018). Infectivity and RNA persistence of a norovirus surrogate, the Tulane virus, in oysters. *Front. Microbiol.* 9:716. doi: 10.3389/fmicb.2018.00716
- Sano, D., Amarasiri, M., Hata, A., Watanabe, T., and Katayama, H. (2016). Risk management of viral infectious diseases in wastewater reclamation and reuse: review. *Environ. Intern.* 91, 220–229. doi: 10.1016/j.envint.2016.03.001
- Santiana, M., Ghosh, S., Ho, B. A., Rajasekaran, V., Du, W.-L., Mutsafi, Y., et al. (2018). Vesicle-cloaked virus clusters are optimal units for inter-organismal viral transmission. *Cell Host Microbe*. 24, 208–220. doi: 10.1016/j.chom.2018.07.006
- Shieh, Y. S. C., Calci, K. R., and Baric, R. S. (1999). A method to detect low levels of enteric viruses in contaminated oysters. *Appl. Environ. Microbiol.* 65, 4709–4714.
- Strubbia, S., Phan, M. V. T., Schaeffer, J., Koopmans, M. P. G., Cotten, M., and Le Guyader, F. S. (2019). Characterization of norovirus and other human enteric viruses in sewage and stool samples through random primed next generation sequencing. *Food Env. Virol.* doi: 10.1007/s12560-019-09402-3 [Epub ahead of print].
- van Beek, J., de Graaf, M., Al-Hello, H., Allen, D. J., Ambert-Balay, K., Botteldoorn, N., et al. (2018). Molecular surveillance of norovirus, 2005–16: an epidemiological analysis of data collected from the NoroNet network. *Lancet Inf. Dis.* 18, 545–553. doi: 10.1016/S1473-3099(18)30059-8
- Verhoef, L., Hewitt, J., Barclay, L., Ahmed, S. M., Lake, R., Hall, A. J., et al. (2015). Norovirus genotype profiles associated with foodborne transmission, 1999–2012. *Emerg. Infect. Dis.* 21, 592–599. doi: 10.3201/eid2104.141073
- Watson, S. J., Welkers, M. R. A., Depledge, D. P., Coulter, E., Breuer, J. M., de Jong, M. D., et al. (2013). Vial population analysis and minority-variant detection using short read next-generation sequencing. *Phil. Trans. R. Soc. B.* 368:20120205. doi: 10.1098/rstb.2012.0205
- Wylezich, C., Papa, A., Beer, M., and Hoper, D. (2018). A versatile sample processing workflow for metagenomic pathogen detection. *Nat. Sci. Rep.* 8:13108. doi: 10.1038/s41598-018-31496-1
- Wylie, K. M., Wylie, T. N., Buller, R., Herter, B., Cannella, M. T., and Storch, G. A. (2018). Detection of viruses in clinical samples by use of metagenomic sequencing and targeted sequence capture. *J. Clin. Microbiol.* 56:e1123-18. doi: 10.1128/JCM.01123-18
- Wylie, T. N., Wylie, K. M., Herter, B. N., and Storch, G. A. (2015). Enhanced virome sequencing using targeted sequence capture. *Genome Res.* 25, 1910–1920. doi: 10.1101/gr.191049.115
- Yu, Y., Cai, H., Hu, L., Lei, R., Pan, Y., Yan, S., et al. (2015). Molecular epidemiology of oyster-related human noroviruses and their global genetic diversity and temporal-geographical distribution from 1983 to 2014. *Appl. Environ. Microbiol.* 81, 7615–7624. doi: 10.1128/AEM.01729-15

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

Copyright © 2019 Strubbia, Schaeffer, Oude Munnink, Besnard, Phan, Nieuwenhuijse, de Graaf, Schapendonk, Wacrenier, Cotten, Koopmans and Le Guyader. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

2.4 DISCUSSION

Il y a un nombre limité de protocoles publiés pour l'analyse metagénomique et n'existe pas une méthode universelle. L'approche utilisée doit impérativement être adaptée en fonction de la matrice analysée et de la question scientifique. Notre objectif était d'évaluer différents protocoles d'élution et de concentration des NoVs pour favoriser leur détection lors du séquençage metagénomique.

Pour les échantillons d'eaux usées, deux méthodes ont été utilisées, appelées méthode PEG et Pyro-PEG. La première est basée sur l'acidification de l'échantillon afin de favoriser l'attachement de la capsid virale aux particules en fonction de leur point isoélectrique avant concentration par addition du Polyéthylène glycol (PEG) (Atmar *et al.*, 1995), alors que l'autre méthode comprend tout d'abord une étape d'élution chimique utilisant du pyrophosphate de sodium associée à une étape de sonication, avant la concentration par PEG (Bisseux *et al.*, 2018). Cette étape de sonication favorise la dissociation des agglomérats, améliore l'élution des particules virales à partir de fragments organiques et réduit la quantité de bactéries (Ettayebi *et al.*, 2016 ; Santiana *et al.*, 2018).

La méthode Pyro-PEG s'est révélée être la plus efficace permettant une meilleure récupération des NoVs lors de l'étape d'élution. Le nombre de particules NoVs GI et GII ont augmenté de 2 et 1 log respectivement en utilisant la méthode Pyro-PEG par rapport à la méthode PEG. Cette concentration plus élevée a permis d'obtenir un nombre plus élevé de reads et des contigs de tailles plus longues, avec six génomes complets identifiés pour NoVs GI, et trois séquences de NoVs GII couvrant plus de la moitié du génome identifié. De plus, un petit nombre de reads appartenant au GIV a été obtenu à partir des échantillons traités avec la méthode Pyro-PEG, mais pas avec la méthode PEG.

Deux méthodes d'élution ont été appliquées aux échantillons d'huitre. La première basée sur une digestion enzymatique des tissus digestifs utilisant la protéinase K (méthode PK), et une deuxième, appelée méthode PK-PEG, qui associe une étape de concentration des particules virale par PEG suite au traitement par PK. Cette deuxième approche a permis d'obtenir des concentrations de NoVs plus élevées par rapport à la méthode PK. Cependant, cette optimisation de la méthode d'élution des particules virales s'est avérée insuffisante pour une analyse metagénomique concluante. En effet, seulement 21 reads de NoVs ont été produits, montrant la nécessité d'une approche metagénomique ciblée. L'utilisation du système de capture VirCapSeq-VERT, a permis de séquencer un nombre de reads suffisant pour poursuivre l'analyse et générer un total de 57 contigs de plus de 500nt dans les échantillons traités selon la méthode PK-PEG, et 19 contigs avec la méthode PK.

Une plus grande diversité de souches du GI plutôt que GII a été détectée dans les échantillons d'eaux usées et parallèlement dans les tissus d'huitre.

Plusieurs souches ont été identifiées dans les eaux et dans les huitres, alors que des souches de NoVs GI.1 étaient identifiées seulement dans les tissus d'huitre confirmant la sélection active souche-dépendante opérée par les huitres lors de la filtration, expliqué par la présence de sites de liaisons similaires aux HBGA chez l'homme (Maalouf *et al.*, 2011 ; Le Guyader *et al.*, 2012).

Bien que les méthodes ont été développées pour cibler les souches de NoV, l'approche metagénomique a permis de détecter un certain nombre de contigs appartenant à d'autres familles de virus entériques humains, telles que *Reoviridae*, *Picornaviridae*, *Caliciviridae* et *Astroviridae*. Le nombre de reads obtenu était beaucoup plus élevé pour toutes les familles virales

détectées dans les échantillons traités avec la méthode Pyro-PEG, pour les eaux usées, et PK-PEG pour les huitres.

La diversité des génotypes de NoVs identifiés dans l'échantillon de référence au cours des trois runs était comparable. Toutefois, un nombre plus élevé de reads ayant été obtenu lors du troisième run de séquençage, cela a permis l'identification des souches GI.3 et GII.P16 non détectées dans les analyses 1 et 2. Cela confirme l'importance de générer une quantité des reads suffisante pour pouvoir détecter la variabilité des souches présentes dans l'échantillon.

2.5 CONCLUSIONS ET PERSPECTIVES

Lors de l'optimisation des méthodes d'élution et de concentration des virus, une approche métagénomique peut permettre d'évaluer les impacts au-delà de la simple quantification des acides nucléiques extraits. En effet, à concentration égale, certaines étapes d'élution et de purification peuvent avoir un effet sur la qualité du matériel génétique et de ce fait sur la qualité et la taille des contigs obtenues après le séquençage haut débit. De la même façon certains traitements peuvent ne pas augmenter les concentrations de NoVs, mais avoir une action plus efficace sur l'élimination des acides nucléiques qui constituent le bruit de fond (hôte, bactéries..) et ainsi augmenter le rapport des reads de NoVs sur la totalité de reads obtenus. Pendant cette étude l'approche Pyro-PEG pour le traitement des eaux usées s'est révélée être prometteur. L'efficacité et la bonne reproductibilité de cette méthode ont été montrées par l'augmentation proportionnelle existante entre la concentration de NoVs et le nombre de reads.

Une retombée importante de ce travail a été la corrélation entre l'obtention d'un plus grand nombre de reads et l'identification d'une plus large diversité de NoVs, point important lorsqu'il s'agit d'échantillons environnementaux faiblement contaminés. Ce point est même critique dans l'analyse des huitres pour lesquelles l'utilisation d'un système de capture s'est révélée indispensable.

Le traitement par PK-PEG a permis l'identification d'un nombre important de reads et une bonne variabilité des souches de NoVs. Toutefois le pourcentage de reads de NoVs par rapport aux reads totaux dans l'échantillon est encore très faible (<0.1%) mettant en évidence la nécessité d'optimiser ultérieurement la méthode d'élution et de concentration des NoVs ou alors de développer une approche visant à réduire le bruit de fond (beaucoup de reads étant identifiés comme du génome de l'huitre).

L'utilisation d'un échantillon de référence nous a permis de comparer les trois runs de séquençage et d'observer un nombre total de reads similaire dans les trois occasions. En termes de variabilité de souches identifiées, considérant la variabilité de contamination l'identification des mêmes séquences était peu envisageable, toutefois les distributions de séquences globales étaient comparables. Dans le futur, des contrôles négatifs pourront être intégrés dans le run afin de vérifier l'absence de contaminations pendant le traitement des échantillons ou lors du séquençage haut débit.

ARTICLE III

3.1 INTRODUCTION A L'ETUDE

Les mollusques bivalves, tels que les huîtres cultivées dans les environnements côtiers, sont considérés depuis de nombreuses années comme un aliment à haut risque. Si les réglementations actuelles basées sur des indicateurs bactériens permettent de réduire efficacement les épidémies dues à la contamination bactérienne, ces mollusques restent en effet l'un des aliments les plus fréquemment identifiés dans les épidémies virales (Yu *et al.*, 2015). Différents facteurs, tels que leur mode de consommation principalement cru et leur capacité à accumuler des virus, expliquent leur rôle dans les épidémies virales, la plupart dues à une contamination par des norovirus (NoVs) (McLeod *et al.*, 2017; Venugopal *et al.*, 2018; Bellou *et al.*, 2013). Appartenant à la famille des *Caliciviridae*, les NoV sont la principale cause de gastro-entérite aiguë chez l'homme et sont donc fréquemment détectés dans les eaux usées (Atmar *et al.*, 2018; Sano *et al.*, 2016).

Le séquençage métagénomique appliqué à une grande variété d'échantillons a démontré son intérêt dans divers domaines (Kallies *et al.*, 2019; Giu *et al.*, 2019; Alberti *et al.*, 2017). Cependant, malgré l'abondance de virus sur la terre, le virome est encore peu connu (Rosani *et al.*, 2019).

Dans la métagénomique virale, la forte abondance de séquences encore inconnues et la prédominance de séquences appartenant à l'hôte ou au microbiote, compliquent l'analyse métagénomique (Hjelmsø *et al.*, 2017 ; Nooij *et al.*, 2018 ; Bibby *et al.*, 2013). Une élimination efficace autant que possible des acides nucléiques de l'hôte et des bactéries, suivie d'un traitement à la nucléase pour éliminer les acides nucléiques libres, sont des étapes indispensables pour identifier la présence des virus entériques humains, qui possèdent pour la plupart des petits génomes à ARN. Cela signifie qu'une conversion en ADN et des étapes d'enrichissement ou d'amplification sont également nécessaires.

Des hexamères aléatoires, appelés communément *random*, sont souvent utilisés pour la synthèse du double brin d'ADN (New England Biolabs (NEB), USA). En 2005, Endoh et collègues, ont sélectionnés un pool d'hexamères non ribosomiques favorables à la retro transcription du virus à ARN. L'utilisation de ces hexamères s'est révélée utile pour la détection de virus de mammifères indépendamment de l'espèce échantillonné (Endoh *et al.*, 2005).

Des systèmes de captures ont également été développés afin de favoriser la métagénomique virale et permettant d'enrichir l'extrait d'acides nucléiques à l'aide d'une capture de séquence ciblée (Brown *et al.*, 2016 ; Wylie *et al.*, 2018).

Enfin, pour exploiter pleinement le potentiel de la métagénomique appliquée aux échantillons faiblement contaminés, des contrôles de qualité et de bonnes pratiques de laboratoire sont essentiels. L'ARN est sensible à la dégradation et peut être facilement perdu au cours des différentes étapes du protocole. Cependant, le principal défi consiste à prévenir la contamination, comme cela a déjà été observé pour de nombreuses autres méthodes de détection moléculaire. Des réplicats techniques et/ou biologiques peuvent en outre être utilisés pour vérifier la bonne répétabilité de détection et pallier aux faibles concentrations virales, permettant de séquencer plusieurs fois le même échantillon et de fusionner les données de sortie suite à l'analyse métagénomique (au niveau des contigs assemblés).

3.2 OBJECTIF DE L'ÉTUDE

L'objectif de ce travail était l'évaluation de la diversité des virus entériques humains, et la caractérisation des souches de NoVs accumulés par des huîtres contaminées par un échantillon d'eau usée. L'impact de différentes amorces utilisées lors de l'étape initiale de transcription inverse a été évalué comparant les hexamères random, le set sélectionné par Endoh *et al.* (2005) et un set conçu en interne (*In-house designed*, I-HD) pour réduire la transcription des ARN des huîtres et des moules au cours de la préparation de l'ADNc en ciblant les séquences des NoVs.

Trois lots d'huîtres ont été utilisés pour étudier l'accumulation souche-dépendante dans les tissus digestifs d'huître : *Ostrea edulis* (*O. edulis*) et *Crassostrea gigas* (*C. gigas-2*) élevées dans une plateforme scientifique avec de l'eau de mer traitée aux UV et nourries avec du phytoplancton et *Crassostrea gigas* (*C. gigas-1*) acheté auprès d'un producteur local (ouest de la France).

Chaque échantillon a été séquencé trois fois afin d'évaluer l'impact des réplicats techniques dans l'analyse de la diversité des NoV par métagénomique.

Plusieurs contrôles négatifs en NoVs ont été ajoutés, incluant des tissus d'huîtres non contaminées (pour chaque lot), l'eau UltraPure™ utilisée en laboratoire pour le traitement des échantillons et l'eau de mer.

Une méthode d'élution et concentration précédemment mise au point a été utilisée : la méthode Pyro-PEG pour les échantillons d'eau et PK-PEG pour les tissus d'huître.

38 genomes and six sequences longer than 3600 bases were obtained allowing a precise strain
39 identification. The use of technical triplicates was found valuable to increase the chances to sequence
40 viral strains present at low concentrations.

41

42

43

44 **keywords:** norovirus, oysters, metagenomic, human enteric viruses.

45

46

47

48 **1. Introduction**

49 Human virus transmission through food consumption has been identified since many years and
50 preventative measures have been undertaken in many countries. Nonetheless, these remain imperfect,
51 and the international trade increases the risk of dissemination of viral pathogens (Li et al., 2018). The
52 main sources of food contamination with viruses are infected food handlers, zoonotic transmission
53 (essentially for hepatitis E virus) or contact with sewage-contaminated waters during the production
54 process (Bosch 2018). For the two first cases, a test, targeting the virus excreted by the food handler or
55 by the animal, may directly identify the contaminant in the foodstuff. However, when dealing with
56 food contaminated during production, the virus identification may be more challenging. Indeed, a
57 variety of human enteric viruses are excreted by the human population and may contaminate the
58 environment mainly through sewage discharge (Metcalf et al., 1995). With the increasing demands of
59 water for agriculture or human activities, wastewater re-use will be necessary, increasing the risk for
60 environmental virus pollution, and subsequent viral transmission through contaminated food (Sano et
61 al., 2016). Being able to identify the virus contaminating a food sample will be of primary importance
62 for risk management and consumer safety (Nieuwenhuijse et al., 2017).

63 Bivalve molluscan shellfish such as oysters are consumed worldwide and considered as a
64 source of important nutrient (Venugopal et al., 2018). However, they have been identified for many
65 years as a high-risk food (Metcalf et al., 1995). Regulations based on bacterial indicators, have
66 lowered the frequency of outbreaks due to bacterial contamination, but oyster remain one of the most
67 frequently associated food in foodborne viral outbreaks (Yu et al., 2015). Although some reporting bias
68 in outbreaks identification contribute to increase their image of risky food, different factors, such as
69 their mode of consumption (mainly raw) and their capacity to accumulate viruses by filtering large
70 amount of water, explain the documented implication of shellfish in viral outbreaks on recurrent
71 occasions (Bellou et al., 2013, Yu et al., 2015, McLeod et al., 2017, Venugopal et al., 2018). Most of
72 these outbreaks are due to Human norovirus (HuNoV) contamination. Belonging to the *Caliciviridae*
73 family, HuNoV rank as the leading cause of acute gastroenteritis in human, and thus are frequently
74 detected in human sewage (Sano et al., 2016, Atmar et al., 2018). Yet, other human enteric viruses
75 such as hepatitis A virus or rotavirus have also been implicated in shellfish borne outbreaks (Kittigul
76 et al., 2014, Boxman et al., 2016). Thus, following a flooding event close to a shellfish production
77 lagoon, we detected up to seven different virus strains in oyster and clinical samples, including an
78 Aichi strain identified for the first time in Europe in an oyster sample (Le Guyader et al., 2008).

79 Metagenomics refer to the description of all nucleic acid sequences present in a sample
80 allowing the description of bacteria, archaea, eukaryotes or viruses (Giu et al., 2019, Kallies et al.,
81 2019). However, despite the abundance of viruses on the earth, the virome is still poorly described
82 (Rosani et al., 2019). As a catch-all, agnostic and sensitive approach potentially allowing the detection
83 and identification of the whole viral genomic material, metagenomic sequencing holds many promises

84 for the surveillance of viruses in food and water (Nieuwenhuijse et al.,2017). Nevertheless, several
85 issues remain, owing to both the diversity of potential viral contaminants and the very low level of
86 contamination in complex food matrices (Nieuwenhuijse et al.,2017). Indeed, viral metagenomics is
87 generally performed by removing as much host, matrix and bacterial nucleic acids as possible
88 followed by nuclease treatment to remove free nucleic acids. These steps are of primary importance to
89 identify human enteric viruses, most of them having a small RNA genome, among the vast amount of
90 DNA derived from the host or the food matrix. Besides, most viruses of interest bearing an RNA
91 genome require a conversion to cDNA. Finally, enrichment or amplification steps are also frequently
92 needed to increase the recovery of viral sequences (Wylie et al., 2018). An approach combines cDNA
93 synthesis and enrichment in virus sequences by using primers selected to avoid reverse-transcription
94 of the host ribosomal RNA while allowing the priming of mammalian viruses (Endoh et al., 2005).
95 Such non-ribosomal primers have been successfully applied to the sequencing of mammalian viruses
96 in clinical samples (Cotten et al., 2014), and could contribute to the detection of human viruses in
97 food.

98 The work presented here first aimed at evaluating the selection of primers used for reverse-
99 transcription in order to assess the diversity of human enteric viruses in artificially contaminated
100 oysters. Secondly, we wanted to explore the use of technical replicates to analyze HuNoV diversity by
101 metagenomic sequencing.

102

103 **2. Materials and methods**

104

105 *2.1. Oyster batches*

106

107 Two large oyster batches raised in a scientific farm with UV treated seawater and fed with
108 phytoplankton produced in the farm were used: one batch of *Ostrea edulis* (called O. edulis) and one
109 of *Crassostrea gigas* (C. gigas-2). A third batch of *Cassostrea gigas* oysters (C. gigas-1) purchased
110 from a producer (western France) was included in the study.

111 Immediately after arrival in the laboratory, *Escherichia coli*, total flora and total marine bacteria were
112 counted in shellfish flesh as described previously. Briefly, 15 oysters of each batch were homogenized
113 (1:10) with Peptone Salt Solution (Oxoid, France) and further two one-Log dilutions were performed.
114 Each of these dilutions were then inoculated on Plate Count Agar (Oxoid) and Marin agar
115 (FisherScientific, France) on duplicate and incubated for 72 h at 30°C and 20°C respectively.
116 *Escherichia coli* were quantified according to the standard NF V08-106 (AFNOR, 2010).

117

118 *2.2. Bioaccumulation*

119

120 A composite 24-h raw sewage sample was collected at the inlet of a waste-water treatment
121 plant in western France, titrated for HuNoV concentration and used to contaminate three aquariums
122 filled with 22 L of natural seawater, to reach a final concentration of 10^7 HuNoV (GI + GII) RNA
123 copies/L. The water temperature was kept at 12 ± 1 °C.

124 Each tank received 100 oysters of each batch for contamination during 24 h. Twenty-five oysters of
125 each batch were kept in 5-L of clean seawater as negative controls. After 24 h, all oysters were
126 collected and flesh weights were recorded to calculate the allometric coefficient, used to describe the
127 physiology of the animal (Polo et al., 2018). Digestive tissues (DT) were dissected using a sterile
128 knife, pooled, chopped and distributed in 2 g aliquots immediately stored at -80°C until analysis.

129

130 2.3. *Quality controls*

131

132 Non-bioaccumulated oysters, seawater, and UltraPure™ DNase/RNase-Free Distilled Water
133 (TechnoFisher, France) used during the experiment were verified for the absence of NoV
134 contamination and included in deep-sequencing analysis. All these controls were analyzed once.

135

136 2.4. *Samples preparation*

137

138 Sewage and control seawater samples were treated with the Pyro-PEG method as previously
139 described (Strubbia et al., 2019). Briefly, 10 mM sodium pyrophosphate was added to 40 mL of
140 sample and incubated for 40 min at room temperature under gentle agitation (Bisseux et al., 2018).
141 After sonication, and centrifugation at low speed to eliminate big particles, supernatants were
142 recovered and the pH adjusted to 7. After incubation with polyethylene glycol 6000 (PEG) (Sigma-
143 Aldrich, St-Quentin, France) and centrifugation, the pellet was resuspended in 2 mL of glycine buffer
144 (0.05 M) pH 9.

145 DT aliquots (2 g) were incubated with 2 mL of proteinase K (Thermo Fisher Scientific) for 15 min at
146 37°C and 15 min at 60°C . After sonication (Bandelin UD 2200 with cup-horn adaptor) for 3 min at
147 maximum power (cycles of 1 min with 1 min on ice), the mixture was centrifugated for 5 min at 3,000
148 x g. Then the supernatant was mixed with 10 mM sodium pyrophosphate (Bisseux et al., 2018) and
149 incubated 40 min at room temperature under gentle agitation. After 20 min of centrifugation at 8,000 x
150 g, the supernatant (3 mL) was recovered and 1.5 mL of a PEG 24% (wt/vol)-sodium chloride (1.2 M)
151 solution was added prior incubation for 1 h at 4°C . After centrifugation for 20 min at 8,000 x g, the
152 pellet was resuspended in 1 mL of glycine buffer (0.05 M) pH 9.

153 All samples were treated with 10 $\mu\text{L}/\text{mL}$ OmniCleave™ Endonuclease (Lucigen Corporation) and 100
154 μL MgCl_2 (100mM) for 1 h at 37°C . As some inhibitors may persists, an additional RNA purification
155 was performed using the RNA Clean & Concentrator™ -5 kit (Zymo Research, Irvine USA). These
156 NA extracts were used for HuNoV quantification and to prepare libraries under technical triplicates.

157

158 *2.5. NoV quantification/detection*

159

160 Each extract was analyzed by real-time RT-PCR using primers and probe targeting the
161 conserved region between ORF1 and ORF2 for HuNoV GI and GII, using a one-step real-time RT-
162 PCR kit (Invitrogen) (Le Guyader et al., 2009).

163

164 *2.6. In-house designed hexamer set*

165

166 Following the method described by Endoh et al. (2005), an In-House Designed (I-HD)
167 hexamer set was selected to lower oyster and mussel rRNA transcription during cDNA preparation. A
168 host rRNA reference set corresponding to 48 sequences of bivalve molluscan shellfish was selected
169 from Silva rRNA database by compiling “ostrea” and “mytilus” sequences (Glockner et al., 2017). A
170 second reference set containing 79 HuNoV sequences from complete or partial genomes available in
171 EBI project Compare ([https://www.ebi.ac.uk.cna/pathognes/xref](https://www.ebi.ac.uk/cna/pathognes/xref)) was selected. A python script
172 (v2.7.9) was used to compute the number of occurrences of the 4096 hexamers in each sequence
173 (reverse complement search) for the two references. The probability was calculated as the number of
174 occurrences in sequence divided by the length of each sequence selected as a reference. This allowed
175 the selection of 166 hexamers (forming the I-HD set) with no (or low) match for oysters and mussels
176 rRNA and with probability between >0.1 and <0.3 to match to HuNoV sequences.

177

178 *2.7. Library preparation and capture-based deep sequencing*

179

180 Technical triplicates were performed for each batch. Two sets of hexamers, random hexamers
181 (New England Biolabs (NEB), USA) and I-HD hexamers were used to prepare the cDNA from the
182 RNA extracted from the three oyster batches, the sewage sample as well as from all controls
183 (seawater, control oysters, and UltraPure™ water). Additionally, RNA from C.gigas-1 batch was also
184 reverse-transcribed using the Endoh hexamer set (Endoh et al., 2005). The reverse transcription was
185 performed using the enzyme SuperScript II (ThermoFisher) and 5 μ L were used to perform a real-time
186 HuNoV GII PCR using primers and probe as described above. A total of 33 cDNA was obtained and
187 organized into three pools for the next step, considering a uniform distribution of technical replicates
188 (Fig. 1). Each pool contained seven cDNA prepared from a bioaccumulated oyster batch, with an even
189 distribution of the hexamer set, two negative oyster cDNA and two libraries prepared from one water
190 sample (either sewage, or seawater or UltraPure™ water sample) (Fig. 1). After double strand DNA
191 synthesis using second strand reaction buffer from NEBNext Ultra RNA Library prep, a physical
192 fragmentation (Ultrasonicator M220, Covaris) was performed for 3 min. To reach the advised
193 quantities preconized by the manufacturer, ten amplification cycles were added to the control sample

194 extracts. Libraries were prepared using the kit KAPA Hyper Prep (Roche), adapter ligation was
195 performed and viral sequences were enriched using VirCapSeq-VERT capture panel (Roche) (Wylie
196 et al., 2015). The capture was done simultaneously on the 3 pools.
197 Sequencing was performed on an Illumina NextSeq 500 using High output kit V2 (Illumina) to
198 generate 2 x 150bp reads.

199

200 2.8. Bioinformatic pipeline

201

202 Metagenomics bioinformatics pipeline included quality trimming, de novo assembly, and
203 sequence alignment and annotation. Illumina adapters were removed from raw reads before trimming
204 from 3' to 5' to reach a Phred score above 25 (TrimGalore v.0.4.1). Trimmed reads were *de novo*
205 assembled using SPAdes v.3.12.0 to obtain contigs (Bankevich et al., 2012). The first step of
206 SPAdes pipeline was error correction (hammer tools) with default setting, then assembly was
207 performed using de Bruijn graph algorithm and a set of k-mer lengths of 21, 33, 55, 77. To ensure a
208 reliable identification, only contigs longer than 500 nt were identified using BLAST v.2.6.0, and only
209 sequences with identity score above 85% were considered during data analysis. The coverages and
210 abundances of selected contigs were further evaluated using Bowtie 2 v2.3.0. (Langmead et al., 2012).
211 Contigs originating from oyster DT sequenced in triplicate were analyzed separately and sequences
212 identified as belonging to *Caliviridae* family were genotyped using HuNoV genotyping tool v.2.0
213 (Kroneman 2011). The high amount of reads produced from sewage samples required a normalization
214 step prior assembling using bbmap v.38.22 to reach a coverage of 40 (Bushnell et al., 2017).

215

216 2.9. Statistical analysis

217

218 The Pearson correlation test was applied to evaluate the relationship between NoV GII C_T
219 values and the number of reads per million (rpm), according to the hexamer set used. RStudio
220 v1.2.1335 was used to calculate the correlation and to plot data.

221

222 3. Results

223

224 3.1. Oyster microbiology and controls

225

226 The mean weight of individual oysters from the two *C. gigas* batches were similar (9 and 9.5 g
227 for *C. gigas*-1 and *C. gigas*-2 respectively) while flat oysters weighted around 5 g. However, the
228 allometric coefficients (flesh weight divided by DT weight) varied from 17 to 19 for the three batches,
229 meaning that their physiological conditions were similar and are thus comparable for the three
230 different bioaccumulations.

231 All three oyster batches were devoid of fecal contamination (concentrations were lower than
232 67 *Escherichia coli* / 100g of total flesh). The marine bacteria concentrations were also identical for all
233 three batches and varied from 1.1×10^7 units forming colony (UFC)/100 g for *C. gigas*-2 batch to $4.9 \times$
234 10^7 UFC/100g to *C. gigas*-1 batch. The only difference was found for total flora with a concentration
235 one Log higher for *C. gigas*-1 batch (5.4×10^5 UFC/100g) compared to concentrations measured in *C.*
236 *gigas* 2 and *O. edulis* batches (5.5×10^4 and 4.8×10^4 UFC/100g respectively) that were raised in clean
237 experimental facilities.

238 HuNoV concentrations in the sewage sample used for the bioaccumulations were 6.7×10^4 and
239 2.9×10^4 for GI and GII respectively. After bioaccumulations, the three oyster batches reached similar
240 NoV contamination levels, as measured by rRT-PCR, with C_T values ranging from 29.8 to 32.7 for
241 both genogroups, corresponding to concentrations from 9.7×10^2 RNAc/g of DT for *O. edulis* batch to
242 2×10^3 RNAc/g of DT for both *C. gigas* batches. For all negative control oyster batches (non-
243 bioaccumulated oysters) and the seawater used for the bioaccumulation, NoV RNA levels were under
244 the limit of detection of the method.

245

246 3.2. Theoretical calculation for the three hexamer sets

247

248 The three hexamer sets used in this study, Endoh, Random and I-HD were evaluated for their
249 theoretical probability to recognize a panel of Human enteric virus sequences, as well as mollusk and
250 rat genomes among eukaryote sequences (Fig. 2). The random set presented the same probability
251 (around 23 %) to match with all sequences selected for our evaluation. Endoh and I-HD sets presented
252 a lower probability to recognize rat sequences than the random hexamer set. As expected the I-HD set
253 did not match with shellfish sequences, while the Endoh set recognized shellfish sequences with a
254 probability around 12 to 15 %. Regarding virus sequences, the Endoh set displayed a large variability
255 with a higher probability to recognize picobirnavirus and rotavirus sequences compared to HuNoV.
256 The I-HD set presented less variability in the probability to match with the different human enteric
257 virus sequences selected, with a slightly higher probability to recognize HuNoV sequences than the
258 Endoh set. Here, we aimed at evaluating the performance of our I-HD set to recover human enteric
259 virus sequences in oyster extracts, by comparing them to random primers on all samples. The Endoh
260 set, which is optimized for mammalian virus in rodent cells, was used to prepare triplicate libraries for
261 only one sample (*C. gigas*-1), as a control.

262

263 3.3. Run outputs

264

265 A total of 432,023,640 raw reads were produced during the Next-Seq500 run, 39% of these
266 reads corresponded to the two libraries prepared with the sewage sample. This high number of reads
267 (167,864,533) obtained from the sewage sample was normalized with BBNorm which induced a loss

268 of 96% of reads, leading to read counts comparable to the other samples. Quality trimming was
269 applied to all samples and eliminated 2% of reads of the sewage sample, and between 1% and 3% of
270 reads from oyster samples (including contaminated oysters and negative controls). The two water
271 controls provided high number of reads presumably due to the additional amplification cycles
272 performed during the library preparation to obtain the same amount of DNA. After trimming the
273 number of reads, the seawater control was lowered by 15% while for the UltraPure water control, the
274 loss represented around 6% of the initial reads.

275

276 3.4. Host abundance and viral reads

277

278 To analyze the impact of the hexamer sets used to prepare the cDNA, we cumulated all the
279 reads obtained from bioaccumulated oyster libraries prepared using either the I-HD or the random
280 hexamer sets (Endo set was excluded as it was used for only one sample) (Fig. 3). Both for Eukaryota
281 and Mollusca reads, the I-HD set gave lower numbers with a greater variability, confirming their
282 capacity to decrease the sequencing of eukaryote and specifically mollusk genomes. In addition, the
283 number of reads targeting virus sequences which was higher using the I-HD set compared to the
284 random set. However, when analyzing reads targeting the HuNoV reads, the random set yielded a
285 higher abundancy than the I-HD set (Fig. 3).

286 The highest amounts of reads attributed to viruses was obtained from the sewage libraries (69 and 63%
287 out of total reads with I-HD and random hexamers, respectively (Table 1). Recovering such a huge
288 amount of reads from one library was not expected, and draws attention on the preparation of mixes
289 into different pools as it may favor one sample and overshadows the others. This also showed the
290 importance of technical triplicates. A low number of mollusk reads was obtained from the seawater
291 libraries (Table 1). This is natural seawater that has been in contact with some mollusks prior
292 sampling. It was more surprising to obtain a high number of reads from the UltraPure water control.
293 This was probably due to non-specific amplification which may have occurred during the additional
294 amplification cycles performed to reach enough DNA concentration during the library preparation, or
295 a tag jump, or due to a possible laboratory contamination during the preparation step.

296

297

298 3.5. Virus diversity in samples

299

300 Sequences of five viral families (*Reoviridae*, *Picornaviridae*, *Picobirnaviridae*, *Caliciviridae*
301 and *Astroviridae*) were identified in the sewage sample and in the three bioaccumulated oyster batches
302 (Fig. 4). In the sewage sample, the two hexamer sets gave similar numbers of reads, with the highest
303 read numbers blasting to the *Astroviridae* family. The same five families were identified in the three
304 oyster batches with the two hexamer sets. The *Picornaviridae* sequences were the less frequently

305 identified sequences (ranging between 0 and 100 rpm) while *Astroviridae* and *Picobirnaviridae*
306 families were the most represented (between 100 and 10000 rpm). The library prepared using the
307 Endoh set, provided the lowest read numbers for all five families. The random set provided the highest
308 read numbers for all five families in agreement with total viral read numbers obtained (Fig. 4). The
309 negative controls included in this study provided no significant read numbers for the five virus
310 families except the *C. gigas-1* negative control. This batch bought to a producer could have been
311 contaminated in the producing area, or it could be a laboratory contamination during the sample
312 preparation or the library preparation, as this negative control was in the pool with the most
313 contaminated sample (sewage).

314

315

316 3.6. Correlation between HuNoV RNA C_T values and number of HuNoV reads

317

318 We evaluated the impact of the set of hexamers used for cDNA synthesis on HuNoV recovery,
319 as measured by rRT-PCR following this first step (Table 1). The random hexamers set provided lower
320 C_T values (meaning higher HuNoV concentrations) for all batches compared to C_T values obtained
321 with the I-HD set. For the *C.gigas-1* nucleic acid extract, the Endoh and I-HD sets provided
322 comparable C_T values. The Endoh set, used to prime cDNA of one sample, provided C_T values in the
323 same range as the I-HD set. Accordingly, the use of random hexamers led to the highest number of
324 HuNoV reads recovered from the *C. gigas-2* and *O. edulis* libraries, but this was less clear for the *C.*
325 *gigas-1* batch. A significant negative correlation between HuNoV reads frequency and C_T value was
326 found when using random hexamers (Pearson correlation value: $r = -0.81$, $p < 0.05$) but not when
327 using the I-HD set (Fig. 5). This statistical analysis was not performed with the Endoh set as data was
328 available for only one sample.

329 No HuNoV reads were obtained from the *O. edulis* or *C. gigas-2* negative control batches.
330 However the *C.gigas-1* negative control libraries allowed the identification of 44 and 179 HuNoV
331 reads, confirming some of the *Caliviridae* reads (Figure 4), despite the absence of HuNoV detection
332 by rRT-PCR. As hypothesized above, this contamination could have occurred during the sample
333 processing, but also in the producing area.

334

335 3.7. Contigs length and HuNoV strain identifications

336

337 HuNoV contig length is an important criterion to select an effective method as longer contigs
338 allow a better strain identification. Reads obtained using the Endoh set for cDNA priming, generated
339 HuNoV contigs shorter than 500 nt, and thus were not considered for further analysis (Fig. 6). The
340 random hexamers set allowed the recovery of the longest HuNoV contigs, for all samples except two
341 contigs obtained from sewage sample. In this sample, the identified contig of 3211nt obtained with I-

342 HD was similar to a 2872 nt contig obtained with the random set. The longest HuNoV contigs were
343 obtained from *O. edulis* libraries, which yielded two whole HuNoV genomes (Fig. 6) (Table 2). From
344 oyster samples, the longest HuNoV contigs obtained using the I-HD set were of 1436, 899 and 1056 nt
345 from *C. gigas*-1, *C. gigas*-2 and *O. edulis* batches respectively. Considering all samples, the I-HD set
346 allowed the identification of 150 and 194 HuNoV GI and GII contigs, while the random hexamers set
347 allowed the identification of 208 and 284 NoV GI and GII contigs. Indeed, when the contig lengths
348 were too short, strain identification was not possible, as observed with libraries built using Endoh set.
349

350 3.8. *HuNoV strain diversity*

351

352 Only contigs obtained from libraries prepared using the random hexamer set were considered
353 for HuNoV genotyping using the online Norovirus typing tool 2.0. For two oyster batches (*C.gigas*-2
354 and *O. edulis*), one library did not provide any contig meeting the criteria set up for the analysis (at
355 least 500 nt length and blast identity score over 85%). For the same two batches, one of the three
356 libraries provided few informative contigs that only allowed genogroup identification (Fig. 7). For the
357 *O. edulis* batch, the contigs obtained for the third replicate were distributed among five HuNoV GII
358 strains in the ORF1 and four in the ORF2 typing region and one GI (both ORF1 and 2). Two complete
359 genomes (one GI and one GII) were identified in one library from the *O. edulis* batch and a third one
360 was almost complete. In this same library, a sequence covering more than half of the genome was also
361 identified (Table 2). For the *C.gigas*-2 batch most of the contigs obtained fell into a non-typing region
362 of HuNoV GII ORF1 and only few contigs allowed the identification of a GII.P16 and a GI.P4 strains.
363 Four sequences covering more than half of the genome were detected, three HuNoV GI and one GII
364 strains. Five other strains longer than 1000 nt could also be identified either in the ORF1 or the ORF2
365 region (Table 2). Libraries obtained from the *C.gigas*-1 batch provided contigs for all three replicates.
366 A GI.P1 sequence was observed in two replicates while the other one only provided genogroup
367 information. GI.P2, GII.P16 and GII.P17 strains were also identified in one of the replicates. For the
368 ORF2 region, a GI.1 sequence was identified in the same two replicates while the other one allowed
369 only genogroup typing (Fig. 7).

370 In the sewage sample, contigs targeting non-typing region of the GII ORF1 and ORF2 were
371 the most prevalent. Thirteen contigs were obtained from the sewage sample, but the longest was less
372 than 3,000 nt and only one allowed genotyping of both the ORF1 and 2 (Table 2).
373 All the strains detected in the sewage sample were detected at least once in one of the oyster batches,
374 while the GI.2[P2] was detected only from oyster samples (full genome in *O. edulis*, and half genome
375 in *C. gigas*-2). Considering the variability observed among the library triplicates, this approach is
376 important to evaluate the diversity of viral strains that may be present in a sample. This may be of
377 particular interest for low contaminated samples such as oyster samples.

378

379 *4. Discussion*

380

381 This study aimed to provide a method to identify HuNoV and other Human enteric viruses
382 among the virome of oysters DT. While nowadays HuNoV constitute the pathogen most frequently
383 implicated in oyster related outbreaks, other viruses such as hepatitis A virus or rotavirus have also
384 been responsible for clinical cases following raw oyster consumption, which highlights the interest of
385 their detection (Bellou et al., 2013, Boxman et al., 2016). Oysters usually get contaminated by filtering
386 seawater contaminated with human sewage, and thus accumulate a large diversity of human enteric
387 viruses. Such diversity and the low copy number of the viral pathogens in contaminated oysters
388 complicate their identification, and the containment and mitigation of foodborne outbreaks (McLeods
389 et al., 2017). Metagenomic approaches, describing all nucleic acid present in a sample, appear as a
390 promising tool to identify the different viruses that may be present without discriminating criteria.
391 The first challenge was to recover HuNoV from oyster tissues. HuNoV, and other enteric viruses, are
392 mainly concentrated in oyster DT where they may bind specific ligands (such as histo-blood group
393 antigen-like glycans, sialic acids) or be retained through non-specific ionic interactions (Di Girolamo
394 et al., 1977, McLeod et al., 2009, Le Guyader et al., 2012). Targeting these tissues improve the method
395 sensitivity and has now been validated worldwide (Atmar et al., 1995, ISO 15216-1, 2017). We used
396 an enzymatic treatment to elute viruses from the DT and we further improved the viral particle
397 recovery using a sonication step and a chemical elution using sodium pyrophosphate before PEG
398 concentration. Sonication was found useful for HuNoV cultivation in enteroids and it may help to
399 disrupt vesicles in which HuNoV may be trapped (Ettayebi et al., 2016, Santiana et al., 2018).
400 Alternatively, it may destroy bacteria to which the virus may bind (Miura et al., 2013). The combined
401 used of sonication and sodium pyrophosphate was found the best method to recover viruses from
402 sewage sludge (Brown et al., 2015). Considering the limited knowledge on the association of HuNoV
403 or other human enteric virus with bacteria, no filtration step was applied after the PEG concentration
404 to prevent any loss of viral particles (Conceicao-Net et al., 2015, Amarasiri et al., 2019).

405 Another challenge faced by metagenomics is the depletion of host or background nucleic
406 acids, that can account for up to 99% of the reads. This issue is encountered in clinical samples either
407 for ethical issues or to improve the sensitivity (Oechslin et al., 2018, Giu et al., 2019), as well as in
408 food samples (Bartsch et al., 2018, Wylezich et al., 2018). The use of non-ribosomal hexanucleotides
409 that favor RNA virus reverse transcription was found useful for species-independent detection of
410 viruses in infected cells (Endoh et al., 2005). Following the same rational, we selected a combination
411 of hexamers with the aim to reduce the binding to oyster and mussels rRNA. This indeed reduced the
412 host background by about 10% compared to random hexamers. However, these I-HD hexamers were
413 less successful to detect our target virus sequences than the random hexamer set, which may be due to
414 the reduced number of sequences from HuNoV and other Human enteric viruses that were used to
415 validate this selection. Thus, to increase the sensitivity and selectivity of our approach towards Human

416 enteric viruses, the future development of this work will be to complete our hexamer selection by
417 extending to more HuNoV strains including new variants, and to other Human enteric virus sequences.
418 Indeed, beside the first step of sample preparation and nucleic acid purification this reverse
419 transcription is the key step for a successful deep sequencing. The control performed by real-time PCR
420 showed that depending on the hexamer set used, a difference of up to 3 C_T can be observed
421 corresponding to a 10 times difference in HuNoV concentration. Such gain is important when looking
422 for low contamination levels as usually found in marketed oysters (EFSA, 2019). In all samples, the
423 use of random hexamers led to a more efficient transcription of HuNoV sequences and to longer
424 contigs allowing a better HuNoV strain identification compared to the two other sets. Indeed,
425 norovirus genotyping requires contigs falling into specific regions of the genome at the end of ORF1
426 and the beginning of ORF2, which is less likely when reads are scarce and contigs, short. While the
427 Endoh set was applied to only one sample, all contigs recovered were short (<500 nt) and did not
428 allowed any HuNoV identification. The I-HD set looks promising as shown by results obtained with
429 the sewage sample, and the reduce mollusk background reads.

430 To fully exploit the potential of metagenomics applied to low contaminated samples, quality
431 controls and good laboratory practices are essential. RNA is sensitive to degradation and may be easily
432 lost during the different steps of the protocol. However, the main challenge is to prevent
433 contamination as already observed for many other molecular detection methods. Despite all
434 precautions (dedicated rooms, filter tips...) we cannot exclude that identification of virus contigs in
435 one of the negative control oyster batch and the UltraPure water resulted from contamination during
436 the process, probably during the library preparation or the flow cell amplification (Schnell et al.,
437 2015). The other negative controls (waters and oysters) presented few HuNoV reads that may be due
438 to PCR errors occurring during the additional amplification cycles needed to reach the 10 ng/ μ L
439 requested to satisfy the library kit instructions. The high number of reads recovered from the random-
440 set primed library for the UltraPure water was unexpected, and among the virus reads no contig
441 matching with Human enteric viruses was identified. None of the contigs detected in our negative
442 controls meet our quality criteria of 500 nt with 85% Blast score (applied for all enteric viruses), and
443 thus they were not considered for data analysis. Considering the huge sequencing capacity of high
444 throughput metagenomic sequencing which is able to detect contaminant from reagents, from past or
445 other experiments undertaken in the laboratory, the inclusion of negative controls that follow all the
446 experiment process is essential (Brown et al., 2016, Nieuwenhuijse et al., 2017, Wylie et al., 2018).

447 In a previous study, a reference sample that we inserted in three separated runs provided
448 similar HuNoV diversity, but with some variation in contigs lengths (Strubbia et al., 2019). In the
449 present study, technical triplicates were used to evaluate the impact of library preparation. Such
450 approach was found valuable to minimize potential bias in aquarium water virome analysis (Kim et
451 al., 2017). Our study confirmed the interest of such replicates as NoV diversity varied from no contig
452 to several sequences assigned to different genotypes, depending on the library considered for one

453 given sample. Replicate libraries may be recommended in the future for low contaminated sample as it
454 enhances the probability to identify rare sequences. When dealing with food samples this is important
455 for further risk analysis. Finding two NoV sequences of less than 200 base pairs in a sample clearly
456 identified as being responsible of an outbreak may be interpreted as a positive argument to destroy the
457 food in question, but it may also be a challenging issue for risk managers in absence of any
458 epidemiological data (Bartsch et al., 2018). Providing solid arguments of repeatable detection in
459 technical or biological replicates may help to increase confidence in this highly sensitive method.

460 Another valuable option to increase method sensitivity towards Human enteric viruses is to
461 enrich the nucleic acid extract using a targeted sequence capture (Brown et al., 2016, Wylie et al.,
462 2018). As mentioned above, HuNoV and most of other human enteric viruses have short RNA
463 genome, making them very difficult to identify among the outnumbering other nucleic acids that may
464 be extracted from the oyster DT. Adding too many steps before nucleic acid extraction may lead to the
465 loss of viral particles or be too selective when based on physical properties (size selection, pH
466 treatment, etc...). An alternative method for enrichment in viral sequences was proposed few years
467 ago by using a targeted sequence capture (Wylie et al., 2015, Brown et al., 2016). We successfully
468 used this capture-based approach in a previous experiment where direct metagenomic detection did
469 not allow HuNoV identification (Strubbia et al., 2019). Here, this capture-based approach was applied
470 to all samples, including the sewage sample used to contaminate the oysters, leading to a massive
471 amount of reads in the two libraries prepared from that sample. Despite the normalization step, the
472 amount of reads produced and the large HuNoV diversity, complicated the *de novo* assembly and
473 impaired the assembly of full HuNoV genome from sewage sample. Detection of viruses can be
474 performed after *de novo* assembly of short reads into longer sequences (contigs) followed by a variety
475 of computational methods to detect known and novel viral sequences. A number of workflows have
476 been published but their selection may be challenging for new users since the bio-informatic pipeline
477 used can make a difference to identify rare sequence or new virus genome (Cotten et al., 2014, Oude
478 Munnink et al., 2016). A proficiency test performed to analyze viral reads highlighted the need to
479 share workflows and to harmonized criteria for their validation (Brown et al., 2015). In a literature
480 review, considering all published workflows for virus metagenomics, a strategy was proposed to select
481 the optimal workflow (Nooij et al., 2018). Such approach is of particular interest for environmental or
482 food samples to detect what is considered as a needle in a haystack (Rosani et al., 2019).

483 In this study, two full HuNoV genome sequences were identified in one of the oyster sample
484 contaminated with a thousand copies of HuNoV GI or GII per g of DT. These promising results
485 demonstrate the possibilities to improve the method sensitivity and its applicability to even less
486 contaminated oyster samples. Considering the variability of HuNoV and their high frequency of
487 genome recombination, it is important to achieve the recovery of long contigs that allows strain
488 identification both in the ORF1 and ORF2 typing regions (Chhabra et al., 2019). HuNoV from
489 different genotypes differ in their ability to cause disease thus a proper identification will help risk

490 manager to take adequate decisions (Parra et al., 2017, Atmar et al., 2018). For food outbreak analysis,
491 the strain identification in the two ORFs will be the only suitable way to identify transmission
492 pathways (Nieuwenhuijse et al., 2017, Nooij et al., 2018). Over the three oyster batches, 28 contigs
493 longer than 1000 nt were obtained, half of them providing the ORF1-ORF2 junction region, showing
494 that the method proposed here may help to investigate shellfish-borne outbreaks.
495 In human or in environmental matrices, the microbiome may play a role for HuNoV selection or
496 persistence (Monedero et al., 2018, Amarasiri et al., 2019, Walker et al., 2019). In this study we
497 compared two batches of *Crassostrea gigas* oysters: one that has never been exposed to human fecal
498 contamination and one bought from a local producer. This producer was selected in a clean area to
499 avoid pre-existing HuNoV or Human enteric virus contamination, but still some differences were
500 observed in bacteria concentrations. Nonetheless, no difference either in terms of HuNoV
501 bioaccumulation or in terms of strains identification were found between the two batches. No
502 difference either was observed with *Ostrea edulis*, except that this batch was the one that provided the
503 longest contigs for HuNoV identification.

504

505 **5. Conclusions**

506

507 Environmental samples such as oyster tissues are complex matrix generally characterized by
508 low virus concentration. The scientific interest on developing an approach able to reduce the host
509 background is high and need to be further investigated. An optimization of the hexamer set we
510 designed can be considered extending the number and the diversity of the reference sequences used for
511 hexamer selection. Furthermore, the use of triplicate libraries is recommended to increase the chances
512 to detect viruses present in low concentrations. Reliable data are also ensured by the use of negative
513 controls which should always be included in a metagenomic run.

514

515 **Data availability**

516 The short reads data for this study has been deposited in the European Nucleotide Archive (ENA)
517 <http://www.ebi.ac.uk/ena/dta/view/PRJEB34625> (ERS3786121-3786154, ERX3562554-3562586 and
518 ERR3564155-3564187). The full genome or longer than half genome sequences for HuNoV were
519 deposited in GenBank and are available under the accession numbers MN416944, MN416945,
520 MN416946, MN416764 and MN416765.

521

522 **Acknowledgments**

523 We are grateful to the sewage treatment plant staff for the sewage sample, and to our colleagues from
524 the Marine Mollusks Experimental platform (PMM Bouin, Ifremer) for providing the two batches of
525 oysters. This study was supported by the European Community's Horizon 2020 research and

526 Innovation Program under the COMPARE project, grant agreement N°643476. SS was supported by a
527 grant from the Ifremer Direction Scientifique.

528

529

530 **References**

531

- 532 AFNOR, 2010, NF V08-106. Dénombrement des *Escherichia coli* dans les coquillages vivants -
533 Technique indirecte par impédancemétrie directe
- 534 Amarasiri, M., Sano, D., 2019. Specific Interactions between Human norovirus and environmental
535 matrices: effects on the virus ecology. *Viruses* 11, 224.
- 536 Atmar, R. L., Neill, F. H., Romalde, J. L., Le Guyader, F., Woodley, C.M., Metcalf, T.G., Estes, M.
537 K., 1995. Detection of Norwalk virus and Hepatitis A virus in shellfish tissues with the PCR. *Appl.*
538 *Environ. Microbiol.* 61, 3014-3018.
- 539 Atmar, R. L., Ramani, S., Estes, M.K., 2018. Human noroviruses: recent advances in a 50-year
540 history. *Curr. Opin. Infect. Dis.* 31, 422-432.
- 541 Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikiv, A.S., Lesin, V.M.,
542 Nikolenko, S.I., Pham, S., Pribelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., tesler, G.,
543 Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: A New Genome Assembly Algorithm and Its
544 Applications to Single-Cell Sequencing. *J. Comp. Biol.* 19, 455-477.
- 545 Bartsch, C., Hiper, D., Made, D., Johne, R., 2018. Analysis of frozen strawberries involved in a large
546 norovirus gastroenteritis outbreak using next generation sequencing and digital PCR. *Food*
547 *Microbiol.* 76, 390-395.
- 548 Bellou, M., Kokkinos, P., Vantarakis, A. 2013. Shellfish-borne viral outbreaks: a systematic review.
549 *Food Environ. Virol.* 5, 13-23.
- 550 Bisseux, M., Colombet, J., Mirand, A., Roque-Alfonso, A-M., Abranavel, F., Izopet, J., Archimbaud,
551 C., Peigue-Lafeuille, H., Debroas, D., Bailly, J-L., Henquell, C., 2018. Monitoring human enteric
552 viruses in wastewater and relevance to infections encountered in the clinical setting: a one-year
553 experiment in central France, 2014 to 2015. *Euro Surveill.* 23, 7.
- 554 Bosch, A., Gkogka, E., Le Guyader, F.S., Loisy-Hamon, F., Lee, A., van Lieshout, L., Marthi, B.,
555 Myrmel, M., Sansom, A., Schultz, A-C., Winkler, A., Zuber, S., Phister, T., 2018. Foodborne
556 viruses : detection, risk assessment, and control options in food processing. *Int. J. Food Microbiol.*
557 285, 110-128.
- 558 Boxman, I., Verhoef, L., Vennema, H., Ngui, S., Friesema, I.H., Whiteside, C., Lees, D., Koopmans,
559 M., 2016. International linkage of two food-borne hepatitis A clusters through traceback of
560 mussels, the Netherlands, 2012. *Euro Surveill.* 21, 3.
- 561 Brown, J.R., Roy, S., Ruis, C., Romero, E.Y., Shah, D., Willimas, R., Breuer, J., 2016. Norovirus
562 whole-genome sequencing by sure select target enrichment: a robust and sensitive method. *J. Virol.*
563 54, 2530-2537.
- 564 Brown, M.R., Camezuli, S., Davenport, R.J., Petelenz-Kurdziel E., Ovreas, L., Curtis, T.P., 2015.
565 Flow cytometric quantification of viruses in activated sludge. *Wat. Res.* 68, 414-422.
- 566 Bushnell, B., Rood, J., Singer, E., 2017. BBMerge-Accurate paired shotgun read merging via overlap.
567 *PLoS ONE.* 12, e0185056.
- 568 Chhabra, P., de Graaf, M., Parra, G.I., Chan, M.C-W., Green, K., Martella, V., Wang, Q., White P.A.,
569 Katayama, K., Vennema, H., Koopmans, M.P.G., Vinje, J., 2019. Updated classification of
570 norovirus genogroups and genotypes. *J. Gen. Virol.* online first, doi, org/10.10999/jgv.0.001318.
- 571 Conceicao-Neto, N., Zeller, M., Lefrere, H., de Bruyn, P., Beller, L., Deboutte, W., Yinda, C.K.,
572 Lavigne, R., Maes, P., Van Ranst, M., Heylen, E., Matthijssens, J., 2015. Molecular approach to
573 customise sample preparation procedures for viral metagenomics: a reproducible protocol for
574 virome analysis. *Scientific Reports* 5,16532.
- 575 Cotten, M., Munnink, B.O., Canuti, M., Deijis, M., Watson, S.J., Kellam, P., van der Hoek, L., 2014.
576 Full genome virus detection in fecal samples using sensitive nucleic acid preparation, deep
577 sequencing, and a novel iterative sequence classification algorithm. *Plos ONE* 9, e93269.

578 Cotten, M., Petrova, V., Phan, M.V.T., Rabaa, M., Watson, S.J., Ong, S.H., Kellam, P., 2014. Deep
579 sequencing of norovirus genomes defines evolutionary patterns in an urban tropical setting. *J.*
580 *Virol.* 88, 11056-11069.

581 Di Girolamo, R., Liston, J., Matches, J., 1977. Ionic binding, the mechanism of viral uptake by
582 shellfish mucus. *Appl. Environ. Microbiol.* 33, 19-25.

583 EFSA, 2019. Analysis of the European baseline survey of norovirus in oysters. *EFSA J.* 17, 5762.

584 Endoh, D., Mizutani, T., Kirisawa, R., Maki, Y., Saito, H., Kon, Y., Morikawa, S., Hayashi, M., 2005.
585 Species-independent detection of RNA virus by representational difference analysis using non-
586 ribosomal hexanucleotides for reverse transcription. *Nucl. Ac. Res.* 33, 1-11.

587 Ettayebi, K., Crawford, S.E., Murakami, K., Broughman, J.R., Karandikar, U., Tenge, V.R., Neill,
588 F.H., Blutt, S.E., Zeng, X-L., Qu, L., Kou, B., Opekum, A.R., Burrin, D., Graham, D.Y., Ramani,
589 S., Atmar, R.L., Estes, M.K., 2016. Replication of human noroviruses in stem cell-derived human
590 enteroids. *Science.* 353, 1387-1393.

591 Giu, W., Miller, S., Chiu, C.Y., 2019. Clinical metagenomic next-generation sequencing for pathogen
592 detection. *Annu. Rev. Pathol. Mech. Dis.* 14, 317-336.

593 Glockner, F.O., Yilmaz, P., Quast, C., Gerken, J., Beccati, A., Ciuprina, A., Bruns, G., Yarza, P.,
594 Peplies, J., Westram, R., Ludwig, W., 2017. 25 years of serving the community with ribosomal
595 RNA gene reference databases and tools. *J. Biotechnol.* 261, 169-176.

596 ISO 15216-1, 2017. Microbiology of food and animal feed -horizontal method for determinaion of
597 hepatitis A virus and norovirus in food using real-time RT-PCR, Part 1: method for quantification.

598 Kallies, R., Holzer, M., Toscan, R.B., da Rocha, U.N., Anders, J., Marz, M., Chatzinotas, A., 2019.
599 Evaluation of sequencing library preparation protocols for viral metagenomic analysis from pristine
600 aquifer groundwaters. *Viruses.* 11, 484.

601 Kim, Y., van Bonn, W., Aw, T.G., Rose, J.B., 2017. Aquarium viromes: viromes of human-managed
602 aquatic systems. *Front. Microbiol.* 8, 1231.

603 Kittigul, L., Panjangampathana, A., Rupprom, K., Pombubpa, K., 2014. Genetic diversity of rotavirus
604 strains circulating in environmental water and bivalve shellfish in Thailand. *Int. J. Environ. Res.*
605 *Public Health.* 11, 1299-1311.

606 Kroneman, A., Vennema, H., Deforche, K., Avoort, H.V.D., Penaranda, S., 2011. An automated
607 genotyping tool for enteroviruses and noroviruses. *J. Clin. Virol.* 51, 121-125.

608 Langmead, B., Salzberg, S., 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods.* 9, 357-
609 359.

610 Le Guyader, F., Parnaudeau S, Schaeffer J, Bosch A, Loisy F, Pommepuy M, Atmar R.L., 2009.
611 Detection and quantification of noroviruses in shellfish. *Appl. Environ. Microbiol.* 74, 618-624.

612 Le Guyader, F.S., Le Saux, J-C., Ambert-Balay, K., Krol, J., Serais, O., Parnaudeau, S., Giraudon, H.,
613 Delmas, G., Pommepuy, M., Pothier, P., Atmar, R.L., 2008. Aichi virus, norovirus, astrovirus,
614 enterovirus and rotavirus involved in clinical cases from a French oyster-related gastroenteritis
615 outbreak. *J. Clin. Microbiol.* 46, 4011-4017.

616 Le Guyader, F. S., Atmar, R.L., Le Pendu, J., 2012. Transmission of viruses through shellfish: when
617 specific ligands come into play. *Cur. Op. Virol.*, 2, 103-110.

618 Li, M., Baker, C.A., Danyluk, M.D. Belanger, Ph., Boelaert, F., Cressey, P., Gheorghe, M.,
619 Polkinghorne, B., Toyofuku, H., Havelaar, A.H., 2018. Identification of biological hazards in
620 produce consumed in industrialized countries: A Review. *J. Food Prot.* 81, 1171-1186.

621 McLeod, C., Hay, B., Grant, C., Greening, G., Day, D., 2009. Localization of norovirus and poliovirus
622 in Pacific oysters. *J. Appl. Microbiol.* 106, 1220-1230.

623 McLeod, C., Polo, D., Le Saux, J-C., Le Guyader, F.S., 2017. Depuration and relaying: a review on
624 potential removal of norovirus from oysters. *Comp. Rev. Food Sci. Food Saf.* 16, 692-706.

625 Metcalf, T., Melnick, J.L., Estes, M.K., 1995. Environmental Microbiology : from detection of virus in
626 sewage and water by isolation to identification by molecular biology- a trip of over 50 years. *Ann.*
627 *Rev. Microbiol.* 49, 461-487.

628 Miura, T., Sano, D., Suenaga, A., Yoshimura, T., Fusawa, M., Nakagomi, T., Nakagomi, O., Okabe,
629 S., 2013. Histo-blood group antigen-like substances of human enteric bacteria as specific
630 adsorbents for human noroviruses. *J. Virol.* 87, 9441-9451.

631 Monedero, V., Buesa, J., Rodriguez-Diaz, J., 2018. The interactions between host glycobiology,
632 bacterial microbiota, and viruses in the gut. *Viruses.* 10, 96.

633 Nieuwenhuijse, D., Koopmans, M.P.G., 2017. Metagenomic sequencing for surveillance of food- and
634 waterborne viral diseases. *Front. Microbiol.* 8, 230.

635 Nooij, S., Schmitz, D., Vennema, H., Kroneman, A., Koopmans, M.P.G., 2018. Overview of virus
636 metagenomic classification methods and their biological applications. *Front. Microbiol.* 9, 749.

637 Oechslin, C.P., Lenz, N., Liechti, N., Ryter, S., Agyeman, P., Bruggmann, R., Leib, S.L., Beuret, C.
638 M., 2018. Limited correlation of shotgun metagenomics following host depletion and routine
639 diagnostics for viruses and bacteria in low concentrated surrogate and clinical samples. *Front. Cel.
640 Inf. Microbiol.* 8: 375.

641 Oude Munnink, B. B., van der Hoek, L., 2016. Viruses causing gastroenteritis : the known, the new
642 and those beyond. *Viruses.* 8, 42.

643 Parra, G. I., Squires, R.B., Karangwa, C.K., Johnson, J.A., Lepore, C.J., Sosnovtsev, S.V., Green, K.Y.,
644 2017. Static and evolving norovirus genotypes: implications for epidemiology and immunity. *PLoS
645 Pathogens.* 13, e1006136.

646 Polo, D., Schaeffer J., Teunis P., Buchet V., Le Guyader F.S. 2018. Infectivity and RNA persistence of
647 a norovirus surrogate, the Tulane virus, in oysters. *Front. Microbiol.* 9, 716.

648 Rosani, U., Shapiro, M., Venier, P., Allam, B., 2019. A needle in a haystack: tracing bivalve-
649 associated viruses in high-throughput transcriptomic data. *Viruses,* 11, 205.

650 Sano, D., Amarasiri, M., Hata, A., Watanabe, T., Katayama, H. 2016. Risk management of viral
651 infectious diseases in wastewater reclamation. *Environ. Int.* 91, 220-229.

652 Santiana, M., Ghosh, S., Ho, B.A., Rajasekaran, V., Du, W-L., Mutsafi, Y., De Jesus-Diaz, D.A.,
653 Sosnovtsev, S.V., Levenson, E.A., Parra, G.I., Takvorian, P.M., Cali, A., Bleck, C., Vlasova, A.N.,
654 saif, L.J., Patton, J.T., Lopalco, P., Corcelli, A., green, K.Y., Altan-Bonnet, N., 2018. Vesicle-
655 cloaked virus clusters are optimal units for inter-organismal viral transmission. *Cell Host Microbe.*
656 24, 208-220.

657 Schnell, I. B., Bohmann, K., Gilbert, T.P., 2015. Tag jumps illuminaad-reducing sequence-to-sample
658 misidentifications in metabarcoding studies. *Mol. Ecology Res.* 15, 1289-1303.

659 Steinegger, M., Mirdita, M., Soding, J., 2019. Protein-level assembly increases protein sequence
660 recovery from metagenomic samples manifold. *Nat. Methods.* 16, 603-606.

661 Strubbia, S., Schaeffer, J., Oude Munnink, B.B., Besnard, A., Phan, M.V.T., Nieuwenhuijse, D.F., de
662 Graaf, M., Schapendonk, C.M.E., Wacrenier, C., Cotton, M., Koopmans, M., Le Guyader, F.S.
663 2019. Metavirome sequencing to evaluate norovirus diversity in sewage and realated
664 bioaccumulated oysters. *Front. Microbiol.* 10, 2394.

665 Venugopal, V., Gopakumar, K., 2018. Shellfish: nutritive value, health benefits, and consumer safety.
666 *Comp. Rev. Food Sci. Food Saf.* 16, 1219-1242.

667 Walker, F. C., Baldrige, M.T., 2019. Interaction between noroviruses, the host, and the microbiota.
668 *Cur. Op. Virol.* 37, 1-9.

669 Wylezich, C., Papa, A., Beer, M., Hoper, D. 2018. A versatile sample processing workflow for
670 metagenomic pathogen detection. *Nature Scien. Rep.* 8, 13108.

671 Wylie, K. M., Wylie T., Buller R., Herter B., Cannella M.T., Storch G.A. (2018). Detection of viruses
672 in clinical samples by use of metagenomic sequencing and targeted sequence capture. *J. Clin.
673 Microbiol.* 56: e01123-01118.

674 Wylie, T. N., Wylie, K.M., Herter, B.N., Storch, G.A., 2015. Enhanced virome sequencing using
675 targeted sequence capture. *Genome Res.* 25, 1910-1920.

676 Yu, Y., Cai, H., Hu, L., Lei, R., Pan, Y., Yan, S., Wang, Y., 2015. Molecular epidemiology of oyster-
677 related human noroviruses and their global genetic diversity and temporal-geographical distribution
678 from 1983 to 2014. *Appl. Environ. Microbiol.* 81, 7615-7624.

679

680

Table

Table 1 : Read counts and HuNoV C_T values obtained for the different samples using the three hexamer sets

Sample	Hexamer set	Pool	Millions of reads				NoV Reads	NoV GH C_T values*
			Trimmed	Eukaryote	Mollusk	Virus		
C. gigas-1	I-HD	1	4	3	3	0.05	1,650	32.2 ± 0.2
		2	27	23	23	0.5	26,084	32.0 ± 0.3
		3	25	22	22	0.3	3,020	32.2 ± 0.3
	Random	1	6	5	5	0.04	1,854	30.6 ± 0.3
		2	30	27	26	0.2	4,743	30.6 ± 0.2
		3	25	23	23	0.2	7,299	30.7 ± 0.1
	Endoh	1	3	3	3	0.03	86	32.4 ± 0.6
		2	191	17	16	0.2	8,887	32.2 ± 0.2
		3	30	28	27	0.2	5,402	32.0 ± 0.5
C. gigas-2	I-HD	1	4	3	3	0.1	1,601	32.7 ± 0.2
		2	17	15	14	0.4	833	32.5 ± 0.3
		3	25	22	21	0.4	4,350	32.7 ± 0.3
	Random	1	35	3	3	0.07	184	31.7 ± 0.3
		2	27	23	22	0.6	17,264	30.8 ± 0.2
		3	28	25	24	0.7	41,583	31.2 ± 0.2
O. edulis	I-HD	1	3	2	2	0.1	3,955	34.8 ± 3.7
		2	22	17	17	0.7	14,454	34.8 ± 0.6
		3	22	18	18	0.5	3,696	32.6 ± 0.3
	Random	1	4	3	3	0.3	17,842	29.9 ± 0.2
		2	26	29	19	0.5	32,774	29.9 ± 0.1
		3	27	21	21	0.7	8,011	30.3 ± 0.1
C. gigas-1 nc ^o	I-HD	1	0.1	0.06	0.065	0.009	44	no C_T
	Random	1	4	2	2	0.3	179	no C_T
C. gigas-2 nc ^o	I-HD	2	20	12	12	2	2	no C_T
	Random	2	18	11	11	1	0	no C_T
O. edulis nc ^o	I-HD	3	23	6	6	6	0	no C_T
	Random	3	26	5	5	7	0	no C_T
Sewage	I-HD	1	6♦	0.1	0.003	1	159,838	32.2 ± 0.4
	Random	1	6♦	0.1	0.002	1	162,168	28.7 ± 0.1
UltraPure	I-HD	3	0.7	0.03	0	0.008	0	no C_T
	Random	3	33	5	0.009	7	0	no C_T
Seawater	I-HD	2	6	0.09	0.01	1	2	no C_T
	Random	2	0.4	0.03	0.007	0.0002	0	no C_T

* NoV C_T values expressed as the mean value of three PCR replicates and the standard deviation, ^onc: negative control, I-HD: In House designed hexamers; ♦Number of reads after normalization with bbnorm

Table 2 : HuNoV strains identified in the samples using random hexamers.

Sample	Node	Reads	Length	ORF 1	ORF 2	Ref. sequence	Identity
Sewage	189	2,993	2872	-	GI.1	MH638229.1	97.08
	237	4,435	2564	GI.P1	-	MH638228.1	98.28
	294	5,557	2299	-	GI.1	MH638229.1	99.76
	601	3,380	1677	GII.P12	-	MG601447.1	98.20
	727	2,939	1528	GII	-	MH218642.1	97.86
	832	1,124	1435	GII	-	MF421538.1	99.03
	879	3,050	1401	-	GII.3	MG601451.1	98.36
	971	5,316	1346	GI.P4	GI.4	LN854563.1	97.70
	1128	2,987	1244	GII	-	MH218642.1	98.12
	1147	1,314	1233	GI	-	LN854563.1	97.49
	1236	425	1191	GII	-	KM198498.1	94.43
	1455	578	1097	GI	-	LN854563.1	96.89
	1656	10,035	1029	GII.P7	GII.6	MH218642.1	98.74
	C.gigas-1	34	560	2243	GII.P7	GII.6	MH218642.1
38		241	2192	GI.P1	GI.4	LN854563.1	87.79
156		1,082	1629	GI.P1	GI.1	MH638229.1	97.85
192		52	1055	GII	-	MK073886.1	98.62
209		94	1016	-	GI.1	MH638229.1	98.13
C.gigas-2	13	1951	4205	GII.P7	GII.6	MH218642.1	98.05
	16	2,341	4023	GI.P4	GI.4	LN854563.1	96.69
	17	1,385	3960	GI.P1	GI.1	MH638228.1	98.84
	20	2,296	3677	GI.P2	GI.2	KF306212.1	99.12
	26	619	3303	-	GII.6	MH218642.1	98.00
	216	176	1371	-	GII.3	MG601451.1	94.39
	77	464	1985	GII	-	MG601447.1	98.39
	346	174	1101	GII.P12	GII.3	MH260511.1	98.37
	85	338	1893	GI	-	MH218649.1	97.68
	86	445	1890	-	GI.2	KF306212.1	99.15
	98	261	1817	GII	-	MG601447.1	98.24
	326	167	1038	-	GI.4	GQ413970.1	97.15
	377	111	1072	GII	-	KU870455.1	94.96
	434	138	1019	GII	-	LN854563.1	95.49
	O.edulis	2	2,858	7445	GII.P7	GII.6	MH218642.1
4		2,261	7288	GI.P2	GI.2	LN854563.1	99.69
5		1,722	6422	GII.P12	GII.3	MG601447.1	97.88
25		1,594	3852	GI.P4	GI.4	LN854563.1	96.72
30		387	3257	GII.P7	-	KU870455.1	95.42
51		641	2479	-	GII.6	KU870455.1	95.60
59		1,487	2274	GI.P1	GI.1	MH638228.1	98.55
95		729	1741	-	GI.1	MH638229.1	97.97
105		120	1618	GI	-	LN854563.1	96.89

Contigs longer than 1000 nt are shown with the node number assigned during assembly, the number of reads that mapped to the assembled contig, the contig length, the ORFs identification according to the NoV typing tool v2.0 (Kroneman et al, 2011), and the reference sequence and percentage of identity obtained using BLAST (NCBI). In bold the sequences longer than half of the complete HuNoV genome.

Figure

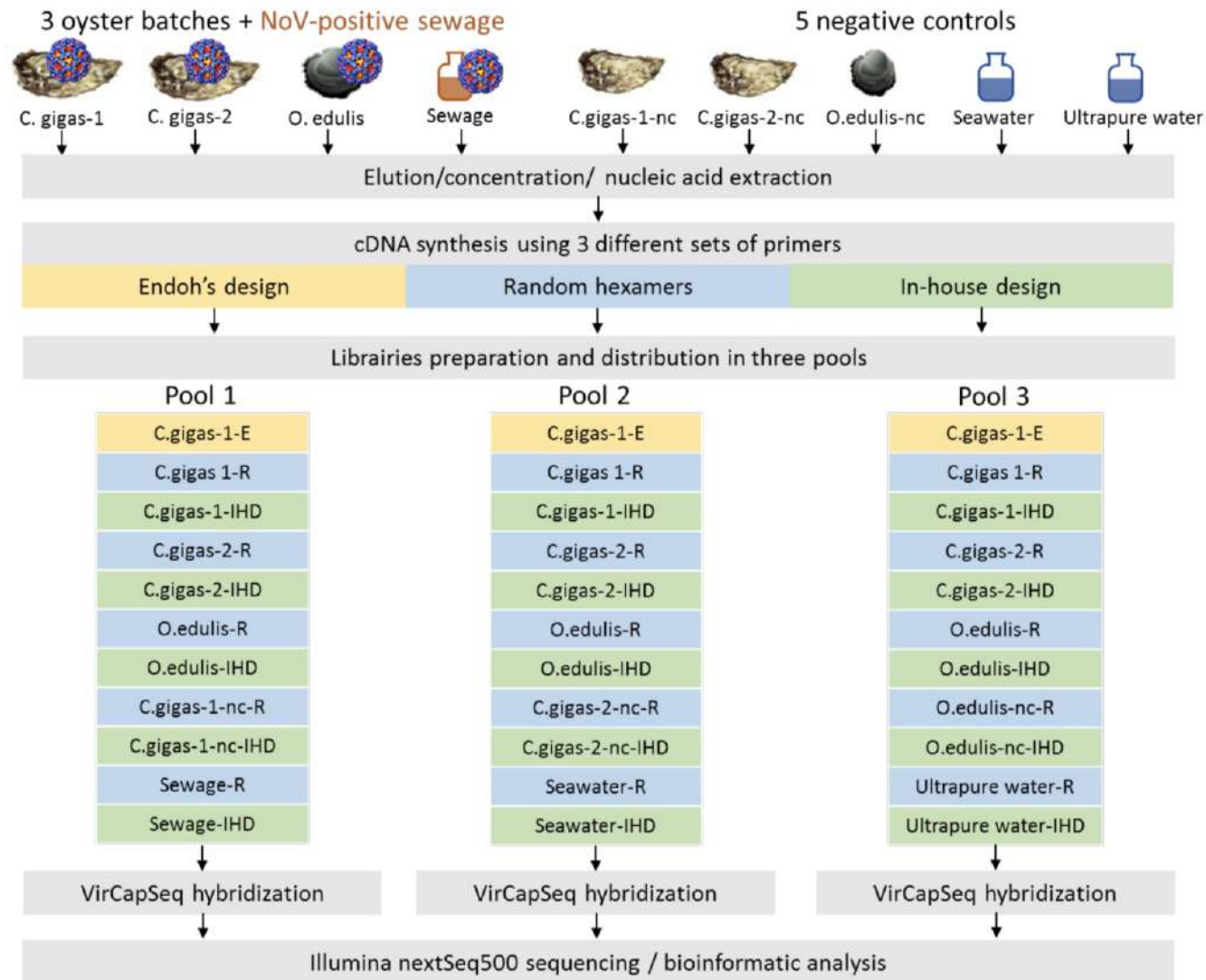


Fig. 1. Workflow of samples and libraries preparation and pool organization prior to hybridization for capture-based deep sequencing.

cDNA were prepared using Endoh (E, yellow), Random (R, blue) or *In-house* designed (I-HD, green) hexamer sets. Negative control (nc) were included in the different pool.

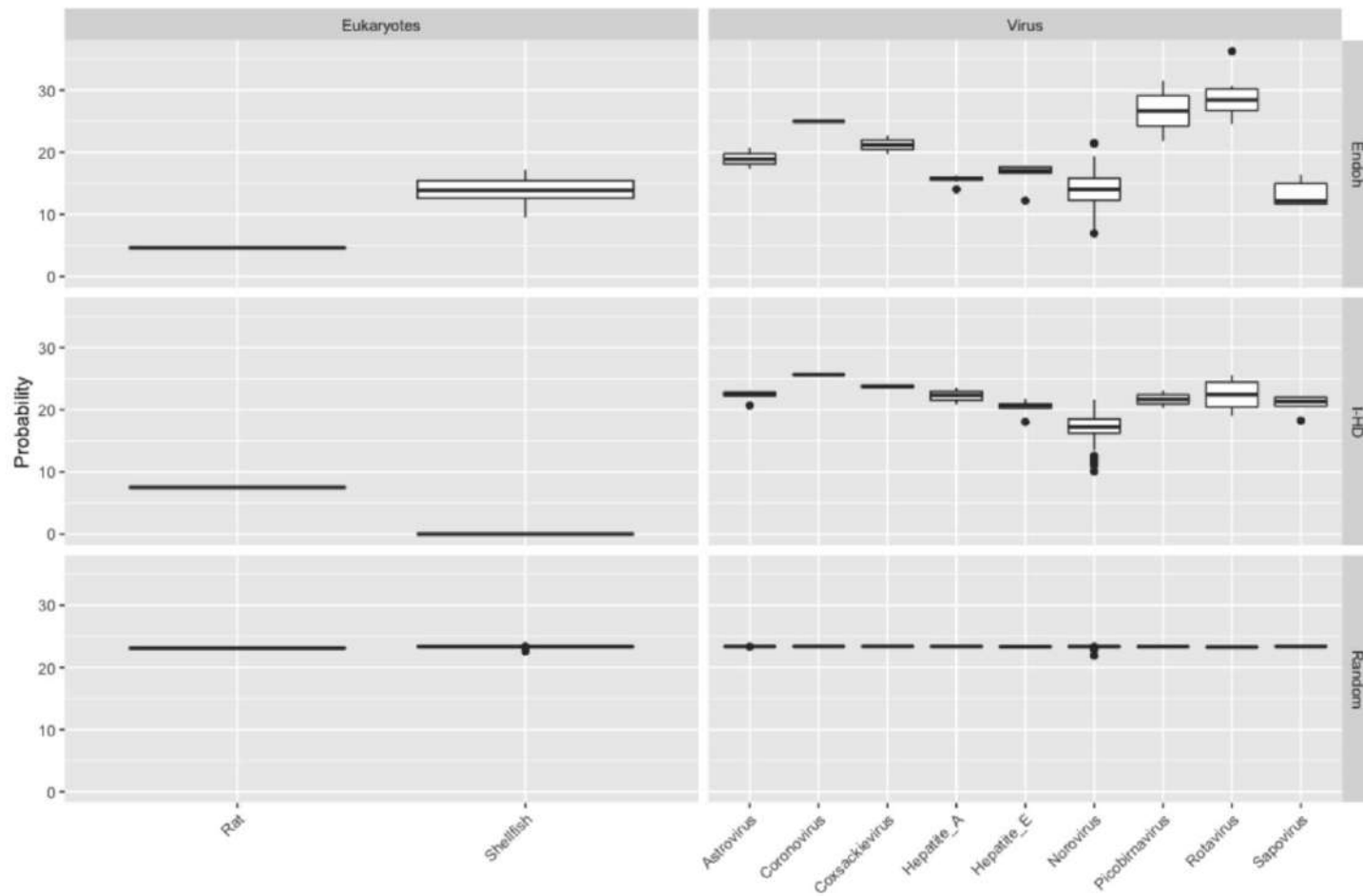


Fig. 2. Frequencies of priming sites in Eukaryote or virus genomes for three sets of hexamers.

The probability was calculated as the sum of probabilities of all non-ribosomal hexanucleotides in sense and complementary sequences to selected genomic sequences.

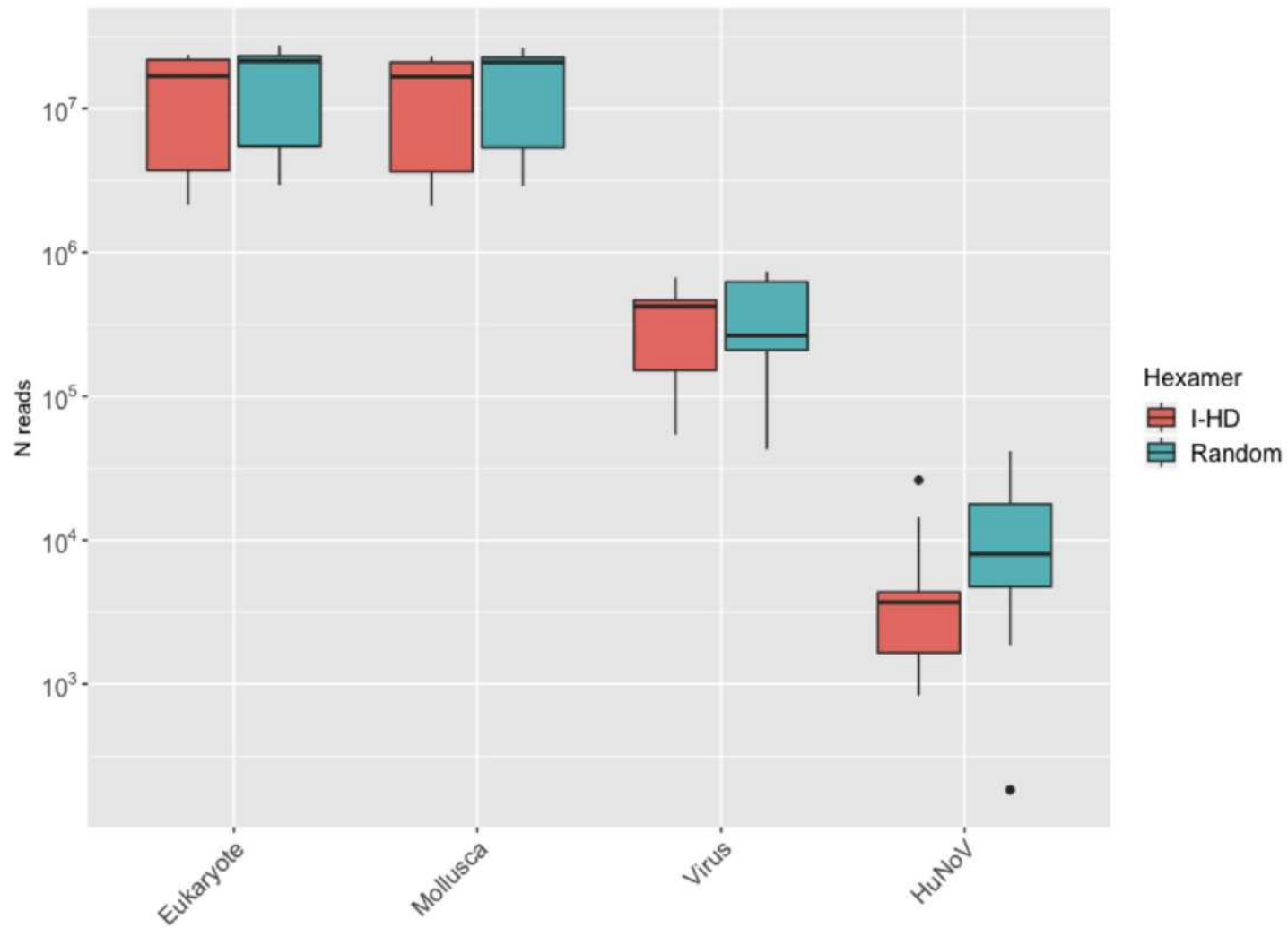


Fig. 3. Reads distribution obtained from bioaccumulated oyster libraries prepared using the I-HD (red) and random (green) sets.

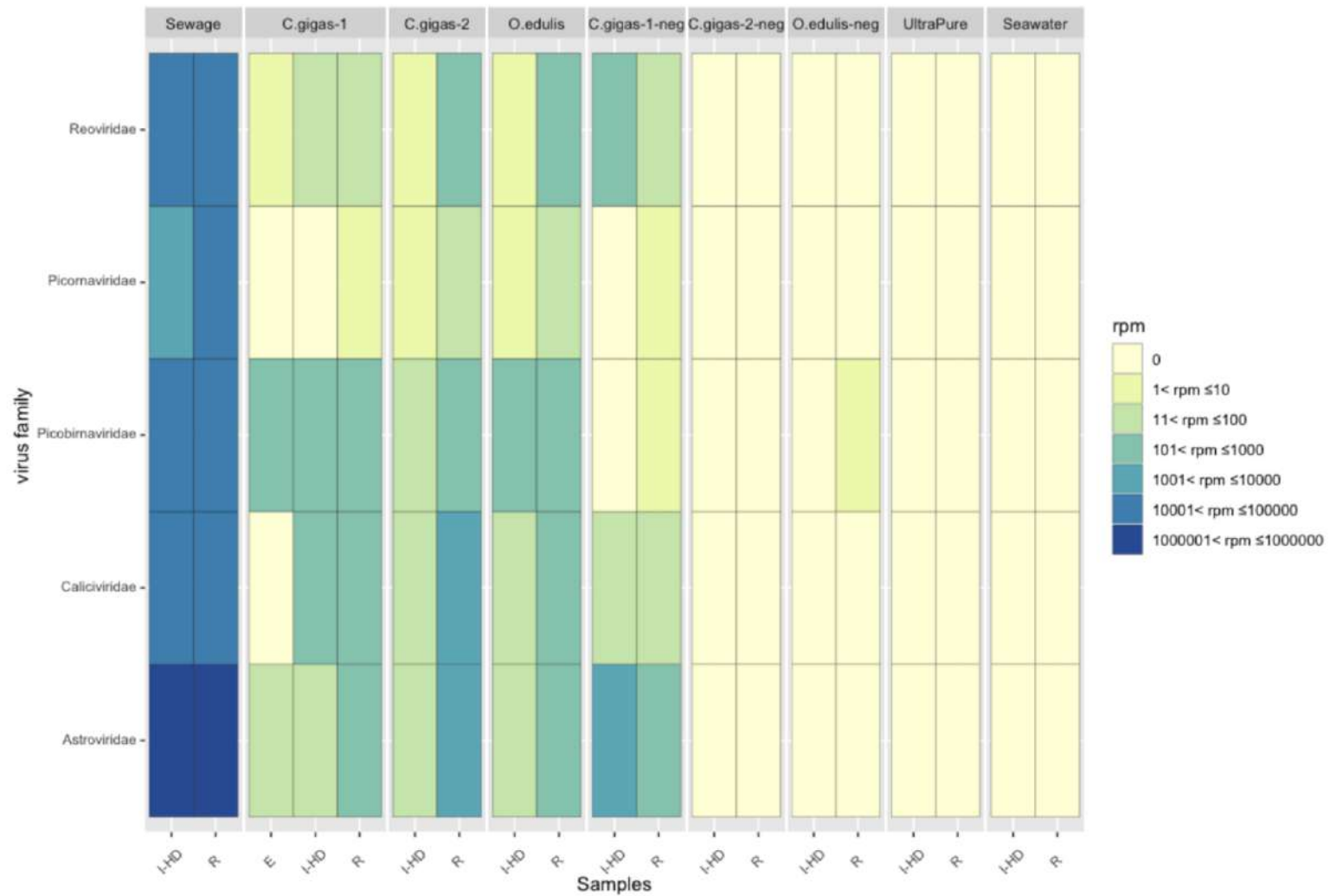


Fig. 4. Human enteric virus associated reads identified according to the different sets of hexamers used for the RT.

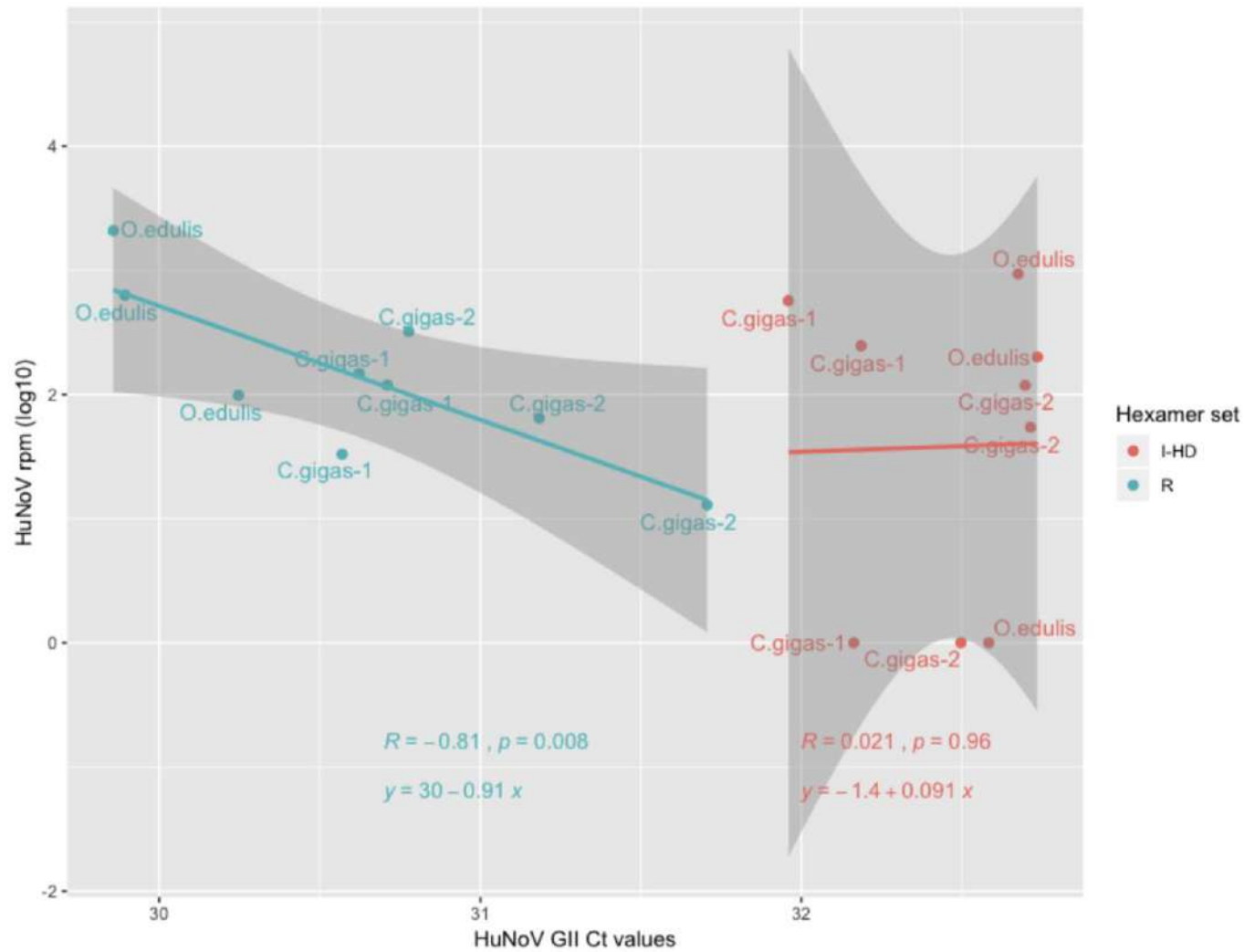


Fig. 5. Relationship between HuNoV concentrations and read expressed in reads per million (rpm), according to the hexamer sets used for cDNA.

Red dots : In-house designed (I-HD) set, blue dots: Random (R) hexamer set.

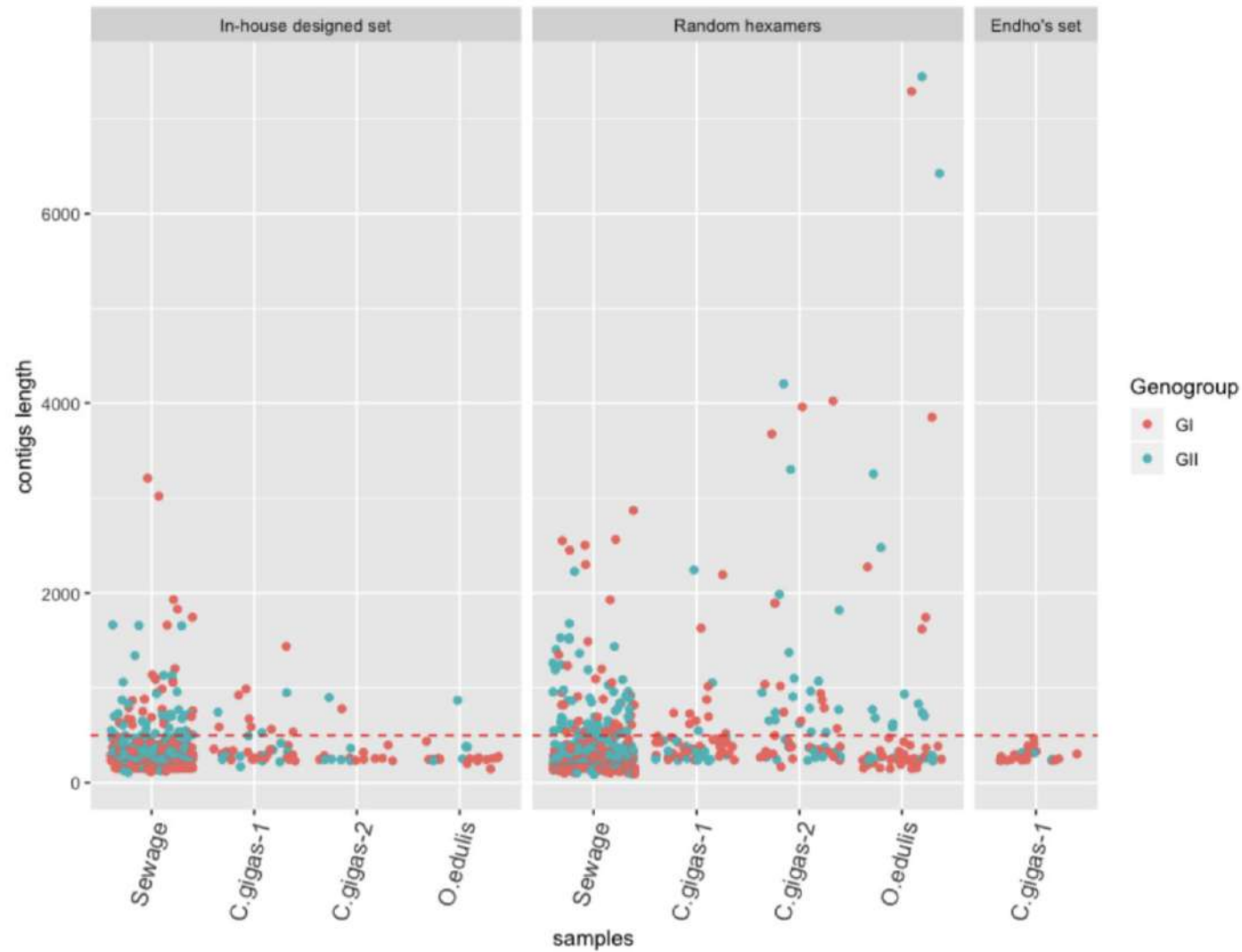


Fig. 6. Scatter plot of jittered data showing the length of contigs obtained using the three hexamer sets.

Genogroup I (GI) (red dots) and GII (blue dots) obtained for the sewage sample and bioaccumulated oyster batches are reported. The red dashed line represents the threshold contigs length fixed for data analysis (500nt)

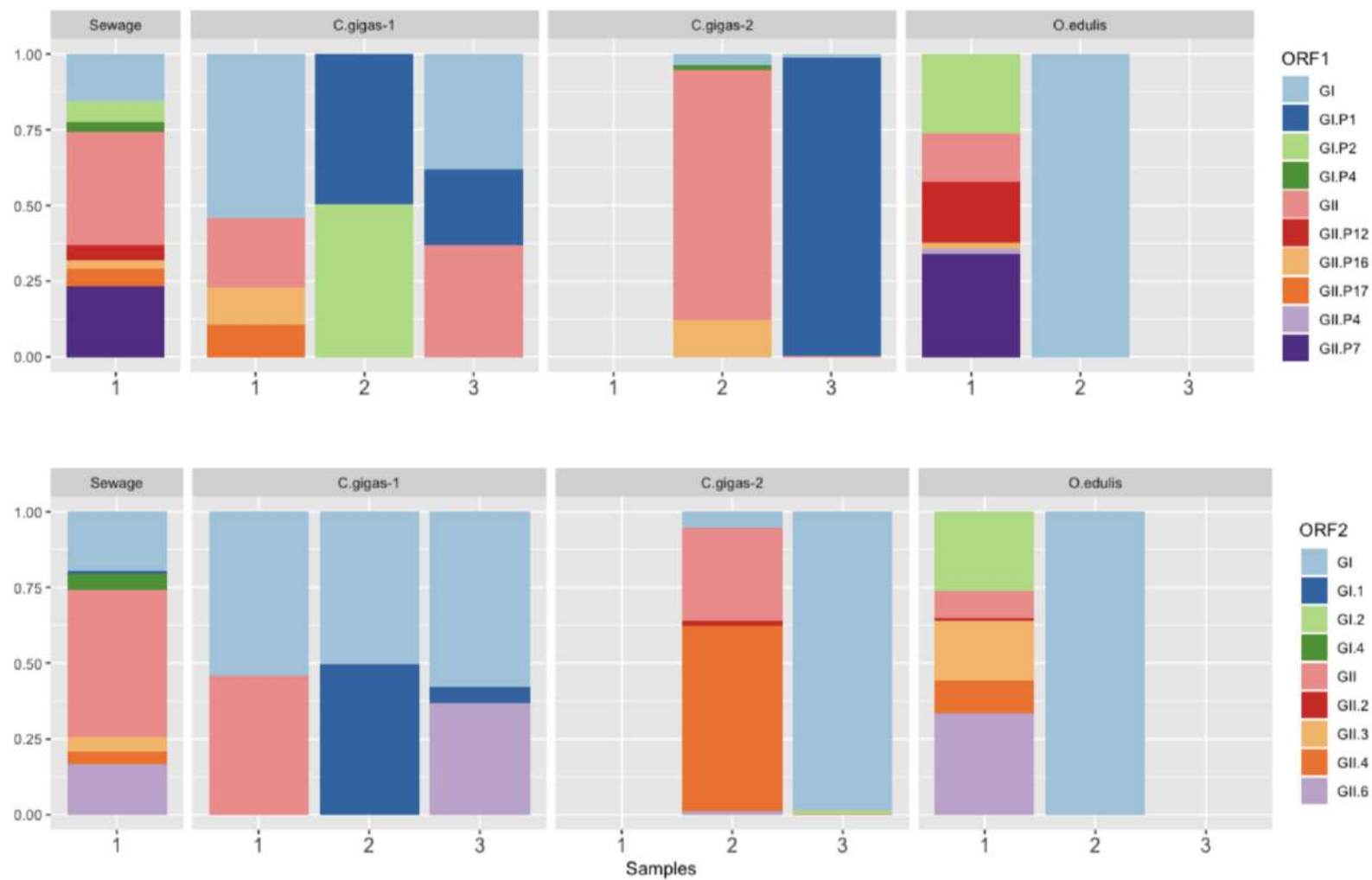


Fig. 7. HuNoV ORF1 and ORF2 identified in sewage and oyster samples using random hexamers during library preparation.

The three technical replicates are reported for each oyster batch.

3.4 DISCUSSION

Cette étude visait à fournir une méthode permettant d'identifier les NoVs et d'autres virus entériques humains dans le virome du tissu digestif d'huître, comparant trois lots d'huîtres. Une approche metagénomique virale inclue toutefois des défis qui peuvent représenter une limite à son applicabilité. L'élution et la récupération des particules de NoVs à partir des tissus d'huîtres représentaient le premier challenge. La présence de liens spécifiques entre les NoVs et les tissus d'huître complique cette étape (Di Girolamo *et al.*, 1977; McLeod *et al.*, 2009; Le Guyader *et al.*, 2012).

L'élution et la récupération des particules de NoVs a nécessité de plusieurs étapes. D'abord un traitement enzymatique pour digérer les tissus et libérer les particules attachées, ensuite une étape de sonication pour réduire la population bactérienne et pour favoriser la désagrégation d'agglomérats viraux. Après une élution chimique avec du pyrophosphate de sodium les particules virales ont été concentrées par le Polyéthylène glycol (PEG).

Compte tenu des connaissances limitées sur l'association de NoVs ou d'autres virus entériques humains aux bactéries, aucune étape de filtration n'a été appliquée après l'étape de concentration en PEG afin de prévenir toute perte de particules virales (Conceicao-Neto N. *et al.*, 2015; Amarasiri *et al.*, 2019). Toutefois une élimination par taille, ajoutant une étape de filtration jusqu'à 0.45 µm, pourrait être considérée.

Ensuite nous avons testé trois pools d'hexamères (Endoh, random et I-HD) pendant la phase de RT-PCR afin de limiter le nombre de séquences bactériennes et de l'hôte qui composent le background. Le pool d'hexamères que nous avons sélectionnés, I-HD, a réduit le nombre de séquences de l'hôte d'environ 10% par rapport aux hexamères random. Toutefois le background était encore extrêmement élevé représentant 87% des reads sur les échantillons traités avec le pool I-HD et 97% sur les échantillons traités avec les hexamères random. Ces résultats mettent en évidence la nécessité d'une ultérieure optimisation dans cette étape.

Par rapport à l'identification du virus cible, les échantillons traités avec les hexamères I-HD ont produit un nombre inférieur de reads de NoVs comparé aux échantillons traités avec les hexamères random, dans les trois lots d'huître.

L'obtention d'un plus grand nombre de reads de NoVs suite au traitement des échantillons avec les hexamères random a confirmé la présence d'une corrélation entre nombre de reads et concentration virale dans l'échantillon.

La taille des contigs est également un critère important pour choisir une méthode efficace, car des contigs plus longs permettent une meilleure caractérisation de la souche. Les reads obtenues en utilisant le pool sélectionné par Endoh et al. ont généré des contigs de NoVs de taille inférieure à 500 nt et n'ont donc pas été pris en compte pour l'analyse metagénomique.

L'utilisation des hexamères random a permis de récupérer les plus longs contigs pour tous les échantillons, avec deux génomes entiers de NoVs identifiés à partir des huitres *O. edulis*.

Pour le lot C. gigas-2, la plupart des contigs obtenus n'ont pas permis le typage des souches de NoVs identifiées. Cependant, dans le même lot d'huître, quatre séquences couvrant plus de la moitié du génome ont été détectées, trois souches NoVs GI et une GII. Cinq autres souches de plus de 1000nt ont permis la caractérisation de la région ORF1 ou ORF2.

Malgré l'abondance de reads obtenues à partir des échantillons d'eau usée (> 63% du totale de reads produites) la plus part des contigs obtenus étaient courts et n'ont pas permis le typage des souches, ce qui laisse penser à des problèmes pendant la phase d'assemblage, probablement à

cause d'une redondance d'information.

Dans l'ensemble, cinq familles virales ont été identifiées : *Reoviridae*, *Picornaviridae*, *Picobirnaviridae*, *Caliciviridae* et *Astroviridae*, dans l'échantillon d'eaux usées et dans les trois lots d'huîtres bioaccumulées.

L'utilisation des hexamères random a fourni les nombres de reads les plus élevés pour les cinq familles virales, avec *Astroviridae* et *Picobirnaviridae* comme les familles les plus représentées.

Concernant les contrôles négatifs utilisés pendant l'étude, seul le contrôle du lot C. gigas-1 a montré des positivités avec des reads identifiées comme NoVs. L'absence de détection de NoV par PCR quantitative lors des analyses préliminaires de ces échantillons, laisse penser à une contamination en laboratoire pendant le traitement de l'échantillon ou de la librairie, car ce contrôle négatif se trouvait dans le pool contenant l'échantillon le plus contaminé (eau usée).

Des reads de coquillages ont été détectés dans l'eau UltraPure, probablement dû à une amplification non spécifique qui peut s'être produite au cours des cycles d'amplification supplémentaires effectués pour atteindre une concentration suffisante en ADN lors de la préparation de la librairie ou en raison d'une contamination de laboratoire pendant l'étape de préparation.

3.5 CONCLUSIONS ET PERSPECTIVES

Les échantillons environnementaux tels que les tissus d'huîtres sont des matrices complexes généralement caractérisées par une faible concentration en virus. Toutefois la grande variabilité des NoVs et leur capacité à générer des recombinaisons, accroît la nécessité de développer une méthode de détection sensible et efficace. L'apport de la métagénomique peut à cet égard être extrêmement intéressant, bien que des améliorations méthodologiques soient encore nécessaires. L'identification des souches nécessite des longs contigs permettant l'identification des régions de type ORF1 et ORF2 (Kroneman *et al.*, 2013).

L'intérêt scientifique pour le développement d'une approche capable de réduire le background de l'hôte est élevé et doit être approfondi. Une optimisation du pool d'hexamères que nous avons conçu peut-être envisagée prenant en compte un nombre plus élevé de séquences de NoV de référence lors de leur sélection.

La variabilité des génomes de NoV, et le nombre élevé des reads produits lors du séquençage métagénomique, peuvent poser un problème pendant la phase d'assemblage des reads de NoV, comme reporté ici pour les échantillons d'eaux usées. Des nouveaux outils et d'autres stratégies pourront être évalués, dont l'assemblage à niveau des protéines plutôt que à niveau des nucléotides (Steinegger *et al.*, 2019).

Enfin, l'utilisation de triplicats est recommandée afin d'augmenter les chances de détecter les virus présents à faible concentration. Des données fiables sont également assurées par l'utilisation de contrôles négatifs qui doivent toujours être inclus dans une analyse métagénomique.

CONCLUSIONS GENERALES ET PERSPECTIVES

Cette thèse s'est développée autour de deux axes principaux: l'utilisation des technologies de séquençage haut débit pour étudier la diversité des souches de NoVs présents dans l'environnement et l'utilisation la culture d'enteroïdes pour étudier la durée de persistance des NoVs infectieux dans l'environnement.

Le premier objectif a été méthodologique. Il s'agissait d'intervenir sur les étapes de préparation des échantillons et des librairies avec une approche polyvalente de concentration des virus entériques, ciblant préférentiellement les NoVs et applicables à l'analyse d'eaux usées ou d'huîtres contaminées. Afin de tirer le meilleur bénéfice de l'approche metagénomique, une adaptation méthodologique spécifique à la matrice analysée et aux micro-organismes ciblés est importante. Sélectionner un volume d'eaux usées (ou de surface) garantissant une bonne représentativité est une question difficile. Le plus souvent l'échantillonnage de grands volumes, qui sont ensuite concentré avant la poursuite de l'analyse, est l'option sélectionnée. Par exemple la concentration de 100 L d'eau de surface en utilisant un système de filtration tangentielle a permis de caractériser la diversité de virus (Bisseux *et al.* 2018). Mais cette approche, applicable essentiellement sur les eaux faiblement chargées en particules, peut conduire à concentrer les inhibiteurs, ou induire une perte de certaines particules virales selon leur fixation à différents types de particules. Une option souvent utilisée au laboratoire consiste dans l'utilisation de préleveur automatique, mais cette technique n'est applicable qu'avec une source d'énergie. Les capteurs passifs, technique développée au laboratoire constitue une approche séduisante. En effet ces capteurs sont des membranes immergées sur lesquelles les micro-organismes vont se fixer. Le choix des membranes (nylon, PVDF..) peut éventuellement privilégier certains micro-organismes (Vincent-Hubert *et al.* 2017). Cette approche vient également d'être proposée avec une détection en utilisant un simple téléphone portable (Chung *et al.*, 2019).

Cette question de l'échantillonnage se pose également pour l'autre matrice étudiée dans ce travail, les huîtres. En effet il y a encore peu de données sur l'homogénéité de la contamination de lot d'huître en milieu naturel (Le Mennec *et al.*, 2017). Ce travail ayant pour objectifs de développer des méthodes nous avons opté pour des huîtres bio-accumulées en laboratoire, technique mise au point depuis de nombreuses années au laboratoire et qui nous permet de travailler avec des lots présentant une contamination homogène, tout en restant proche des concentrations observées en milieu naturel (Maalouf *et al.*, 2010). La prochaine étape de ce travail sera de vérifier sur des échantillons prélevés sur le littoral.

Les virus entériques sont le plus souvent rejetés sous formes d'agglomérats, ou fixés sur des particules ou encore des bactéries (Miura *et al.*, 2013). Une des premières étapes pour optimiser leur récupération est de favoriser leur élution en jouant sur le pH, la conductivité ou en favorisant leur adsorption sur des molécules connues telles que les mucines gastriques porcines (PGM) (Mullendore *et al.* 2001 ; da Silva *et al.*, 2011). Bien que plusieurs publications aient mentionné l'intérêt de ces mucines pour la sélection des NoVs nous avons été surpris du manque de reproductibilité de cette approche. En effet, si nos essais *in vitro* avec les VLPs de NoV GI.1 et GII.4 ont montré un effet dose-dépendant, l'application à des souches de NoVs purifiées de selles

humaines n'a pas donné de résultat satisfaisant. Un autre intérêt des PGM est que la fixation des NoVs se réalisant sur des sites spécifiques présents sur la capsid, des informations sur l'intégrité du virus pouvaient être obtenus (Afolayan *et al.*, 2016 ; Tian *et al.*, 2012). Mes travaux, bien que nécessitant des compléments, montrent de nombreuses limites à leur utilisation et à cette interprétation. Par exemple, selon la ou les souches de NoVs présents dans l'échantillon, une sélection préférentielle pourrait se faire et ainsi biaiser le résultat final. Les interactions aspécifiques et la présence de nombreux acides sialiques laissent également présager de nombreux résultats difficilement interprétables. Lorsque nous avons utilisés cette approche pour enrichir les NoVs dans des échantillons de selles ou d'eaux usées avant analyse métagénomique si le nombre de reads de NoVs a été augmenté l'effet a également été notable sur d'autres familles virales. En particulier les résultats observés sur les astrovirus nous ont surpris (Article N°1). Néanmoins l'approche de capture sur des billes permettant de concentrer les virus est séduisante et les développements futurs pourraient considérer l'utilisation de ces PGM en associant d'autres molécules telles que des antigènes synthétiques de groupes sanguins, ou des polysaccharides extraits de bactéries marines ou entériques portant des structures glycaniques proches reconnues par les NoVs. Cette approche sélective pourrait aider en effet à une détection plus spécifique des NoVs ou d'autres virus reconnaissant ces sites et donc présentant un éventuel risque en santé publique. Associé à des détections en temps réel (type smartphone) cette approche constituerait un outil d'alerte pour prévenir les phénomènes de contamination.

La concentration par le PEG est une technique ancienne utilisée pour la mise en évidence de virus entériques dans l'environnement. Son utilité s'est vérifiée dans mes travaux et cette technique simple nous a permis l'identification de longues séquences de NoVs ainsi que d'autres familles virales (Article N°1 et N°2). Cependant, le nombre de reads correspondant aux séquences virales représentait moins de 1% du nombre total de reads, montrant la nécessité d'optimiser la méthode afin de diminuer ce 'bruit de fond'. L'addition d'étape de traitement chimique ou physique a eu un effet bénéfique aussi bien en terme de récupération de quantité de NoVs qu'en terme de qualité des séquences obtenues après traitement bio-informatique (article N°2). Cependant chaque étape doit être considérée avec précaution car, par exemple, appliquer une filtration pour éliminer les bactéries peut conduire à une perte en NoVs si un traitement pour les éluer n'a pas été appliqué au préalable.

Nous avons suivi la même logique de mise au point pour optimiser la récupération des virus à partir des tissus d'huitres. Ainsi l'association de la digestion enzymatique par protéinase K, la concentration par le PEG et un traitement chimique a donné des résultats satisfaisant mais le % de reads correspondant à la composante bactérienne et au génome de la matrice était encore trop abondante. D'autres traitements tels que la sonication et/ou une filtration pourraient être envisagés.

Les étapes de préparation des échantillons sont importantes pour enrichir en particules virales et éliminer autant que possible la matrice relative à l'échantillon, mais d'autres approches d'enrichissement sont possibles après extraction et purification des acides nucléiques. Lorsque l'on cible les virus entériques humains, constitués essentiellement de virus à ARN, la première étape est la synthèse de l'ADNc avant la préparation des bibliothèques. Si cette étape de transcription inverse est réalisée le plus souvent en utilisant un mélange de courtes amorces aléatoires (hexamers random), l'utilisation d'hexamères ciblant les virus infectant les mammifères a été

publié (Endoh *et al.*, 2005). Basé sur ce principe nous avons sélectionné des hexamères reconnaissant des génomes viraux d'intérêts pour notre thématique tout en présentant une reconnaissance la plus faible possible des génomes de mollusques bivalves. Le pool que nous avons sélectionné (I-HD) a réduit le nombre de séquences d'huitre d'environ 10% par rapport aux hexamères random (Article N°3) nous confortant dans cette approche, malgré un bruit de fond encore trop élevé. Sélectionner un nombre plus important d'hexamères reconnaissant un panel de génomes viraux plus importants semble donc une piste prometteuse. Cependant un outil de ce type se limiterait à l'analyse des données aux mollusques marins et ne nous permettrait pas de comparer avec les eaux environnantes. En effet le biais de sélection pourrait être différents et il est important de considérer les facteurs pouvant influencer les résultats.

Une autre option, applicable à tout type d'échantillon, est d'enrichir en séquence virale au cours de la librairie comme par exemple la puce développée spécifiquement pour cibler les génomes de NoVs (SureSelect, Brown *et al.*, 2016) ou VirCap-SeqVERT, ciblant 207 taxons viraux infectant des vertébrés (Briese *et al.*, 2015). Si la puce spécifique des génomes de NoVs nous a donné des résultats intéressants (Article N°1), son application limitée aux séquences connues de NoVs et prévenant la détection des autres virus entériques semble inappropriée pour les échantillons de l'environnement. La seconde approche, ciblant un large panel de virus à ARN, semble plus en adéquation avec les potentialités du séquençage metagénomique, même si des génomes très divergents ou non identifiés dans les bases de données, pourraient échapper à la capture et donc ne pas être détectés. Cependant les résultats obtenus (Article N° 2 et N° 3) sont prometteurs tant par la qualité des séquences obtenues (longs fragments) que par la diversité des génomes identifiés. L'optimisation des sondes fixées ciblant davantage de virus entériques est un axe à développer.

La préparation de l'échantillon et les 'défis de la pailleasse' sont importants, mais le challenge de l'analyse des reads obtenus est loin d'être négligeable. Parmi les étapes qui constituent le pipeline d'analyse metagénomique, la phase d'assemblage des reads en contigs constitue une des difficultés majeures principalement à cause de l'abondance de données de départ et de la variabilité des séquences. Ainsi, l'abondance des reads peut parfois présenter un écueil comme observé lorsque nous avons réalisé le séquençage en utilisant la technologie Illumina NextSeq, qui a produit 120 Gb de données (au lieu de 15 Gb obtenus avec un MiSeq). Nous avons dû adopter une normalisation *in silico* (BBnorm) pour réduire cette complexité en filtrant la couverture en excès par élimination des séquences redondantes (Rose *et al.*, 2016) (Article N°3). Une autre stratégie possible aurait pu être le sous-échantillonnage aléatoire d'un nombre défini de reads sur la totalité des reads de l'échantillon, mais avec le risque de perdre des séquences uniques. Malgré l'étape de normalisation, aucun génome complet de NoVs n'a pas pu être assemblé, alors que lors de l'analyse précédente, utilisant le même échantillon, plusieurs génomes complets avaient été obtenus (Article N°2). D'autres stratégies comme la conversion des reads en protéines (algorithme Plass) afin de réduire la variabilité pourrait faciliter l'assemblage (Steinegger *et al.*, 2019) ou encore procéder à l'identification directe des courts reads (Kaiju), préalablement convertis en protéines, mais dépassant l'étape d'assemblage (Menzel *et al.*, 2016). Ces approches sont séduisantes pour l'analyse de génomes viraux à ARN tels que ceux des NoVs, présentant un taux de mutation élevé. Cependant, la simplification de l'information obtenue par ces deux approches peut également diminuer l'information sur la diversité observée. Cet inconvénient possible

pourrait être lié au fait que Kaiju se limite à la caractérisation de séquences très courtes (reads) et que Plass réduit la variabilité du génome lors du passage en séquence protéique. Cet aspect d'analyse des données est très importants et commence à être considéré par la comparaison des différentes pipelines publiés (Nooij *et al.*, 2018) et par des essais inter-laboratoires d'analyse de jeu de données (Sala *et al.*, submitted).

Détecter de faibles concentrations d'acides nucléiques par biologie moléculaire a toujours constitué un enjeu pour l'analyse d'échantillon de l'environnement. Au sein de notre laboratoire, l'utilisation de triplicat d'extraction a été appliqué lors de diverses études (Le Mennec *et al.*, 2017, Maalouf *et al.*, 2011, Grodzki *et al.*, 2014). Considérant la sensibilité importante du séquençage haut débit et supposant que des variations sont inévitables d'un run à l'autre ou d'une librairie à l'autre, nous avons dans un premier temps inséré un échantillon de référence lors de trois runs de séquençage (Article N°2). Comme attendu des différences ont été observées, liées à la complexité de la matrice, à l'opérateur et au caractère aléatoire lors du pipetage des petits volumes requis pour la préparation de la librairie, à l'amplification *random* et la pré-amplification mais également aux erreurs qui peuvent se produire pendant le séquençage. Cet échantillon contrôle nous a semblé important, et l'intérêt des répliquats biologiques a été démontré lors de ma dernière étude (Article N°3). Ces approches montrent l'importance de la conception du plan d'expérience selon la question posée et de l'échantillon à analyser. L'utilisation des répliquats soulève également une question analytique. En effet, si des différences peuvent se présenter lors de la préparation de l'échantillon et lors du séquençage, l'introduction d'une certaine variabilité est également possible lors du traitement bio-informatique des données. La question « *A quel moment est-il approprié d'unifier les informations provenant des différentes répliquats ?* » peut alors se poser. Effectivement les informations peuvent être fusionnées dès le début de l'analyse (traitement des reads des triplicats en même temps), ou analyser chaque librairie séparément et avoir une vue d'ensemble sur la totalité des séquences produites à la fin de l'assemblage. Nous avons sélectionné cette dernière option mais il pourrait être intéressant d'évaluer d'autres approches comme l'assemblage protéique ou l'identification des reads cumulés.

Au final mes travaux ont montré l'importance des étapes de préparation de l'échantillon pour enrichir en particules virales, et l'intérêt des contrôles et des répliquats biologiques et techniques. La métagénomique appliquée aux virus est encore peu développée et il est important de vérifier les points-clés de l'analyse qui permettront l'interprétation des résultats. Cette considération implique également l'analyse bio-informatique, domaine en pleine évolution, indispensable à une bonne interprétation des données. L'analyse d'échantillons de l'environnement ou d'aliments, sans contexte biologique et sans données sur les sources éventuelles de contamination constitue un challenge pour les années futures. Si la métagénomique, par sa sensibilité et sa capacité à permettre l'identification des agents pathogènes sans a priori, ouvre de nouvelles opportunités pour l'analyse, il est important de considérer d'ors et déjà son utilisation et son application pour d'éventuelles analyse de routine. En effet le développement technologique est rapide et l'utilisation d'unité de séquençage de type MinION (Oxford Nanopore) au sein de laboratoires de contrôle peut être envisagée dans un futur assez proche. Il est donc important que les techniques de préparation de l'échantillon permettent d'obtenir des ARN nettoyés autant que possible des matrices et de bonne qualité afin d'obtenir des séquences de bonne qualité également. Considérant la variabilité des virus à ARN, l'identification est plus précise sur des séquences

longues et avec une bonne profondeur. Trouver deux reads d'un virus dans un extrait d'aliment impliqué dans une épidémie peut conforter l'étude épidémiologique, mais en revanche, peut poser des questions pour la gestion du risque en dehors de tout contexte.

Considérant les NoVs, sujet de ma thèse, le développement de la mise en évidence de l'infectiosité pourra aider à répondre en partie à ces questions. En effet si la technique sur entéroïde est encore complexe et délicate à appliquer, elle va permettre d'acquérir des données sur la persistance de ce virus dans certaines conditions. Comme démontré lors de mon travail de thèse, en collaboration avec Marion Desdouits, les NoVs peuvent rester infectieux pendant plus de 20 jours dans l'eau de mer à 12°C. Cette grande résistance supposée par les observations épidémiologiques, va ainsi pouvoir être caractérisée. Un des prochains défis sera de mettre au point cette méthode pour suivre l'infectiosité dans les huîtres. Pour cela un développement méthodologique pour purifier et concentrer les NoVs est nécessaire. Ces développements pourront s'appuyer sur les approches développées pour l'analyse par métagénomique (élimination de la matrice, élution des bactéries...). L'autre point d'écueil réside dans le peu de souches des NoVs cultivables à l'heure actuelle. Ce modèle va certainement se simplifier dans les années à venir et permettra l'identification du récepteur cellulaire, important pour comprendre la physiologie de ce virus. Pour notre thématique, ces aspects fondamentaux aideront aux développements de tests complémentaires pour caractériser l'état physique de la capsid (intégrité du site reconnaissant le récepteur cellulaire par exemple). Ils devraient également faciliter la validation de virus de substitution présentant des caractéristiques de résistance similaire mais plus facilement multipliable en culture cellulaire.

Après des années où seule la technique de PCR permettait de détecter les NoVs dans l'environnement, ces avancées technologiques que constituent la métagénomique et les entéroïdes ouvrent des perspectives nouvelles pour analyser leur devenir et évaluer l'impact du microbiome de l'environnement sur leur capacité de persistance. Ces outils vont permettre de répondre plus efficacement au challenge que constitue l'étude de ces virus à ARN présentant une forte diversité génomique, des différences d'évolution selon les souches, une sensibilité génétique humaine, et pour lesquels l'environnement joue un rôle essentiel.

REFERENCES

- Abdessemed, D., Nezzal, G. and Aïm, R. Ben (1999) 'Treatment of wastewater by ultrafiltration', *Desalination*, 126(1–3), pp. 1–5. doi: 10.1016/S0011-9164(99)00149-6.
- Adriaenssens, E. M. et al. (2018) 'Viromic Analysis of Wastewater Input to a River Catchment Reveals a Diverse Assemblage of RNA Viruses', *mSystems*, 3(3), pp. 1–18. doi: 10.1128/msystems.00025-18.
- Afolayan, O.T., Webb, C.C., Cannon, J.L. (2016). Evaluation of a Porcine Gastric Mucin and RNase A Assay for the Discrimination of Infectious and Non-infectious GI.1 and GI.4 Norovirus Following Thermal, Ethanol, or Levulinic Acid Plus Sodium Dodecyl Sulfate Treatments. *Food Environ Virol* 8, pp. 70-78.
- Ahmed, S. M., Lopman, B. A. and Levy, K. (2013) 'A Systematic Review and Meta-Analysis of the Global Seasonality of Norovirus', *PLoS ONE*, 8(10), pp. 1–7. doi: 10.1371/journal.pone.0075922.
- Aigrain, L., Gu, Y. and Quail, M. A. (2016) 'Quantitation of next generation sequencing library preparation protocol efficiencies using droplet digital PCR assays - a systematic comparison of DNA library preparation kits for Illumina sequencing', *BMC Genomics*. *BMC Genomics*, 17(1), pp. 1–11. doi: 10.1186/s12864-016-2757-4.
- Almand, E. A., Moore, M. D. and Jaykus, L.-A. (2017) 'Norovirus Binding to Ligands Beyond Histo-Blood Group Antigens', *Frontiers in Microbiology*, 8(December), pp. 1–10. doi: 10.3389/fmicb.2017.02549.
- Aoki, Y. et al. (2010) 'Duration of norovirus excretion and the longitudinal course of viral load in norovirus-infected elderly patients', *Journal of Hospital Infection*. Elsevier Ltd, 75(1), pp. 42–46. doi: 10.1016/j.jhin.2009.12.016.
- Apone, L., Dimalanta E., and Stewart F. (2017). 'Improving Enzymatic DNA Fragmentation for Next Generation Sequencing Library Construction'. Feature Article. New England Biolabs, Inc. https://international.neb.com/-/media/nebus/files/feature-articles/nebexpressions_feature_nebnext_fs_issueiii_2017.pdf?la=en&rev=fd17aa21b2a54859badffd521653026a&hash=8CFA22E95B085233BD55FDFCAC462790C6338767
- Arthur, S. E. and Gibson, K. E. (2015) 'Physicochemical stability profile of Tulane virus: A human norovirus surrogate', *Journal of Applied Microbiology*, 119(3), pp. 868–875. doi: 10.1111/jam.12878.
- Ashbolt, N. J., Grabow, W. O. K. and Snozzi, M. (2001) 'World Health Organization (WHO). Water Quality: Guidelines, Standards and Health', in Fewtrell, L. and Bartram, J. (eds) *Indicators of microbial water quality*. London, UK: IWA Publishing, pp. 289–316.
- Atmar, R. L. (2010) 'Noroviruses - State of the Art', *Food Environ Virology*, 2(3), pp. 117–126. doi: 10.1007/s12560-010-9038-1.Noroviruses.
- Atmar, R.L., Baric, R.S., Estes, M.K., Green, K.C., Glass, R.I., Hall, A.J., Kang, C., Lee, B.A., Riddle, M., & Vinjé, J. (2015). 'Global Burden of Norovirus and Prospects for Vaccine Development'. CDC Report. https://pdfs.semanticscholar.org/0f7c/9599363cb3fbdd0081377f22f3d934c68be6.pdf?_ga=2.160271737.516663841.1573555211-330753142.1573555211
- Atmar, R. L. et al. (2014) 'Determination of the 50% human infectious dose for norwalk virus', *Journal of Infectious Diseases*, 209(7), pp. 1016–1022. doi: 10.1093/infdis/jit620.

- Atmar, R. L., Ramani, S. and Estes, M. K. (2018) 'Human noroviruses: recent advances in a 50-year history', *Current Opinion in Infectious Diseases*, p. 1. doi: 10.1097/qco.0000000000000476.
- Atmar, R.L., Neil, F.H., Romalde, J.L., Le Guyader, F., Woodley, C.M., Metcalf, T.G., Estes, M.K., (1995). Detection of Norwalk virus and Hepatitis A Virus in Shellfish Tissues with the PCR. *Applied and Environmental Microbiology* 61, pp. 3014-3018.
- Ayling, M., Clark, M. D. and Leggett, R. M. (2019) 'New approaches for metagenome assembly with short reads', *Briefings in Bioinformatics*, 00(February), pp. 1–11. doi: 10.1093/bib/bbz020.
- Bankevich, A. et al. (2012) 'SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing', *Journal of Computational Biology*, 19(5), pp. 455–477. doi: 10.1089/cmb.2012.0021.
- Bartsch, C. et al. (2018) 'Analysis of frozen strawberries involved in a large norovirus gastroenteritis outbreak using next generation sequencing and digital PCR', *Food Microbiology*. Elsevier Ltd, 76(July), pp. 390–395. doi: 10.1016/j.fm.2018.06.019.
- Batovska, J. et al. (2017) 'Metagenomic arbovirus detection using MinION nanopore sequencing', *Journal of Virological Methods*. Elsevier, 249(August), pp. 79–84. doi: 10.1016/j.jviromet.2017.08.019.
- Beier, R. et al. (2014) 'Selection of a DNA aptamer against norovirus capsid protein VP1', *FEMS Microbiology Letters*, 351(2), pp. 162–169. doi: 10.1111/1574-6968.12366.
- Belliot, G. et al. (2010) 'Evidence of emergence of new GGII.4 norovirus variants from Gastroenteritis Outbreak Survey in France during the 2007-to-2008 and 2008-to-2009 winter seasons', *Journal of Clinical Microbiology*, 48(3), pp. 994–998. doi: 10.1128/JCM.01888-09.
- Bisseux, M. et al. (2018) 'Monitoring human enteric viruses in wastewater and relevance to infections encountered in the clinical setting: A one-year experiment in central France, 2014 to 2015', *Eurosurveillance*, 23(7), pp. 1–11. doi: 10.2807/1560-7917.ES.2018.23.7.17-00237.
- Bonnardin, S. (2012) 'Parures de coquillages du néolithique en Europe (vie-ve millénaires av. J.-C.)', *Techniques & culture*, (59), pp. 26–43. doi: 10.4000/tc.6518.
- Boom, R., Sol, C.J., Salimans, M.M., Jansen, C.L., Wertheim-van Dillen, P.M., van der Noordaa, J., (1990). Rapid and simple method for purification of nucleic acids. *J Clin Microbiol* 28, pp. 495-503.
- Boone, M., De Koker, A. and Callewaert, N. (2018) 'Survey and summary capturing the “ome”: The expanding molecular toolbox for RNA and DNA library construction', *Nucleic Acids Research*. Oxford University Press, 46(6), pp. 2701–2721. doi: 10.1093/nar/gky167.
- Borrego, J.J., Cornax, R., Preston, D.R., Farrah, S.R., McEl-haney, B., and Bitton, G. (1991) Development and application of new positively charged filters for recovery of bacteriophages from water. *Appl Environ Microbiol* 57, pp. 1218–1222.
- Bosch, A., Guix, S., Sano, D., & Pintó, R. M. (2008). New tools for the study and direct surveillance of viral pathogens in water. *Current Opinion in Biotechnology*, 19(3), pp. 295–301. <https://doi.org/10.1016/j.copbio.2008.04.006>
- Boutin, C. (2009) 'Technologies d'épuration en vue d'une réutilisation des eaux usées traitées (REUT) Rapport final', *Convention de partenariat ONEMA-Cemagref 2008; action 28 : 'Réutilisation des eaux traitées'*, p. 100. Available at: http://www.onema.fr/IMG/pdf/2009_038.pdf.
- Brake F, Ross T, Holds G, Kiermeier A, McLeod C. (2015). A survey of Australian oysters for the presence of noroviruses. *Food Microbiol*. 44: 264-270.

- Branco, P. M. P. et al. (2016) 'Efficiency of biological treatment for the removal of total and thermotolerant coliform bacteria from domestic sewage', *Bioscience Journal*, pp. 1572–1577. doi: 10.14393/bj-v32n6a2016-33292.
- Briese, T. et al. (2015) 'Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis', *mBio*, 6(5), pp. 1–11. doi: 10.1128/mBio.01491-15.
- Brinkmann, A. et al. (2019) 'Proficiency Testing of Virus Diagnostics Based on Bioinformatics Analysis of Simulated In Silico High-Throughput Sequencing Data Sets', 57(8), pp. 1–12.
- Bronner, I. F. et al. (2013) Improved protocols for Illumina sequencing, *Current Protocols in Human Genetics*. doi: 10.1002/0471142905.hg1802s80.
- Brown, J. R. et al. (2016) 'Norovirus Whole-Genome Sequencing by SureSelect Target Enrichment : a Robust and Sensitive Method', 54(10), pp. 2530–2537. doi: 10.1128/JCM.01052-16.Editor.
- Buermans, H. P. J. and den Dunnen, J. T. (2014) 'Next generation sequencing technology: Advances and applications', *Biochimica et Biophysica Acta - Molecular Basis of Disease*. Elsevier B.V., 1842(10), pp. 1932–1941. doi: 10.1016/j.bbadis.2014.06.015.
- Burke, R. M. et al. (2018) 'The Norovirus Epidemiologic Triad: Predictors of Severe Outcomes in US Norovirus Outbreaks, 2009–2016', *The Journal of Infectious Diseases*, 30329, pp. 1–9. doi: 10.1093/infdis/jiy569.
- Bustin, S. et al. (2015) 'Variability of the reverse transcription step: Practical implications', *Clinical Chemistry*, 61(1), pp. 202–212. doi: 10.1373/clinchem.2014.230615.
- Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J., Wittwer, C.T. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* 55, pp. 611–622.
- Calgua, B. et al. (2008) 'Development and application of a one-step low cost procedure to concentrate viruses from seawater samples', *Journal of Virological Methods*, 153(2), pp. 79–83. doi: 10.1016/j.jviromet.2008.08.003.
- Cannon, J.L., Vinjé, J. (2008). Histo-Blood Group Antigen Assay for Detecting Noroviruses in Water. *Applied and Environmental Microbiology* 74, pp. 6818-6819.
- Cashdollar, J. L., & Wymer, L. (2013). Methods for primary concentration of viruses from water samples: A review and meta-analysis of recent studies. *Journal of Applied Microbiology*, 115(1), pp. 1–11. <https://doi.org/10.1111/jam.12143>
- Chan, M. C. W., Kwan, H. S. and Chan, P. K. S. (2016) Structure and Genotypes of Noroviruses, *The Norovirus: Features, Detection, and Prevention of Foodborne Disease*. Elsevier Inc. doi: 10.1016/B978-0-12-804177-2.00004-X.
- Chan, M. C.-W. et al. (2019) 'Use of Human Intestinal Enteroids to Detect Human Norovirus Infectivity', *Emerging Infectious Diseases*, 25(9), pp. 1–4. doi: 10.3201/eid2509.190205.
- Cheetham, S. et al. (2006) 'Pathogenesis of a Genogroup II Human Norovirus in Gnotobiotic Pigs', *Journal of Virology*, 80(21), pp. 10372–10381. doi: 10.1128/JVI.00809-06.
- Chhabra, P. et al. (2019) 'Updated classification of norovirus genogroups and genotypes', *Journal of General Virology*, 1–14, p. jgv001318. doi: 10.1099/JGV.0.001318.
- Choi, J.-M. et al. (2008) 'Atomic resolution structural characterization of recognition of histo-blood

- group antigens by Norwalk virus.', *Proceedings of the National Academy of Sciences of the United States of America*, 105(27), pp. 9175–9180. doi: 10.1073/pnas.0803275105.
- Chung, S. et al. (2019) 'Smartphone-Based Paper Microfluidic Particulometry of Norovirus from Environmental Water Samples at the Single Copy Level', *ACS Omega*, 4(6), pp. 11180–11188. doi: 10.1021/acsomega.9b00772.
- Conceição-Neto, N. et al. (2015) 'Modular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis', *Scientific reports*. Nature Publishing Group, 5(November), p. 16532. doi: 10.1038/srep16532.
- Cook, N., Knight, A. and Richards, G. P. (2016) 'Persistence and Elimination of Human Norovirus in Food and on Food Contact Surfaces: A Critical Review', *Journal of Food Protection*, 79(7), pp. 1273–1294. doi: 10.4315/0362-028X.JFP-15-570.
- Cortés-Sánchez, M. et al. (2011) 'Earliest known use of marine resources by Neanderthals.', *PLoS one*, 6(9), p. e24026. doi: 10.1371/journal.pone.0024026.
- Costantini, V. et al. (2018) 'Human norovirus replication in human intestinal enteroids as model to evaluate virus inactivation', *Emerging Infectious Diseases*, 24(8), pp. 1453–1464. doi: 10.3201/eid2408.180126.
- Cotten, M. and Koopmans, M. (2016) 'Next-generation sequencing and norovirus', *Future Virology*, 11(11), pp. 719–722. doi: 10.2217/fvl-2016-0099.
- Cotten, M. et al. (2014) 'Full genome virus detection in fecal samples using sensitive nucleic acid preparation, deep sequencing, and a novel iterative sequence classification algorithm', *PLoS ONE*, 9(4). doi: 10.1371/journal.pone.0093269.
- Coudray-Meunier, C. et al. (2016) 'A novel high-throughput method for molecular detection of human pathogenic viruses using a nanofluidic real-time PCR system', *PLoS ONE*, 11(1), pp. 1–17. doi: 10.1371/journal.pone.0147832.
- Cromeans, T. et al. (2014) 'Comprehensive comparison of cultivable norovirus surrogates in response to different inactivation and disinfection treatments', *Applied and Environmental Microbiology*, 80(18), pp. 5743–5751. doi: 10.1128/AEM.01532-14.
- Da Silva, A. K. et al. (2011) 'Adsorption and aggregation properties of norovirus GI and GII virus-like particles demonstrate differing responses to solution chemistry', *Environmental Science and Technology*, 45(2), pp. 520–526. doi: 10.1021/es102368d.
- Dancho, B.A., Chen, H., Kingsley, D.H., 2012. Discrimination between infectious and non-infectious human norovirus using porcine gastric mucin. *Int J Food Microbiol* 155, pp. 222-226.
- De Graaf, M., Van Beek, J. and Koopmans, M. P. G. (2016) 'Human norovirus transmission and evolution in a changing world', *Nature Reviews Microbiology*. Nature Publishing Group, 14(7), pp. 421–433. doi: 10.1038/nrmicro.2016.48.
- De Graaf, M., Villabruna, N. and Koopmans, M. P. (2017) 'Capturing norovirus transmission', *Current Opinion in Virology*. The Author(s), 22, pp. 64–70. doi: 10.1016/j.coviro.2016.11.008.
- Dekker, J. et al. (2002) 'The MUC family: An obituary', *Trends in Biochemical Sciences*, 27(3), pp. 126–131. doi: 10.1016/S0968-0004(01)02052-7.
- Desselberger, U. (2019) 'Caliciviridae other than noroviruses', *Viruses*, 11(3). doi: 10.3390/v11030286.

- Diez-Valcarce M, Kokkinos P, Soderberg K, Bouwknecht M, Willems K, de Roda-Husman A-M, von Bonsdorff C-H, Bellou M, Hernandez M, Maunula L, Vantarakis A, Rodriguez-Lazaro D. (2012). Occurrence of human enteric viruses in commercial mussels at retail level in three European countries. *Food Environ. Virol.* 4: 73-80.
- Dolin, R. et al. (1975) 'Viral gastroenteritis induced by the Hawaii agent. Jejunal histopathology and serologic response', *The American Journal of Medicine*, 59(6), pp. 761–768. doi: 10.1016/0002-9343(75)90461-1.
- Doré, B. et al. (2010) 'Management of health risks associated with oysters harvested from a norovirus contaminated area, Ireland, February-March 2010', *Eurosurveillance*, 15(19), pp. 1–4.
- Drouaz, N. et al. (2015) 'Tulane virus as a potential surrogate to mimic norovirus behavior in oysters', *Applied and Environmental Microbiology*, 81(15), pp. 5249–5256. doi: 10.1128/AEM.01067-15.
- EFSA (2019) 'Analysis of the European baseline survey of norovirus in oysters', *EFSA Journal*, 17(7), pp. 1–99. doi: 10.2903/j.efsa.2019.5762.
- Eftim, S. E. et al. (2017) 'Occurrence of norovirus in raw sewage – A systematic literature review and meta-analysis', *Water Research*. Elsevier Ltd, 111, pp. 366–374. doi: 10.1016/j.watres.2017.01.017.
- Eid, J. et al. (2009) 'Single Polymerase Molecules', *Science (New York, N.Y.)*, 323(January), pp. 133–138. doi: 10.1126/science.1162986.
- El Moqri N, El Mellouli F, Hassou N, Benhafid M, Abouchoaib N, Etahiri S. (2019). Norovirus detection at Oualidia lagoon, a Moroccan shellfish harvesting area, by reverse transcription PCR analysis. *Food Env. Virol.* on line firsts
- Endoh, D. et al. (2005) 'Species-independent detection of RNA virus by representational difference analysis using non-ribosomal hexanucleotides for reverse transcription', *Nucleic Acids Research*, 33(6), pp. 1–11. doi: 10.1093/nar/gni064.
- Escudero-Abarca, B. I. et al. (2014) 'Selection, characterization and application of nucleic acid aptamers for the capture and detection of human norovirus strains', *PLoS ONE*, 9(9). doi: 10.1371/journal.pone.0106805.
- Estes, M. K. et al. (2019) 'Human Norovirus Cultivation in Nontransformed Stem Cell-Derived Human Intestinal Enteroid Cultures : Success and Challenges', 11.
- Ettayebi, K. et al. (2016) 'Replication of human noroviruses in stem cell – derived human enteroids', 5211(August). doi: 10.1126/science.aaf5211.
- European Food Safety Authority (2012) 'Scientific Opinion on Norovirus (NoV) in oysters : methods , limits and control options', *EFSA Journal*, 10(1), pp. 2500–2539. doi: 10.2903/j.efsa.2012.2500.Available.
- F Sanger and AR Coulson (1975) 'A Rapid Method for Determining Sequences in DNA by Primed Synthesis with DNA Polymerase', *Journal of Molecular Biology*, 94(3), pp. 441–448.
- Falman, J. C. et al. (2019) 'Evaluation of Secondary Concentration Methods for Poliovirus Detection in Wastewater', *Food and Environmental Virology*. Springer US, 11(1), pp. 20–31. doi: 10.1007/s12560-018-09364-y.
- Farkas, K., McDonald, J., Malham, S., & Jones, D. (2018). Two-Step Concentration of Complex Water Samples for the Detection of Viruses. *Methods and Protocols*, 1(3), 35.

<https://doi.org/10.3390/mps1030035>

Farkas, T. (2015) 'Rhesus enteric calicivirus surrogate model for human norovirus gastroenteritis', *Journal of General Virology*, 96(7), pp. 1504–1514. doi: 10.1099/jgv.0.000020.

Farkas, T. et al. (2008) 'Characterization of a Rhesus Monkey Calicivirus Representing a New Genus of Caliciviridae', *Journal of Virology*, 82(11), pp. 5408–5416. doi: 10.1128/JVI.00070-08.

Fernandez-Cassi, X. et al. (2017) 'Metagenomics for the study of viruses in urban sewage as a tool for public health surveillance', *Science of the Total Environment*, 618(November), pp. 870–880. doi: 10.1016/j.scitotenv.2017.08.249.

Fonager, J. et al. (2017) 'A universal primer-independent next-generation sequencing approach for investigations of norovirus outbreaks and novel variants', *Scientific Reports*, 7(1), pp. 1–11. doi: 10.1038/s41598-017-00926-x.

Fujii, M., Clevers, H. and Sato, T. (2019) 'Modeling Human Digestive Diseases With CRISPR-Cas9–Modified Organoids', *Gastroenterology*. Elsevier, Inc, 156(3), pp. 562–576. doi: 10.1053/j.gastro.2018.11.048.

Fumian, T. M. et al. (2019) 'Detection of norovirus epidemic genotypes in raw sewage using next generation sequencing', *Environment International*. Elsevier, 123(November 2018), pp. 282–291. doi: 10.1016/j.envint.2018.11.054.

Fusco G, di Bartolo H, Cioffi B, Ianiro G, Palermo P, Monini M, Amoroso MG. (2017). Prevalence of foodborne viruses in mussels in southern Italy. *Food Env. Virol.* 9: 187-194.

Fusco, G. et al. (2017) 'Prevalence of Foodborne Viruses in Mussels in Southern Italy', *Food and Environmental Virology*. Springer US, 9(2), pp. 187–194. doi: 10.1007/s12560-016-9277-x.

Garalde, D. R. et al. (2018) 'Highly parallel direct RNA sequencing on an array of nanopores', *Nature Methods*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 15, p. 201. Available at: <https://doi.org/10.1038/nmeth.4577>.

Garry Pascal (2015). Synthèse de la journée Santé Environnement et Microbiologie 2015. Ifremer, Laboratoire Santé Environnement et Microbiologie ; Laboratoire National de Référence de Microbiologie des coquillages. <https://archimer.ifremer.fr/doc/00302/41358/40550.pdf>

Gaudin, M. and Desnues, C. (2018) 'Hybrid capture-based next generation sequencing and its application to human infectious diseases', *Frontiers in Microbiology*, 9(NOV), pp. 1–9. doi: 10.3389/fmicb.2018.02924.

George, J. (1916) 'The influence of the weather on the incidence of certain infectious disease', *The british medical journal*.

Gerba, C. P. et al. (2018) 'Reducing uncertainty in estimating virus reduction by advanced water treatment processes', *Water Research*. Elsevier Ltd, 133, pp. 282–288. doi: 10.1016/j.watres.2018.01.044.

Giamberardino, A., Labib, M., Hassan, E.M., Tetro, J.A., Springthorpe, S., Sattar, S.A., Berezovski, M.V., DeRosa, M.C. (2013). Ultrasensitive norovirus detection using DNA aptasensor technology. *PLoS One* 8, e79087.

Gibbons, C. D. et al. (2010) 'Evaluation of positively charged alumina nanofibre cartridge filters for the primary concentration of noroviruses, adenoviruses and male-specific coliphages from seawater', *Journal of Applied Microbiology*, 109(2), pp. 635–641. doi: 10.1111/j.1365-2672.2010.04691.x.

- Gibson, K.E. and Schwab, K.J. (2011) Tangential-flow ultrafiltration with integrated inhibition detection for recovery of surrogates and human pathogens from large-volume source water and finished drinking water. *Appl Environ Microbiol* 77, pp. 385–391.
- Graziano, V. R., Wei, J. and Wilen, C. B. (2019) 'Norovirus Attachment and Entry', *Viruses*, 11(6), pp. 1–13. doi: 10.3390/v11060495.
- Greninger, A. L. et al. (2015) 'Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis', *Genome Medicine*. *Genome Medicine*, 7(1), pp. 1–13. doi: 10.1186/s13073-015-0220-9.
- Grøndahl-Rosado, R. C. et al. (2014) 'A One Year Study on the Concentrations of Norovirus and Enteric Adenoviruses in Wastewater and A Surface Drinking Water Source in Norway', *Food and Environmental Virology*, 6(4), pp. 232–245. doi: 10.1007/s12560-014-9161-5.
- Gruber, K. (2015). Here, there, and everywhere: from PCRs to next-generation sequencing technologies and sequence databases, DNA contaminants creep in from the most unlikely places. *EMBO Rep.* 16, pp. 898–901. doi: 10.15252/embr. 201540822
- Gyawali, P. et al. (2019) 'Current and Emerging Technologies for the Detection of Norovirus from Shellfish', *Foods*, 8(6), p. 187. doi: 10.3390/foods8060187.
- Hagström, A. E. V. et al. (2015) 'Sensitive detection of norovirus using phage nanoparticle reporters in lateral-flow assay', *PLoS ONE*, 10(5), pp. 1–11. doi: 10.1371/journal.pone.0126571.
- Hall, A. J. (2012) 'Noroviruses: The perfect human pathogens?', *Journal of Infectious Diseases*, 205(11), pp. 1622–1624. doi: 10.1093/infdis/jis251.
- Hardstaff, J. L. et al. (2018) 'Foodborne and Food-Handler Norovirus Outbreaks: A Systematic Review', *Foodborne Pathogens and Disease*, 15(10), pp. 589–597. doi: 10.1089/fpd.2018.2452.
- Hardy, M. E. (2005) 'Norovirus protein structure and function', *FEMS Microbiology Letters*, 253, pp. 1–8. doi: 10.1016/j.femsle.2005.08.031.
- Hassard, F. et al. (2017) 'Critical Review on the Public Health Impact of Norovirus Contamination in Shellfish and the Environment: A UK Perspective', *Food and Environmental Virology*. Springer US, 9(2), pp. 123–141. doi: 10.1007/s12560-017-9279-3.
- Hjelmsø, M. H. et al. (2017) 'Evaluation of methods for the concentration and extraction of viruses from sewage in the context of metagenomic sequencing', *PLoS ONE*, 12(1), pp. 1–17. doi: 10.1371/journal.pone.0170199.
- Hoque, Sheikh Ariful et al. (2019) 'Alarming Situation of Spreading Enteric Viruses Through Sewage Water in Dhaka City: Molecular Epidemiological Evidences', *Food and Environmental Virology*. Springer US, 11(1), pp. 65–75. doi: 10.1007/s12560-018-09363-z.
- Huson, D. H. et al. (2016) 'MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data', *PLoS Computational Biology*, 12(6), pp. 1–12. doi: 10.1371/journal.pcbi.1004957.
- Ignatov, K. B. et al. (2019) 'Fragmentation through polymerization (FTP): A new method to fragment DNA for next-generation sequencing', *PLoS ONE*, 14(4), pp. 1–12. doi: 10.1371/journal.pone.0210374.
- Ikner, L. A., Gerba, C. P., & Bright, K. R. (2012). Concentration and Recovery of Viruses from Water: A Comprehensive Review. *Food and Environmental Virology*, 4(2), pp. 41–67. <https://doi.org/10.1007/s12560-012-9080-2>

- Imamura S, Kanezashi H, Goshima T, Haruna M, Okada T, Inagaki N, Uema M, Noda M, Akimoto K. (2017). Next-generation sequencing analysis of the diversity of human noroviruses in Japanese oysters. *FoodBorne Path. Dis.* 14, pp. 465-471.
- Imamura, S. et al. (2016) 'Application of next-generation sequencing to investigation of norovirus diversity in shellfish collected from two coastal sites in Japan from 2013 to 2014', *Japanese Journal of Veterinary Research*, 64(2), pp. 113–122. doi: 10.14943/jjvr.64.2.113.
- Ishii, S. et al. (2014) 'Microfluidic quantitative PCR for simultaneous quantification of multiple viruses in environmental water samples', *Applied and Environmental Microbiology*, 80(24), pp. 7505–7511. doi: 10.1128/AEM.02578-14.
- ISO (2017) 'Microbiology of the food chain — Horizontal method for determination of hepatitis A virus and norovirus using real-time RT-PCR'.
- Jacobsson, K. and Schultz, A. C. (2014) 'Sample Processing', *Waterborne Pathogens*. Academic Press, pp. 67–114. doi: 10.1016/B978-0-444-59543-0.00004-9.
- Jain, M. et al. (2015) 'Improved data analysis for the MinION nanopore sequencer', *Nature Methods*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 12, p. 351. Available at: <https://doi.org/10.1038/nmeth.3290>.
- Jennings, L. J. et al. (2017) 'Guidelines for Validation of Next-Generation Sequencing–Based Oncology Panels: A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists', *Journal of Molecular Diagnostics*. American Society for Investigative Pathology and the Association for Molecular Pathology, 19(3), pp. 341–365. doi: 10.1016/j.jmoldx.2017.01.011.
- Jiang, X. et al. (1992) 'Expression, self-assembly, and antigenicity of the Norwalk virus capsid protein.', *Journal of virology*, 66(11), pp. 6527–32. doi: 10.1021/la8017419.
- John, S. G., Mendez, C. B., Deng, L., Poulos, B., Kauffman, A. K. M., Kern, S., ... Sullivan, M. B. (2011). A simple and efficient method for concentration of ocean viruses by chemical flocculation. *Environmental Microbiology Reports*, 3(2), pp. 195–202. <https://doi.org/10.1111/j.1758-2229.2010.00208>.
- Jothikumar, N., Lowther, J.A., Henshilwood, K., Lees, D.N., Hill, V.R., Vinje, J. (2005). Rapid and sensitive detection of noroviruses by using TaqMan-based one-step reverse transcription-PCR assays and application to naturally contaminated shellfish samples. *Appl Environ Microbiol* 71, pp. 1870-1875.
- Kageyama, T. et al. (2003) 'Broadly reactive and highly sensitive assay for Norwalk-like viruses based on real-time quantitative reverse transcription-PCR', *Journal of Clinical Microbiology*, 41(4), pp. 1548–1557. doi: 10.1128/JCM.41.4.1548-1557.2003.
- Karandikar, U. C. et al. (2016) 'Detection of human norovirus in intestinal biopsies from immunocompromised transplant patients', *Journal of General Virology*, 97(9), pp. 2291–2300. doi: 10.1099/jgv.0.000545.
- Karim, M. R. et al. (2015) 'Propidium monoazide reverse transcriptase PCR and RT-qPCR for detecting infectious enterovirus and norovirus', *Journal of Virological Methods*. Elsevier B.V., 219, pp. 51–61. doi: 10.1016/j.jviromet.2015.02.020.
- Karst, S. M. (2010) Pathogenesis of noroviruses, emerging RNA viruses, *Viruses*. doi: 10.3390/v2030748.
- Karst, S. M. and Tibbetts, S. A. (2016) 'Recent advances in understanding norovirus pathogenesis',

- J Med Virol., 88(11), pp. 1837–1843. doi: 10.1002/jmv.24559.Recent.
- Karst, S. M. and Wobus, C. E. (2015) 'A Working Model of How Noroviruses Infect the Intestine', *PLoS Pathogens*, 11(2), pp. 1–7. doi: 10.1371/journal.ppat.1004626.
- Kauppinen, A., Pitkänen, T. and Miettinen, I. T. (2018) 'Persistent Norovirus Contamination of Groundwater Supplies in Two Waterborne Outbreaks', *Food and Environmental Virology*. Springer US, 10(1), pp. 39–50. doi: 10.1007/s12560-017-9320-6.
- Kazama, S. et al. (2016) 'Temporal dynamics of norovirus determined through monitoring of municipal wastewater by pyrosequencing and virological surveillance of gastroenteritis cases', *Water Research*. Elsevier Ltd, 92, pp. 244–253. doi: 10.1016/j.watres.2015.10.024.
- Kchouk, M., Gibrat, J. F. and Elloumi, M. (2017) 'Generations of Sequencing Technologies: From First to Next Generation', *Biology and Medicine*, 09(03). doi: 10.4172/0974-8369.1000395.
- Kittigul L, Thamjaroen A, Chiawchan S, Chavalitshewinkoon-Petmitr P, Pombubpa K, Diraphat P. (2016). Prevalence and molecular genotyping of noroviruses in market oysters, mussels, and cockles in Bangkok, Thailand. *Food Environ. Virol.* 8, pp. 113-140.
- Koo, E. S. et al. (2017) 'Occurrence of novel GII.17 and GII.21 norovirus variants in the coastal environment of South Korea in 2015', *PLoS ONE*, 12(2), pp. 1–16. doi: 10.1371/journal.pone.0172237.
- Kraut, R. Y. et al. (2017) 'Influence of School Year on Seasonality of Norovirus Outbreaks in Developed Countries', *Canadian Journal of Infectious Diseases and Medical Microbiology*. Hindawi Publishing Corporation, 2017, pp. 1–8. doi: 10.1155/2017/9258140.
- Krishnamurthy, S. R. and Wang, D. (2017) 'Origins and challenges of viral dark matter', *Virus Research*. Elsevier B.V., 239, pp. 136–142. doi: 10.1016/j.virusres.2017.02.002.
- Kroneman, A. et al. (2011) 'An automated genotyping tool for enteroviruses and noroviruses', *Journal of Clinical Virology*, 51(2), pp. 121–125. doi: 10.1016/j.jcv.2011.03.006.
- Kroneman, A. et al. (2013) 'Proposal for a unified norovirus nomenclature and genotyping', *Archives of Virology*, 158(10), pp. 2059–2068. doi: 10.1007/s00705-013-1708-5.
- Kuypers, J., and Jerome, K. R. (2017). *crossm Applications of Digital PCR for Clinical*. 55(6), pp. 1621–1628.
- La Rosa G, Della Libera S, Iaconelli M, Proroga YTR, De Medici D, Martella V, Suffredini E. (2017). Detection of norovirus GII.17 Kawasaki 2014 in shellfish, marine water and underwater discharges in Italy. *Food Env. Virol.* 9, pp. 326-333.
- Laurence, M., Hatzis, C., & Brash, D. E. (2014). Common contaminants in next-generation sequencing that hinder discovery of low-abundance microbes. *PLoS ONE*, 9(5), pp. 1–8. <https://doi.org/10.1371/journal.pone.0097876>
- Le Guyader, F. S., Atmar, R. L. and Le Pendu, J. (2012) 'Transmission of viruses through shellfish: When specific ligands come into play', *Current Opinion in Virology*, 2(1), pp. 103–110. doi: 10.1016/j.coviro.2011.10.029.
- Le Guyader, F. S., Bon, F., et al. (2006) 'Detection of multiple noroviruses associated with an international gastroenteritis outbreak linked to oyster consumption', *Journal of Clinical Microbiology*. American Society for Microbiology (ASM), 44(11), pp. 3878–3882. doi: 10.1128/JCM.01327-06.

- Le Guyader, F. S., Loisy, F., et al. (2006) 'Norwalk virus-specific binding to oyster digestive tissues', *Emerging Infectious Diseases*, 12(6), pp. 931–936. doi: 10.3201/eid1206.051519.
- Le Guyader, F. S., Parnaudeau, S., Schaeffer, J., Bosch, A., Loisy, F., Pommepeuy, M., & Atmar, R. L. (2009). Detection and quantification of noroviruses in shellfish. *Applied and Environmental Microbiology*, 75(3), pp. 618–624. <https://doi.org/10.1128/AEM.01507-08>
- Le Mennec, C., Parnaudeau, S., Rumebe, M., Le Saux, J.-C., Piquet, J.-C., & Le Guyader, S. F. (2017). Follow-Up of Norovirus Contamination in an Oyster Production Area Linked to Repeated Outbreaks. *Food and Environmental Virology*, 9(1), pp. 54–61. <https://doi.org/10.1007/s12560-016-9260-6>
- Le Pendu, J. et al. (2006) 'Mendelian resistance to human norovirus infections', *Seminars in Immunology*, 18(6), pp. 375–386. doi: 10.1016/j.smim.2006.07.009.
- Le Pendu, J., Nyström, K. and Ruvoën-Clouet, N. (2014) 'Host-pathogen co-evolution and glycan interactions', *Current Opinion in Virology*, 7(1), pp. 88–94. doi: 10.1016/j.coviro.2014.06.001.
- Lee, L. J., and Abdullah, M. (2014). Optimization of DNA shearing by sonication *AsPac*. *AsPac J. Mol. Biol. Biotechnol*, 22(3), 200–208. Retrieved from <https://pdfs.semanticscholar.org/b8bc/e77382a2c47c52bbbf38aa400a6bbcb4b6a.pdf>
- Lees, D. (2010) 'International standardisation of a method for detection of human pathogenic viruses in molluscan shellfish', *Food and Environmental Virology*, 2(3), pp. 146–155. doi: 10.1007/s12560-010-9042-5.
- Lescroart, M. (2017) *Les huîtres*. Quae. Collection clés pour comprendre.
- Lewandowska, D. W. et al. (2017) 'Optimization and validation of sample preparation for metagenomic sequencing of viruses in clinical samples', *Microbiome*. *Microbiome*, 5(1), p. 94. doi: 10.1186/s40168-017-0317-z.
- Lewis, G. D. and Metcalf, T. G. (1988) 'Polyethylene glycol precipitation for recovery of pathogenic viruses, including hepatitis A virus and human rotavirus, from oyster, water, and sediment samples', *Applied and Environmental Microbiology*, 54(8), pp. 1983–1988. doi: 0099-2240/88/081983-06\$02.00/0.
- Li, M. et al. (2018) 'Identification of Biological Hazards in Produce Consumed in Industrialized Countries: A Review', *Journal of Food Protection*, 81(7), pp. 1171–1186. doi: 10.4315/0362-028X.JFP-17-465.
- Limacher, J.M and Acres, M. 2007. La protéine MUC1, cible thérapeutique en oncologie. *Bulletin du cancer* 94(3), 253-7. DOI : 10.1684/bdc.2007.0214
- Lin, Y. Y. et al. (2017) 'De novo assembly of highly polymorphic metagenomic data using in situ generated reference sequences and a novel BLAST-based assembly pipeline', *BMC Bioinformatics*. *BMC Bioinformatics*, 18(1), pp. 1–10. doi: 10.1186/s12859-017-1630-z.
- Liu, L. et al. (2014) 'Comparison of next-generation sequencing systems', *The Role of Bioinformatics in Agriculture*, 2012, pp. 1–25. doi: 10.1201/b16568.
- Loisy, F., Atmar, R. L., Guillon, P., Le Cann, P., Pommepeuy, M., & Le Guyader, F. S. (2005). Real-time RT-PCR for norovirus screening in shellfish. *Journal of Virological Methods*, 123(1), pp. 1–7. <https://doi.org/10.1016/j.jviromet.2004.08.023>
- Lowther J, Cross L, Stapleton T, Gustar NE, Walker DI, Sills M, Treagus S, Pollington V, Lees DN. (2019). "Use of F-specific RNA bacteriophage to estimate infectious norovirus levels in oysters."

Lowther JA, Gustar NE, Powell AL, Hartnell RE, Lees DN. (2012). Two-year systematic study to assess norovirus contamination in oysters from commercial harvesting areas in the United Kingdom. *Appl. Environ. Microbiol.* 78, pp. 5812-5817.

Lowther, J. A. et al. (2019) 'Validation of EN ISO method 15216 - Part 1 – Quantification of hepatitis A virus and norovirus in food matrices', *International Journal of Food Microbiology*. Elsevier, 288(November 2017), pp. 82–90. doi: 10.1016/j.ijfoodmicro.2017.11.014.

Lugli, N. et al. (2016) 'R-Spondin 1 and noggin facilitate expansion of resident stem cells from non-damaged gallbladders', pp. 769–779. doi: 10.15252/embr.201642169.

Lusk, R. W. (2014). Diverse and widespread contamination evident in the unmapped depths of high throughput sequencing data. *PLoS ONE* 9:e110808. doi: 10.1371/journal.pone.0110808

Maalouf, H. et al. (2010) 'Distribution in tissue and seasonal variation of norovirus genogroup I and II ligands in oysters', *Applied and Environmental Microbiology*, 76(16), pp. 5621–5630. doi: 10.1128/AEM.00148-10.

Maalouf, H. et al. (2011) 'Strain-dependent norovirus bioaccumulation in oysters', *Applied and Environmental Microbiology*, 77(10), pp. 3189–3196. doi: 10.1128/AEM.03010-10.

Madsen, J. B. et al. (2016) 'Structural and mechanical properties of thin films of bovine submaxillary mucin versus porcine gastric mucin on a hydrophobic surface in aqueous solutions', *Langmuir*, 32(38), pp. 9687–9696. doi: 10.1021/acs.langmuir.6b02057.

Mah, A. T., Yan, K. S. and Kuo, C. J. (2016) 'Wnt pathway regulation of intestinal stem cells', *Journal of Physiology*, 594(17), pp. 4837–4847. doi: 10.1113/JP271754.

Mahe, M. M. et al. (2015) 'Establishment of human epithelial enteroids and colonoids from whole tissue and biopsy', *Journal of Visualized Experiments*, 2015(97), pp. 1–13. doi: 10.3791/52483.

Makison Booth, C. (2014) 'Vomiting Larry: a simulated vomiting system for assessing environmental contamination from projectile vomiting related to norovirus infection', *Journal of Infection Prevention*, 15(5), pp. 176–180. doi: 10.1177/1757177414545390.

Marine, R. L. et al. (2019) 'Comparison of Illumina MiSeq and the Ion Torrent PGM and S5 platforms for whole-genome sequencing of picornaviruses and caliciviruses Running Title: Comparison of NGS platforms for sequencing RNA viruses Rachel L. Marine', pp. 1–36. doi: <http://dx.doi.org/10.1101/705632>.

Masago, Y. et al. (2016) 'Comparative evaluation of real-time PCR methods for human noroviruses in wastewater and human stool', *PLoS ONE*, 11(8), pp. 1–10. doi: 10.1371/journal.pone.0160825.

Matrajt, G. et al. (2018) 'A Review of the Most Commonly Used Methods for Sample Collection in Environmental Surveillance of Poliovirus', *Clinical Infectious Diseases*, 67(Suppl 1), pp. S90–S97. doi: 10.1093/cid/ciy638.

McGuckin, M. A. et al. (2011) 'Mucin dynamics and enteric pathogens', *Nature Reviews Microbiology*. Nature Publishing Group, 9(4), pp. 265–278. doi: 10.1038/nrmicro2538.

McIntre, L. et al. (2012) 'Multiple Clusters of Norovirus among Shellfish Consumers Linked to Symptomatic Oyster Harvesters', *Journal of Food Protection*, 75(9), pp. 1715–1720. doi: 10.4315/0362-028x.jfp-12-113.

McLeod, C. et al. (2017) 'Depuration and Relaying: A Review on Potential Removal of Norovirus

- from Oysters', *Comprehensive Reviews in Food Science and Food Safety*, 16(4), pp. 692–706. doi: 10.1111/1541-4337.12271.
- Menzel, P., Ng, K. L. and Krogh, A. (2016) 'Fast and sensitive taxonomic classification for metagenomics with Kaiju', *Nature Communications*. Nature Publishing Group, 7, pp. 1–9. doi: 10.1038/ncomms11257.
- Metcalf, T. G. (1995) 'Environmental Virology: From Detection of Virus in Sewage and Water by Isolation to Identification by Molecular Biology--A Trip of Over 50 Years', *Annual Review of Microbiology*, 49(1), pp. 461–487. doi: 10.1146/annurev.micro.49.1.461.
- Metzker, M. L. (2010) 'Sequencing technologies the next generation', *Nature Reviews Genetics*. Nature Publishing Group, 11(1), pp. 31–46. doi: 10.1038/nrg2626.
- Miclotte, G. et al. (2016) 'Jabba: Hybrid error correction for long sequencing reads', *Algorithms for Molecular Biology*. BioMed Central, 11(1), pp. 1–12. doi: 10.1186/s13015-016-0075-7.
- Miura, T. et al. (2013) 'Histo-Blood Group Antigen-Like Substances of Human Enteric Bacteria as Specific Adsorbents for Human Noroviruses', *Journal of Virology*, 87(17), pp. 9441–9451. doi: 10.1128/JVI.01060-13.
- Montazeri, N. et al. (2015) 'Pathogenic Enteric Viruses and Microbial Indicators during Secondary Treatment of Municipal Wastewater', *Applied and Environmental Microbiology*, 81(18), pp. 6436–6445. doi: 10.1128/aem.01218-15.
- Monteiro, S., & Santos, R. (2017). Nanofluidic digital PCR for the quantification of Norovirus for water quality assessment. *PLoS ONE*, 12(7), pp. 1–12. <https://doi.org/10.1371/journal.pone.0179985>
- Moore, M. D. et al. (2015) 'Generation and characterization of nucleic acid aptamers targeting the capsid P domain of a human norovirus GII.4 strain', *Journal of Biotechnology*. Elsevier B.V., 209, pp. 41–49. doi: 10.1016/j.jbiotec.2015.06.389.
- Moore, M. D. et al. (2016) 'Human Norovirus Aptamer Exhibits High Degree of Target Conformation- Dependent Binding Similar to That of Receptors and Discriminates Particle Functionality', 1(6), pp. 1–15. doi: 10.1128/mSphere.00298-16.Editor.
- Mori, K. et al. (2017) 'Comparison of genetic characteristics in the evolution of Norovirus GII.4 and GII.17', *Journal of Medical Virology*, 89(8), pp. 1480–1484. doi: 10.1002/jmv.24791.
- Morillo, S. G. and Timenetsky, M. do C. S. T. T. (2011) 'Norovirus: An overview', *Revista da Associacao Medica Brasileira*, 57(4), pp. 453–458. doi: 10.1016/S0104-4230(11)70094-X.
- Mukherjee, S. et al. (2015) 'Large-scale contamination of microbial isolate genomes by Illumina PhiX control', *Standards in Genomic Sciences*, 10(1), pp. 1–4. doi: 10.1186/1944-3277-10-18.
- Mullendore, J. L., Sobsey, M. D. and Carol Shieh, Y. S. (2001) 'Improved method for the recovery of hepatitis A virus from oysters', *Journal of Virological Methods*, 94(1–2), pp. 25–35. doi: 10.1016/S0166-0934(01)00263-4.
- Nappier, S. P., Graczyk, T. K. and Schwab, K. J. (2008) 'Bioaccumulation, retention, and depuration of enteric viruses by *Crassostrea virginica* and *Crassostrea ariakensis* oysters', *Applied and Environmental Microbiology*, 74(22), pp. 6825–6831. doi: 10.1128/AEM.01000-08.
- Nasheri, N. et al. (2017) 'Characterization of the Genomic Diversity of Norovirus in Linked Patients Using a Metagenomic Deep Sequencing Approach', 8(January), pp. 1–14. doi: 10.3389/fmicb.2017.00073.

- Nieuwenhuijse, D. F. and Koopmans, M. P. G. (2017) 'Metagenomic sequencing for surveillance of food- and waterborne viral diseases', *Frontiers in Microbiology*, 8(FEB), pp. 1–11. doi: 10.3389/fmicb.2017.00230.
- Nooij, S. et al. (2018) 'Overview of virus metagenomic classification methods and their biological applications', *Frontiers in Microbiology*, 9(APR). doi: 10.3389/fmicb.2018.00749.
- Oka, T. et al. (2015) 'Comprehensive review of human sapoviruses', *Clinical Microbiology Reviews*, 28(1), pp. 32–53. doi: 10.1128/CMR.00011-14.
- Okoh, A. I., Sibanda, T. and Gusha, S. S. (2010) 'Inadequately treated wastewater as a source of human enteric viruses in the environment', *International Journal of Environmental Research and Public Health*, 7(6), pp. 2620–2637. doi: 10.3390/ijerph7062620.
- Oristo, S., Lee, H. J. and Maunula, L. (2018) 'Performance of pre-RT-qPCR treatments to discriminate infectious human rotaviruses and noroviruses from heat-inactivated viruses: applications of PMA/PMAxx, benzonase and RNase', *Journal of Applied Microbiology*, 124(4), pp. 1008–1016. doi: 10.1111/jam.13737.
- Oshiki, M. et al. (2018) 'Microfluidic PCR Amplification and MiSeq Amplicon Sequencing Techniques for High-Throughput Detection and Genotyping of Human Pathogenic RNA Viruses in Human Feces , Sewage , and Oysters', 9(April), pp. 1–10. doi: 10.3389/fmicb.2018.00830.
- Osunmakinde, C. O. et al. (2018) 'Overview of trends in the application of metagenomic techniques in the analysis of human enteric viral diversity in Africa's environmental regimes', *Viruses*, 10(8). doi: 10.3390/v10080429.
- Ozsolak, F. et al. (2009) 'Direct RNA sequencing', *Nature*. Nature Publishing Group, 461(7265), pp. 814–818. doi: 10.1038/nature08390.
- Parnaudeau, P. S. et al. (2014) 'Virus entériques humains et coquillages', *Bulletin de l'Academie Veterinaire de France*, 167(3), pp. 227–234. doi: 10.4267/2042/54194.
- Parra, G. I. et al. (2017) 'Static and Evolving Norovirus Genotypes: Implications for Epidemiology and Immunity', *PLoS Pathogens*, 13(1), pp. 1–22. doi: 10.1371/journal.ppat.1006136.
- Persson, S., Eriksson, R., Lowther, J., Ellström, P., & Simonsson, M. (2018). Comparison between RT droplet digital PCR and RT real-time PCR for quantification of noroviruses in oysters. *International Journal of Food Microbiology*, 284(July), pp. 73–83. <https://doi.org/10.1016/j.ijfoodmicro.2018.06.022>
- Petterson, S., Grøndahl-Rosado, R., Nilsen, V., Myrmel, M., & Robertson, L. J. (2015). Variability in the recovery of a virus concentration procedure in water: Implications for QMRA. *Water Research*, 87(April 2016), pp. 79–86.
- Piquet Jean-Come (2018). Procédure nationale de la surveillance sanitaire microbiologique des zones de production de coquillages. Prescriptions du réseau de surveillance microbiologique des zones de production (REMI). <https://archimer.ifremer.fr/doc/00461/57260/>
- Polo D, Varela MF, Romalde JL. (2015). Detection and quantification of hepatitis A virus and norovirus in Spanish authorized shellfish harvesting areas. *Int. J. Food Microbiol.* 193, pp. 43-50.
- Polo, D. et al. (2014) 'Depuration kinetics of hepatitis A virus in clams', *Food Microbiology*. Elsevier Ltd, 39, pp. 103–107. doi: 10.1016/j.fm.2013.11.011.
- Polo, D. et al. (2016) 'Digital PCR for quantifying norovirus in oysters implicated in outbreaks, France', *Emerging Infectious Diseases*, 22(12), pp. 2189–2191. doi: 10.3201/eid2212.160841.

- Polo, D. et al. (2018) 'Infectivity and RNA persistence of a norovirus surrogate, the Tulane virus, in oysters', *Frontiers in Microbiology*, 9(APR), pp. 1–8. doi: 10.3389/fmicb.2018.00716.
- Pompey, M. et al. (2005) 'Fecal contamination in coastal areas: An engineering approach', *Oceans and Health: Pathogens in the Marine Environment*, pp. 331–359.
- Ponty Y. (2014). Bio-algorithmique des ARN : petite promenade aux interfaces. Eric Sopena. 1024 – Bulletin de la société informatique de France, 4, SIF, pp.23-53, 2014, <http://www.societe-informatique-de-France.fr/bulletin/1024-numero-4/>. <hal-01077506v3>
- Poptsova, M. S. et al. (2014) 'Non-random DNA fragmentation in next-generation sequencing', *Scientific Reports*, 4, pp. 1–6. doi: 10.1038/srep04532.
- Prieur, A. (2008) 'Les coquillages du Paléolithique à l'âge du Bronze au Moyen-Orient et en Méditerranée orientale : interprétations environnementales et utilisation humaine', *Paléorient*, 31(1), pp. 158–168. doi: 10.3406/paleo.2005.4794.
- Rasmussen, L. D. et al. (2016) 'Molecular evidence of oysters as vehicle of norovirus GII.P17-GII.17', *Emerging Infectious Diseases*, 22(11), pp. 2024–2025. doi: 10.3201/eid2211.161171.
- Quail, M. A. et al. (2012) 'A tale of 3 NGS sequencing platforms.pdf', *BMC Genomics* 2012, 13(341), p. 13. doi: 10.1186/1471-2164-13-341.
- Quick, J. et al. (2016) 'Real-time, portable genome sequencing for Ebola surveillance', *Nature*. 2016/02/03, 530(7589), pp. 228–232. doi: 10.1038/nature16996.
- Quijada, N. M. et al. (2016) 'Propidium monoazide integrated with qPCR enables the detection and enumeration of infectious enteric RNA and DNA viruses in clam and fermented sausages', *Frontiers in Microbiology*, 7(DEC), pp. 1–6. doi: 10.3389/fmicb.2016.02008.
- Randazzo, W. et al. (2018) 'Optimization of PMAxx pretreatment to distinguish between human norovirus with intact and altered capsids in shellfish and sewage samples', *International Journal of Food Microbiology*. Elsevier, 266(August 2017), pp. 1–7. doi: 10.1016/j.ijfoodmicro.2017.11.011.
- Reed, L. J., and H. A. Muench. (1938). A simple method of estimating fifty percent endpoints. *Am. J. Hyg. (London)* 27, pp. 493–497.
- Rincé A, Balière Ch, Hervio-Heath D, Cozien J, Lozach S, Parnaudeau S, Le Guyader FS et al. (2018). Occurrence of bacterial pathogens and human noroviruses in shellfish- harvesting areas and their catchments in France. *Front. Microbiol.* 9: 2443.
- Rincé, A., Balière, C., Hervio-Heath, D., Cozien, J., Lozach, S., Parnaudeau, S., ... Gourmelon, M. (2018). Occurrence of bacterial pathogens and human noroviruses in shellfish- Harvesting areas and their catchments in France. *Frontiers in Microbiology*, 9(OCT). <https://doi.org/10.3389/fmicb.2018.02443>
- Robilotti, E., Deresinski, S. and Pinsky, B. A. (2015) 'Norovirus', *Clinical Microbiology Reviews*, 28(1), pp. 134–164. doi: 10.1128/CMR.00075-14.
- Rocha-Pereira, J., Van Dycke, J. and Neyts, J. (2016) 'Norovirus genetic diversity and evolution: implications for antiviral therapy', *Current Opinion in Virology*. Elsevier B.V., 20, pp. 92–98. doi: 10.1016/j.coviro.2016.09.009.
- Rogers, J.D., Ajami, N.J., Fryszczyn, B.G., Estes, M.K., Atmar, R.L., Palzkill, T., (2013). Identification and Characterization of a Peptide Affinity Reagent for Detection of Noroviruses in Clinical Samples. *Journal of Clinical Microbiology* 51, 1803-1808.

- Ronholm, J. et al. (2016) 'Navigating microbiological food safety in the era of whole-genome sequencing', *Clinical Microbiology Reviews*, 29(4), pp. 837–857. doi: 10.1128/CMR.00056-16.
- Rose, R. et al. (2016) 'Challenges in the analysis of viral metagenomes', *Virus Evolution*, 2(2), p. vew022. doi: 10.1093/ve/vew022.
- Rougemont, A. De et al. (2010) 'Actualités sur Les Norovirus', 19(Figure 1).
- Rupnik A, Keaveney S, Devilly L, Butler F, Doré W. (2018). The impact of winter relocation and depuration on norovirus concentrations in pacific oysters harvested from a commercial production site. *Food Env. Virol.* 10, pp. 288-296.
- Ruvoën-Clouet, N., Belliot, G. and Le Pendu, J. (2013) 'Sensibilité génétique aux infections à norovirus et protection innée de population', *Virologie*, 17(4), pp. 264–277. doi: 10.1684/vir.2013.0509.
- Ruvoën, N. and Le Pendu, J. (2013) '[Genetic susceptibility to norovirus infection].', *Pathologie-biologie*, 61(1), pp. 28–35. doi: 10.1016/j.patbio.2013.01.002.
- Sabrià, A. et al. (2018) 'Characterization of intra- and inter-host norovirus P2 genetic variability in linked individuals by amplicon sequencing', *PLoS ONE*, 13(8), pp. 1–14. doi: 10.1371/journal.pone.0201850.
- Sadeuh-Mba, S. A. et al. (2019) 'Genetic landscape and macro-evolution of cocirculating Coxsackieviruses A and Vaccinaderived Polioviruses in the Democratic Republic of Congo, 2008–2013', *PLoS Neglected Tropical Diseases*, 13(4), pp. 2008–2013. doi: 10.1371/journal.pntd.0007335.
- Sala C., et al. (Submitted). 'Comparative analysis of shotgun metagenomic sequencing results obtained for spiked smoked salmon using different wet lab protocols and data analysis tools'.
- Salladini, A., Prisciandaro, M. and Barba, D. (2007) 'Ultrafiltration of biologically treated wastewater by using backflushing', *Desalination*, 207(1–3), pp. 24–34. doi: 10.1016/j.desal.2006.02.078.
- Sangiovanni, M. et al. (2019) 'From trash to treasure: Detecting unexpected contamination in unmapped NGS data', *BMC Bioinformatics*. *BMC Bioinformatics*, 20(Suppl 4), pp. 1–12. doi: 10.1186/s12859-019-2684-x.
- Sano, D. et al. (2016) 'Risk management of viral infectious diseases in wastewater reclamation and reuse: Review', *Environment International*. Elsevier Ltd, 91, pp. 220–229. doi: 10.1016/j.envint.2016.03.001.
- Santiago-Rodriguez, T. M. and Hollister, E. B. (2019) 'Human Virome and Disease: High-Throughput Sequencing for Virus Discovery, Identification of Phage-Bacteria Dysbiosis and Development of Therapeutic Approaches with Emphasis on the Human Gut', *Viruses*, 11(656). doi: 10.3390/v11070656.
- Santiana, M. et al. (2018) 'Vesicle-Cloaked Virus Clusters Are Optimal Units for Inter-organismal Viral Transmission', *Cell Host & Microbe*, 24, pp. 208–220. Available at: <https://doi.org/10.1016/j.chom.2018.07.006>.
- Sato, T. and Clevers, H. (2013) 'Growing self-organizing mini-guts from a single intestinal stem cell: Mechanism and applications', *Science*, 340(6137), pp. 1190–1194. doi: 10.1126/science.1234852.
- Schadt, E. E., Turner, S. and Kasarskis, A. (2010) 'A window into third-generation sequencing', *Human Molecular Genetics*, 19(R2), pp. 227–240. doi: 10.1093/hmg/ddq416.

- Schaeffer, J. et al. (2018) 'Improving the efficacy of sewage treatment decreases norovirus contamination in oysters', *International Journal of Food Microbiology*. Elsevier, 286(July), pp. 1–5. doi: 10.1016/j.ijfoodmicro.2018.07.016.
- Scheuch, M., Höper, D. and Beer, M. (2015) 'RIEMS: A software pipeline for sensitive and comprehensive taxonomic classification of reads from metagenomics datasets', *BMC Bioinformatics*, 16(1). doi: 10.1186/s12859-015-0503-6.
- Schirmer, M. et al. (2015) 'Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform', *Nucleic Acids Research*, 43(6). doi: 10.1093/nar/gku1341.
- Schirmer, M. et al. (2016) 'Illumina error profiles: Resolving fine-scale variation in metagenomic sequencing data', *BMC Bioinformatics*. *BMC Bioinformatics*, 17(1), pp. 1–15. doi: 10.1186/s12859-016-0976-y.
- Schomig, V.J., Kasdorf, B.T., Scholz, C., Bidmon, K., Lieleg, O., Berensmeier, S. (2016). An optimized purification process for porcine gastric mucin with preservation of its native functional properties. *RSC Advances* 6, 44932-44943.
- Sedji, M. I. et al. (2018) 'Quantification of human adenovirus and norovirus in river water in the north-east of France', *Environmental Science and Pollution Research*. *Environmental Science and Pollution Research*, 25(30), pp. 30497–30507. doi: 10.1007/s11356-018-3045-4.
- Shendure, J. and Ji, H. (2008) 'Next-generation DNA sequencing', *Nature Biotechnology*, 26(10), pp. 1135–1145. doi: 10.1038/nbt1486.
- Shi, H., Pasco, E. V., & Tarabara, V. V. (2017). Membrane-based methods of virus concentration from water: A review of process parameters and their effects on virus recovery. *Environmental Science: Water Research and Technology*, 3(5), pp. 778–792. <https://doi.org/10.1039/c7ew00016b>
- Sidhu, S.S., Fairbrother, W.J., Deshayes, K. (2003). Exploring Protein–Protein Interactions with Phage Display. *ChemBioChem* 4, pp. 14-25.
- Sima, L. C. et al. (2011) 'Calicivirus removal in a membrane bioreactor wastewater treatment plant', *Applied and Environmental Microbiology*, 77(15), pp. 5170–5177. doi: 10.1128/AEM.00583-11.
- Simmons, K. et al. (2013) 'Duration of immunity to norovirus gastroenteritis', *Emerging Infectious Diseases*, 19(8), pp. 1260–1267. doi: 10.3201/eid1908.130472.
- Simon-Loriere, E. and Holmes, E. C. (2011) 'Why do RNA viruses recombine?', *Nature Reviews Microbiology*. Nature Publishing Group, 9(8), pp. 617–626. doi: 10.1038/nrmicro2614.
- Sims, D. et al. (2014) 'Sequencing depth and coverage: Key considerations in genomic analyses', *Nature Reviews Genetics*. Nature Publishing Group, 15(2), pp. 121–132. doi: 10.1038/nrg3642.
- Steinegger, M., Mirdita, M. and Söding, J. (2019) 'Protein-level assembly increases protein sequence recovery from metagenomic samples manyfold', *Nature Methods*, 16(7), pp. 603–606. doi: 10.1038/s41592-019-0437-4.
- Strazzulli, A. et al. (2017) 'Metagenomics of microbial and viral life in terrestrial geothermal environments', *Reviews in Environmental Science and Bio/Technology*. Springer Netherlands, 16(3), pp. 425–454. doi: 10.1007/s11157-017-9435-0.
- Sutton, T. D. S. et al. (2019) 'Choice of assembly software has a critical impact on virome characterisation', *Microbiome*. *Microbiome*, 7(1), pp. 1–15. doi: 10.1186/s40168-019-0626-5.

- Tan DM, Lyu SL, Liu W. et al. (2018). Utility of droplet digital PCR assay for quantitative detection of norovirus in shellfish, from production to consumption in Guangxi, China. *Biomed. Environ. Sci.* 31: 713-720.
- Tan, M. and Jiang, X. (2014) 'Histo-blood group antigens: A common niche for norovirus and rotavirus', *Journal of the Australian Mathematical Society*, 16(2). doi: 10.1017/erm.2014.2.
- Teunis, P. F. M. et al. (2008) 'Norwalk Virus: How Infectious is It?', *Journal of Medical Virology*, 80, pp. 1468–1476. doi: 10.1002/jmv.
- Teunis, P. F. M. et al. (2015) 'Shedding of norovirus in symptomatic and asymptomatic infections', *Epidemiology and Infection*, 143(8), pp. 1710–1717. doi: 10.1017/S095026881400274X.
- Thebault, P. A. et al. (2014) 'Modélisation dynamique théorique de la transmission de norovirus par la consommation d'huîtres contaminées et par transmission interhumaine', *Bulletin de l'Académie Vétérinaire de France*, 167(3), pp. 235–240. doi: 10.4267/2042/54195.
- Thorne, L. G. and Goodfellow, I. G. (2014) 'Norovirus gene expression and replication', *Journal of General Virology*, 95(PART 2), pp. 278–291. doi: 10.1099/vir.0.059634-0.
- Tian, P., Brandl, M., Mandrell, R. (2005). Porcine gastric mucin binds to recombinant norovirus particles and competitively inhibits their binding to histo-blood group antigens and Caco-2 cells. *Lett Appl Microbiol* 41, pp. 315-320.
- Tian, P., Jiang, X., Zhong, W., Jensen, H.M., Brandl, M., Bates, A.H., Engelbrektsen, A.L., Mandrell, R. (2007). Binding of recombinant norovirus like particle to histo-blood group antigen on cells in the lumen of pig duodenum. *Res Vet Sci* 83, 410-418.
- Tian, P., Yang, D., Mandrell, R. (2011). A simple method to recover Norovirus from fresh produce with large sample size by using histo-blood group antigen-conjugated to magnetic beads in a recirculating affinity magnetic separation system (RCAMS). *Int J Food Microbiol* 147, pp. 223-227.
- Tian, P., Yang, D., Pan, L., Mandrell, R. (2012). Application of a receptor-binding capture quantitative reverse transcription-PCR assay to concentrate human norovirus from sewage and to study the distribution and stability of the virus. *Appl Environ Microbiol* 78, pp. 429-436.
- Torner, N. et al. (2016) 'Epidemiology of acute gastroenteritis outbreaks caused by human calicivirus (Norovirus and Sapovirus) in Catalonia: A two year prospective study, 2010-2011', *PLoS ONE*, 11(4). doi: 10.1371/journal.pone.0152503.
- Torok V, Hodgson K, McLeod C, Tan J, Malhi N, Turnbull A. (2018). National survey of foodborne viruses in Australian oysters at production. *Food Microbiol.* 69, pp. 196-203.
- Troeger, H. et al. (2009) 'Structural and functional changes of the duodenum in human norovirus infection', *Gut*, 58(8), pp. 1070–1077. doi: 10.1136/gut.2008.160150.
- van Dijk, E. L. et al. (2018) 'The Third Revolution in Sequencing Technology', *Trends in Genetics*. Elsevier Ltd, 34(9), pp. 666–681. doi: 10.1016/j.tig.2018.05.008.
- Venkataram Prasad, B. V., Hardy, M. E. and Estes, M. K. (2000) 'Structural Studies of Recombinant Norwalk Capsids', *The Journal of Infectious Diseases*, 181(s2), pp. S317–S321. doi: 10.1086/315576.
- Verhoef, L. et al. (2015) 'Norovirus genotype profiles associated with foodborne transmission, 1999–2012', *Emerging Infectious Diseases*, 21(4), pp. 592–599. doi: 10.3201/eid2104.141073.
- Vibin, J. et al. (2018) 'Metagenomics detection and characterisation of viruses in faecal samples

- from Australian wild birds', *Scientific Reports*, 8(1), pp. 1–23. doi: 10.1038/s41598-018-26851-1.
- Vincent-Hubert, F. et al. (2017) 'Adsorption of norovirus and ostreid herpesvirus type 1 to polymer membranes for the development of passive samplers', *Journal of Applied Microbiology*, 122(4), pp. 1039–1047. doi: 10.1111/jam.13394.
- Vinje, J. 0007335. pd. (2014) 'Advances in Laboratory Methods for Detection and Typing of Norovirus.', *Journal of clinical microbiology*, 53(July), pp. 373–381. doi: 10.1128/JCM.01535-14.
- Vongpunsawad, S., Venkataram Prasad, B. V., & Estes, M. K. (2013). Norwalk Virus Minor Capsid Protein VP2 Associates within the VP1 Shell Domain. *Journal of Virology*, 87(9), pp. 4818–4825. <https://doi.org/10.1128/JVI.03508-12>
- Wang, J. and Deng, Z. (2016) 'Modeling and prediction of oyster norovirus outbreaks along Gulf of Mexico Coast', *Environmental Health Perspectives*, 124(5), pp. 627–633. doi: 10.1289/ehp.1509764.
- Warner, M. L., Bell, T. and Pioszak, A. A. (2015) 'Engineering high-potency r-spondin adult stem cell growth factors', *Molecular Pharmacology*, 87(3), pp. 410–420. doi: 10.1124/mol.114.095133.
- Wells, J. M. and Spence, J. R. (2014) 'How to make an intestine', *Development (Cambridge)*, 141(4), pp. 752–760. doi: 10.1242/dev.097386.
- White, D. J., Wang, J. and Hall, R. J. (2017) 'Assessing the Impact of Assemblers on Virus Detection in a De Novo Metagenomic Analysis Pipeline', *Journal of Computational Biology*, 24(9), pp. 874–881. doi: 10.1089/cmb.2017.0008.
- White, P. A. (2014) 'Evolution of norovirus', *Clinical Microbiology and Infection. European Society of Clinical Infectious Diseases*, 20(8), pp. 741–745. doi: 10.1111/1469-0691.12746.
- Willert, K. et al. (2003) 'Wnt proteins are lipid-modified and can act as stem cell growth factors', *Nature*, 423(6938), pp. 448–452. doi: 10.1038/nature01611.
- Workman, M. J. et al. (2017) 'Engineered human pluripotent-stem-cell-derived intestinal tissues with a functional enteric nervous system', *Nature Medicine. Nature Publishing Group*, 23(1), pp. 49–59. doi: 10.1038/nm.4233.
- World Health Organization. (2003). Guidelines for environmental surveillance of poliovirus circulation. <https://apps.who.int/iris/handle/10665/67854>
- Wylie, T. N. et al. (2015) 'Enhanced Virome Sequencing Using Targeted Sequence Capture Enhanced virome sequencing using sequence capture', *Cold Spring Harbor Laboratory Press on June, 2*, pp. 1910–1920. doi: 10.1101/gr.191049.115.Freely.
- Wylie, T. N. et al. (2018) 'Detection of Viruses in Clinical Samples by Use of Metagenomic Sequencing and Targeted Sequence Capture', *Journal of Clinical Microbiology*, 56(12), pp. 1–10. doi: 10.1128/jcm.01123-18.
- Yin, X. et al. (2016) 'Engineering Stem Cell Organoids', *Cell Stem Cell. Elsevier Inc.*, 18(1), pp. 25–38. doi: 10.1016/j.stem.2015.12.005.
- Yin, Y. and Zhou, D. (2018) 'Organoid and Enteroid Modeling of Salmonella Infection', *Frontiers in cellular and infection microbiology*, 8(April), p. 257. doi: 10.3389/fcimb.2018.00257.
- Yu, Y. et al. (2015) 'Molecular Epidemiology of Oyster-Related Human Noroviruses and Their Global Genetic Diversity and Temporal-Geographical Distribution from 1983 to 2014', *Applied and Environmental Microbiology*, 81(21), pp. 7615–7624. doi: 10.1128/aem.01729-15.

Zachos, N. C. et al. (2016) 'Human enteroids/colonoids and intestinal organoids functionally recapitulate normal intestinal physiology and pathophysiology', *Journal of Biological Chemistry*, 291(8), pp. 3759–3766. doi: 10.1074/jbc.R114.635995.

Zakhour, M. et al. (2010) 'Bovine norovirus: Carbohydrate ligand, environmental contamination, and potential cross-species transmission via oysters', *Applied and Environmental Microbiology*, 76(19), pp. 6404–6411. doi: 10.1128/AEM.00671-10.

Zhang, B. et al. (2015) 'Comparison of droplet digital PCR and conventional quantitative PCR for measuring EGFR gene mutation', *Experimental and Therapeutic Medicine*, 9(4), pp. 1383–1388. doi: 10.3892/etm.2015.2221.

Zhou, Z. et al. (2017) 'In situ capture RT-qPCR: A new simple and sensitive method to detect human norovirus in oysters', *Frontiers in Microbiology*, 8(APR), pp. 1–8. doi: 10.3389/fmicb.2017.00554.

Zhu, S. et al. (2013) 'Identification of Immune and Viral Correlates of Norovirus Protective Immunity through Comparative Study of Intra-Cluster Norovirus Strains', 9(9). doi: 10.1371/journal.ppat.1003592.

LIENS INTERNET

<http://www.ostreiculteurtraditionnel.fr/histoire/histoire.html>

https://fr.wikipedia.org/wiki/Score_de_qualit%C3%A9_phred

<https://github.com/Ecological-and-Evolutionary-Genomics/eeg2016/wiki/Mar-21-Exercise-7---->

[SPAdes-assembler](#)

<https://international.neb.com/-/media/catalog/datacards-or-manuals/s1330datasheet-lot0041208.pdf?rev=33bbe02bd0684b90983f259786047360>

<https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbnorm-guide/>

<https://research.a-star.edu.sg/articles/highlights/expanding-the-power-of-pores/>

<https://www.genomedetective.com/app/typingtool/virus/introduction>

<https://www.medgadget.com/2019/08/smartphone-and-paper-microfluidic-device-for-norovirus-detection.html>

<https://www.ncbi.nlm.nih.gov/genbank/>

<https://www.ncbi.nlm.nih.gov/refseq/>

<https://www.stemcell.com/technical-resources/area-of-interest/organoid-research/intestinal-research/overview.html>

ANNEXE I

Résumé et caractéristiques des plateformes HTS des différents générations (Adapté de Kchouk, Gibrat and Elloumi, 2017)

Plateforme	Technologie	Reads par run	Taille des reads (moyenne)	Type de reads	Type d'erreur	Taux d'erreur	Données générées par run (Gb)	Année
Première génération								
ABI Sanger	3730xl	96	650	SE	NA	0.3	0.00069 à 0.0021	2002
Deuxième génération								
454	GS20	200	100	SE, PE	Indel	1	0.02	2005
454	GS FLX	400	250	SE, PE	Indel	1	0.1	2007
454	GS FLX Titanium	1 M	450	SE, PE	Indel	1	0.45	2009
454	GS Junior ⁺	100	700	SE, PE	Indel	1	0.07	2014
Illumina	MiniSeq	25M (max)	150	SE, PE	Mismatch	1	7.5 (max)	2013
Illumina	MiSeq	25M (max)	150 ou 300	SE, PE	Mismatch	0.1	15 (max)	2011
Illumina	NextSeq	400M (max)	150	SE, PE	Mismatch	1	120 (max)	2014
Illumina	HiSeq	5B (max)	150	SE, PE	Mismatch	0.1	1.5 Tb (max)	2012
Illumina	HiSeq X	6B (max)	150	SE, PE	Mismatch	0.1	1.8 Tb (max)	2014
SOLiD	5500 W	3B	75	SE	Mismatch	~0.1	160	2011
SOLiD	5500xl W	6B	75	SE	Mismatch	~0.1	320	2013
Ion Torrent	PGM 314 chip v2	400.000-550.000	400	SE	Indel	1	0.06 à 0.1	2011
Ion Torrent	PGM 318 chip v2	4M – 5.5M	400	SE	Indel	1	1.2 à 2	2013
Ion Torrent	Ion Proton	60M - 80M	200	SE	Indel	1	10	2012
Ion Torrent	Ion S5/S5XL 520	3M – 5M	400	SE	Indel	1	1.2 à 2	2015

Ion Torrent	Ion S5/S5XL 530	15M – 20M	400	SE	Indel	1	0.3 à 0.5	2015
Ion Torrent	Ion S5/S5XL 540	60M – 80M	400	SE	Indel	1	ND	2015
Troisième génération								
PacBio	RS C1	432	1300	SE	Indel	15	0.5	2011
PacBio	RS C2	432	2500	SE	Indel	15	0.5 à 1	2012
PacBio	RS C2 XL	432	4300	SE	Indel	15	0.5 à 1	2012
PacBio	RS II C2 XL	564	4600	SE	Indel	15	0.5 à 1	2013
PacBio	RS II P5 C3	528	8500	SE	ND	ND	7	2016
PacBio	RS II P6 C4	660	13500	SE	Indel	12	0.5 à 1	2014
PacBio	Sequel	350	10000	SE	Indel	12	0.5 à 1	2014
Oxford Nanopore	MiniION Mk	100	9545	1D, 2D	Indel/mismatch	12	1.5	2015
Oxford Nanopore	PromethION	ND	9846	1D, 2D	ND	ND	2Tb à 4Tb	2016
ND : Non disponible ; SE : Single End ; PE : Paired End ; M : Million ; B : Billion ; GB : Gigabytes ; Tb : Terabytes								

ANNEXE II

Souches de NoV testées en culture cellulaire

Genotype_variante	P-type*	Réplication	Moyenne RNA copies/ μ L	Taux de réplication	Reference
GI.1	GI.P1	???	0.5- 19.3 x 10 ³	???	Costantini <i>et al</i> , 2018
GI.3	GI.P3	???	0.3-59.4 x 10 ⁴	???	Costantini <i>et al</i> , 2018
GI.3	GI.Pd	???	1.2-8.7 x 10 ⁴	???	Costantini <i>et al</i> , 2018
GI.4	GI.P4	???	2.2 x 10 ²	???	Costantini <i>et al</i> , 2018
GI.7	GI.P7	???	2.6-16.4 x 10 ³	???	Costantini <i>et al</i> , 2018
GII.1	GII.Pg	oui	1.4 x 10 ⁵	696	Costantini <i>et al</i> , 2018
GII.2	GII.P16	oui	0.2-52.6 x 10 ³	56	Costantini <i>et al</i> , 2018
GII.3	GII.P21	???	1.0-6.4 x 10 ⁵	???	Costantini <i>et al</i> , 2018
GII.3	GII.P12	oui	1.6-4.4 x 10 ⁶	181	Costantini <i>et al</i> , 2018
GII.3	GII.P16	???	0.2-141.1 x 10 ⁴	???	Costantini <i>et al</i> , 2018
GII.4_DenHaag	GII.P4	oui	1.3-161.6 x 10 ⁴	155	Costantini <i>et al</i> , 2018
GII.4_NewOrleans	GII.P4	oui	4.1 x 10 ⁵	227	Costantini <i>et al</i> , 2018
GII.4_Sydney	GII.Pe	oui	3.5 x 10 ³ – 2.1 x 10 ⁷	121-1236	Costantini <i>et al</i> , 2018
GII.4_Sydney	GII.P16	oui	0.4 x 10 ³ – 6.1 x 10 ⁶	340	Costantini <i>et al</i> , 2018
GII.4_Sydney	GII.P4	???	0.6-14.9 x 10 ⁴	???	Costantini <i>et al</i> , 2018
GII.5	GII.P22	???	1.1 x 10 ⁴	???	Costantini <i>et al</i> , 2018
GII.6B	GII.P7	???	7.6 x 10 ⁴	???	Costantini <i>et al</i> , 2018
GII.6	GII.P7	???	0.1-8.4 x 10 ⁶	???	Costantini <i>et al</i> , 2018
GII.7	GII.P7	???	0.3-9.1 x 10 ⁶	???	Costantini <i>et al</i> , 2018
GII.14	GII.P7	oui	6.1 x 10 ⁴	47	Costantini <i>et al</i> , 2018
GII.17	GII.Pe	oui	3.4 x 10 ⁵	96	Costantini <i>et al</i> , 2018
GIV	GIV.P1	???	0.3-13.6 x 10 ³	???	Costantini <i>et al</i> , 2018
GI.1	GI.P1	oui	5.7 x 10 ⁶	ND	Ettayebi <i>et al</i> , 2016
GII.3	GII.P21	oui	8.5 x 10 ⁶	10-173	Ettayebi <i>et al</i> , 2016
GII.4/2006a	GII.P4	oui	7.0 x 10 ⁷	34-6730	Ettayebi <i>et al</i> , 2016
GII.4/2006b-1	GII.P4	oui	1.5 x 10 ⁷	34-6730	Ettayebi <i>et al</i> , 2016
GII.4/2006b-2	GII.P4	oui	5.3 x 10 ⁶	34-6730	Ettayebi <i>et al</i> , 2016
GII.4/2006b-3	GII.P4	oui	1.1 x 10 ⁷	34-6730	Ettayebi <i>et al</i> , 2016
GII.4/2009	GII.P4	oui	3.0 x 10 ⁷	34-6730	Ettayebi <i>et al</i> , 2016
GII.4/2012-1	GII.Pe	oui	1.8 x 10 ⁸	34-6730	Ettayebi <i>et al</i> , 2016
GII.4/2012-2	GII.Pe	oui	3.0 x 10 ⁶	34-6730	Ettayebi <i>et al</i> , 2016
GII.17	GII.Pn	???	1.4 x 10 ⁷	ND	Ettayebi <i>et al</i> , 2016

*Polymérase ; ND : non disponible

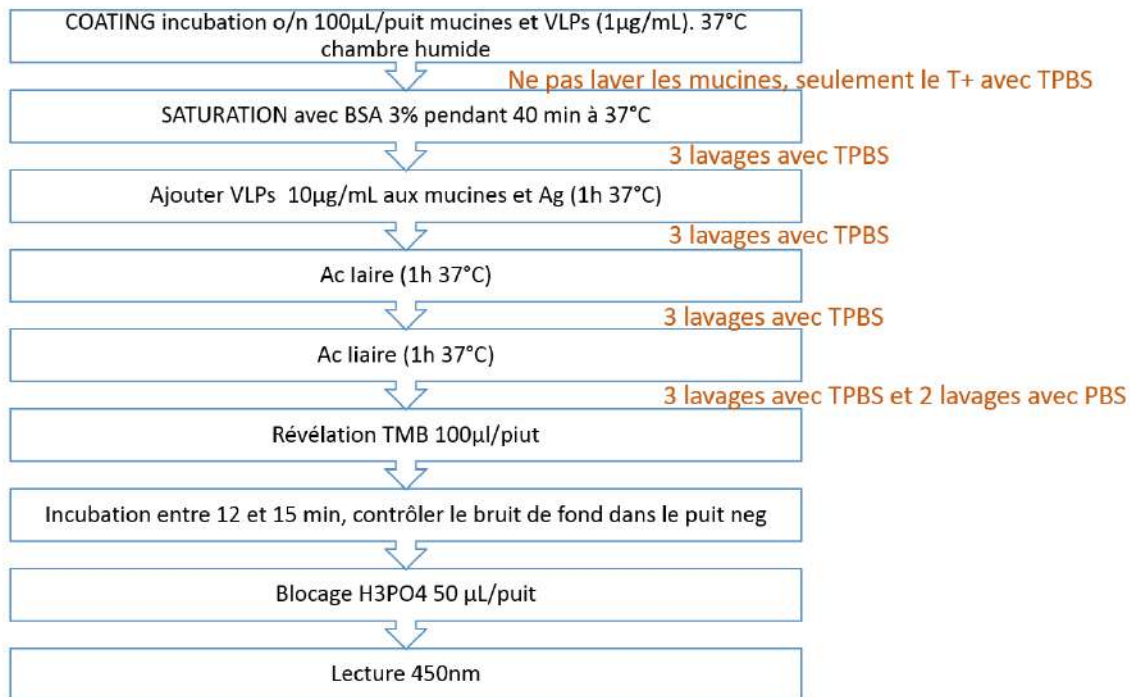
ANNEXE III

Souches NoV prévenantes da selles purifiées et utilisées pour ISC-PCR

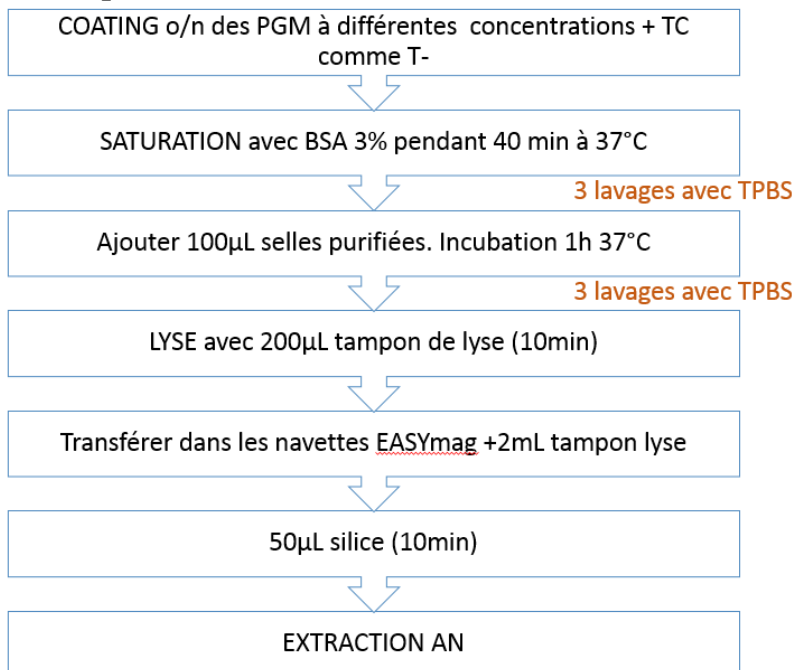
Selles	ORF2	ORF1	Ct approx	Année	Cg/mL (30/10/17)
s55	GI.3	GI.P3	30,6	2003	$6,36 \times 10^4$
s318	GI.2	GI.P2	30,8	2009	$5,52 \times 10^4$
s382	GI.6	GI.Pb	33,4	2008	$9,58 \times 10^3$
s383	GI.7	X	32,2	2008	$4,41 \times 10^4$
s505	GI.4	GI.P4	25,4	2012	$2,01 \times 10^6$
s538	GI.1	GI.P1	29,1	2014	$1,73 \times 10^5$
S582	GII.17	GII.P17	17	2016	$7,55 \times 10^8$
S287	GII.3	GII.P21	12	2009	$2,88 \times 10^7$
S596	GII.1	GII.Pg	23	2017	$1,16 \times 10^7$
S597	GII.4 (Sydney)	GII.P4 (New Orleans)	18	2017	$1,99 \times 10^8$
S598	GII.2	GII.P16	19	2017	$1,87 \times 10^7$
S570	GII.4 (Sydney)	GII.Pe	19	2015	$2,68 \times 10^6$
S555	GII.6	GII.Pe	19	2014	$1,56 \times 10^6$
S514	GII.2	GII.P2	24	2013	$5,46 \times 10^6$
S510	GII.7	GII.P7	19	2012	$8,07 \times 10^6$
S23b	GII.12	GII.Pr	20	2007	$1,98 \times 10^9$
S53	GII.7	GII.P7	27	2003	
s602	GII.14	x	19	2018	$9,40 \times 10^7$

ANNEXE IV

Protocole pour le test ELISA:



Protocole pour ISC-PCR :



ANNEXE V

Protocole de quantification virale pour TV selon la méthode TCID₅₀ (50% tissue culture infective dose)

- Préparation plaque 96 puits pour TCID₅₀

1. Attendre que la lignée des cellules LLC-MK2, cultivé en flasque T75, atteigne une confluence de 90% (plus ou moins 3 jours)
2. Aspirer le milieu avec le VacuSafe et laver les cellules avec 10mL PBS à 37°C
3. Ajouter 5mL trypsine à 37°C et incuber 5 min à 37°C
4. Ajouter 10mL de milieu de croissance à 37°C et à la pipette aspirer - refouler plusieurs fois afin de bien séparer les cellules
5. Ajouter encore 30mL milieu de croissance et bien mélanger
6. A la pipette multicanaux déposer 200µL/puit dans une plaque 96 puits
7. Incuber les plaques 96 puits à 37°C o/n
8. Les cellules pourront être infectées 2 - 3 jours après ensemencement.

- Préparation de l'inoculum

9. Filtrer le surnageant sur filtres 5 – 1,2 – 0,45 et 0.22µm montés sur seringue en transférant directement le filtrat sur un tube Amicon 100kDa
10. Centrifuger l'Amicon® pendant 5 à 10 min à 4300g à 4°C
11. Pendant ce temps conditionner les colonnes Shephadex en laissant passer 5x5mL de milieu d'infectivité. Jeter le milieu qui coule par gravité. Avec les deniers 5mL, centrifuger 1000g 2min et transférer les colonnes dans de nouveaux tubes.
12. Suite à la centrifugation des Amicon, récupérer le rétentât, ajuster son volume à 1,5ml avec du milieu d'infection et transférer sur la colonne Shephadex conditionnée.
13. Centrifuger les colonnes 1000g pendant 2min et récupérer l'éluat (inoculum)
14. A partir de l'inoculum pur, préparer dilutions dégressives 1/10 (tot. 8 dilutions) : 4.5mL milieu d'infectivité plus 500µL inoculum pure.
15. Laisser en incubation à 37°C pendant la préparation/lavages de la plaque 96 puits

- Infection des plaques 96 puits

16. Laver les plaques 96 puits précédemment préparées avec une monocouche cellulaire. Vider les plaques du milieu de croissance et rincer avec du PBS à 37°C à l'aide de la pipette multicanaux. Répéter 3 fois.
17. Ajouter 100µL/puit dans dix puits d'une même ligne pour chaque dilution en faisant couler le liquide sur la paroi des puits pour ne pas déranger le tapis cellulaire.
18. Garder les derniers deux puits de chaque ligne de dilution comme contrôle négatif (100µL milieu d'infectivité sans virus). Préparer la deuxième plaque en duplicat.
19. Incuber à 37°C pendant 1h puis rajouter 100µL/puit de milieu de maintien pauvre en sérum fœtal bovin.
20. Lecture des effets cytopathiques 5 jours après infection.

ANNEXE VI

Protocole d'infection d'entéroïdes différenciés en monocouche par NoV GII (Ettayebi et al., 2016)

Différents types de milieux ont été utilisés pour établir, maintenir ou différencier les enteroïdes :

- Milieu complet sans facteurs de croissance (milieu **CMGF -**) est composé de milieu commercial Advanced DMEM-F-12 (Invitrogen) complété de 100 U/mL de pénicilline-streptomycine (Invitrogen), tampon HEPES 10 mM (Invitrogen) et 1 × GlutaMAX (Invitrogen).

- Milieu complet avec facteurs de croissance (**CMGF +**), constitué de milieu CMGF (-) additionné de 50 ng/mL de facteur de croissance épidermique (EGF) (Invitrogen), 10% de Noggin, 20% de R-spondin et 50% Wnt3A (milieux conditionnés fournis par M.Estes, BCM, Houston), nicotinamide 10 mM (Sigma-Aldrich, Saint Louis, Missouri), 10nM gastrine I (Sigma-Aldrich), 500 nM A-83-01 (Tocris Bioscience, Bristol, Royaume-Uni), 10 µM SB202190 (Sigma-Aldrich), B27 (Invitrogen), N2 (Invitrogen) et 1 mM d'acétylcystéine (Sigma-Aldrich).

Le **milieu de différenciation** est constitué des mêmes composants que ceux du milieu CMGF + sans addition de Wnt3A, SB202190 et nicotinamide, ainsi que d'une réduction de 50% des concentrations de Noggin et de R-spondin.

Le **milieu d'infection** : CMGF- avec 1/100 de GCDCA

Culture d'entéroïdes intestinaux humains :

Les entéroïdes, des cellules souches intestinales humaines issues du jéjunum d'un donneur de phénotype sécréteur (lignée J2), et fournies par M. Estes (Baylor College of Medicine, Houston, USA) sont cultivées en 3 dimensions dans du Matrigel (Corning) en présence de milieu CMGF+ (renouvelé tous les 2-3 jours, dans un incubateur à 37°C et 5% CO₂).

Au bout de 7 jours les entéroïdes sont dissociées par la trypsine 0,05% EDTA (Invitrogen), les cellules sont re-suspendues en milieu CMGF+ additionné de Y-27632 (10 µM, Sigma Aldrich) et déposées en deux plaques 96 puits recouvertes de collagène humain (type IV, Sigma Aldrich) à raison de 100 à 200x10³ cellules/puit, afin d'obtenir des monocouches cellulaires indifférenciées. La différenciation des cellules est ensuite obtenue en remplaçant le lendemain le milieu CMGF+ par du milieu de différenciation, renouvelé tous les deux jours pendant 4 jours

Préparation de l'inoculum :

Comme décrit pour le virus Tulane (Annexe 3)

Infection des entéroïdes en monocouche :

Les cellules en monocouche différenciées sont lavées une fois avec 200µl de CMGF- froid. Le milieu est retiré avant ensemencement de l'inoculum (100µL/puits). Pour chaque point de mesure, 6 puits sont inoculés avec le même inoculum : 3 puits dans chacune des deux plaques, qui constituent 3 réplicats. Les plaques sont incubées 1h à 37°C avec 5% CO₂. Ensuite les puits sont lavés 3 fois avec du CMGF- (température ambiante) et 100µL de milieu d'incubation (milieu de différenciation + GCDCA) sont ajoutés délicatement dans chaque puit des deux plaques. Une plaque sera alors placée à -20°C (annoter : 1hpi) et l'autre à 37°C 5% CO₂ pendant 3 jours (72h) puis congelée à -20°C (annoter : 72hpi).

Mesure de la réplication virale par qRT-PCR :

Après avoir décongelé les plaques, 100µL de tampon de lyse BioMérieux sont déposés dans

chaque puit et incubés 10 min à température ambiante. Le mélange milieu de culture – cellules – tampon de lyse est homogénéisé et transféré dans 2mL de tampon de lyse. L'extraction des acides nucléiques a été automatisé en utilisant EasyMag suivant le programme d'extraction classique avec silice magnétique.

Après extraction les acides nucléiques sont transférés en tube eppendorf et conservés à 4°C si analyses dans les 24h, sinon sont conservés à -20°C ou -80°C.

Le génome viral est quantifié par rRT-PCR NoV GII (Méthode ISO, kit UltraSens), chaque point en duplicat avec une gamme 6 points en duplicats également.

Pour chaque duplicat, la moyenne des Ct est utilisé pour calculer la concentration dans chaque puit lors de l'infection.

Analyse des résultats

La moyenne des 2 Ct est calculé pour chaque duplicat de PCR quantitative, est utilisé pour calculer le ΔCt entre 1h et 72h pour chaque condition et chaque point de triplicat de culture cellulaire. Ceci permet de calculer une ratio d'amplification, égal à $2^{\Delta Ct}$, pour chaque réplicat de culture cellulaire. La moyenne géométrique des 3 ratios d'amplification est ensuite calculée et si le résultat est > 3 , on considère que le virus s'est répliqué et qu'il est bien infectieux.



Capture-based deep sequencing in sewage and in three oyster types using In-house designed hexamer set



Sofia Strubbia, Julien Schaeffer, Alban Besnard, Marions Desdouts, Soizick F. Le Guyader.
Ifremer, laboratoire de Microbiologie, LSEM/SG2M, Nantes France

INTRODUCTION

- A great human norovirus (HuNoV) diversity circulate in the marine environment when water is contaminated by human sewage,
- Oysters farmed in coastal areas can accumulate NoV in their digestive tissues and eventually generate outbreaks,
- The low NoV concentration, the complexity of oyster tissues as sample matrix and the presence of specific binding between oysters and NoV, represent some of the main difficulties for a metagenomic approach

AIMS

- 1) Obtain long sequences, or full HuNoV genomes for strain identification
- 2) Evaluate strain-dependent accumulation patterns in three oyster types.

STRATEGIES

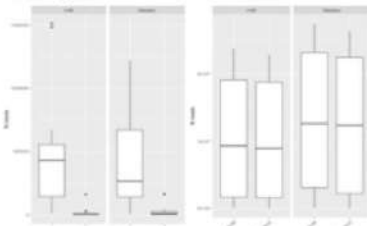
- 1) Develop specific hexamer set to reduce oyster background
- 2) Capture-based deep sequencing

METHOD

- Three oyster batches (*Crassostrea gigas* from 2 different origins and *Ostrea edulis*) were contaminated for 24-hrs in seawater contaminated with a NoV positive sewage.
- HuNoV were extracted using an optimized method based on the ISO-CEN, and quantified by RT-PCR (one step or after cDNA synthesis)¹.
- A set of In-house designed hexamers (I-HD) were used for library preparation in comparison with random hexamers.
- Triplicate libraries were performed for each oyster batch, and a capture-based deep sequencing (VirCapSeq by Roche)² was applied. Negatives controls were included in the run performed using NextSeq Illumina technologies.

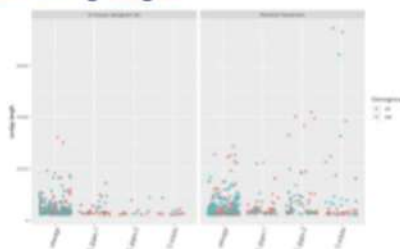
RESULTS

A. Eukaryota and virus reads distribution



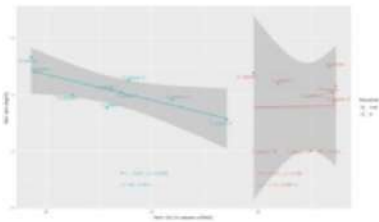
- 166 hexamers were retained for the I-HD set.
- The use of I-HD set reduced reads attributed to mollusk sequences by 10% compared to random hexamers.
- HuNoV reads vary between 0.01% and 0.8%.

C. NoV contig length



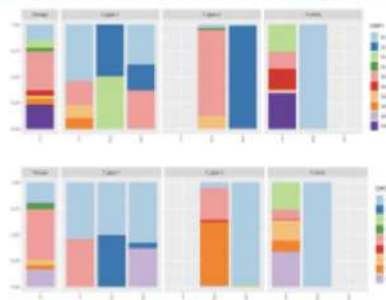
- The I-HD set allowed the identification of 150 and 194 NoV GI and GII contigs respectively, while the random hexamers set allowed the identification of 208 and 284 NoV GI and GII contigs, respectively.
- Longer reads were obtained with random hexamers.

B. Correlation between Ct & number of reads



A linear correlation between Ct values and number of reads was evidenced with random hexamers.

D. HuNoV ORF1 and ORF2 in sewage and oysters



- The best results were obtained using random hexamers
- The triplicate libraries were useful to identify all huNoV diversity in each sample,
- Two full NoV genomes (GI.P2-GI.2 and GII.P7-GII.6) were obtained from the *Ostrea edulis* batch.

DISCUSSION

- Several HuNoV were identified (including complete genomes) showing a greater variability of GI compared to GII HuNoV strains in the three oyster types. The highest strain variability was identified from *Ostrea edulis* batch.
- Technical replicates were found useful to answer our goals, as well as the capture-based approach.
- Developments are still needed to reduce host background during deep sequencing and to obtain longer NoV sequences.

Ref: 1: Strubbia et al. Submitted to Front. Microbiol., 2: Wylie et al. 2015, Genome Res, 25: 1910-20

Acknowledgments: Compare project, H2020 N°643476, S. Strubbia was supported by a fellowship from Ifremer.

Soizick Le Guyader¹, Marion Desdouts¹, Sofia Strubbia¹, Cécile Le Menec¹, David Polo¹, Xi-Lei Zeng², Khalil Ettayebi², Mary K. Estes²

¹ Ifremer, laboratoire de Microbiologie, LSEM/SG2M, Nantes France;

² Molecular Virology and Microbiology, Baylor College of Medicine, Houston, TX, USA.

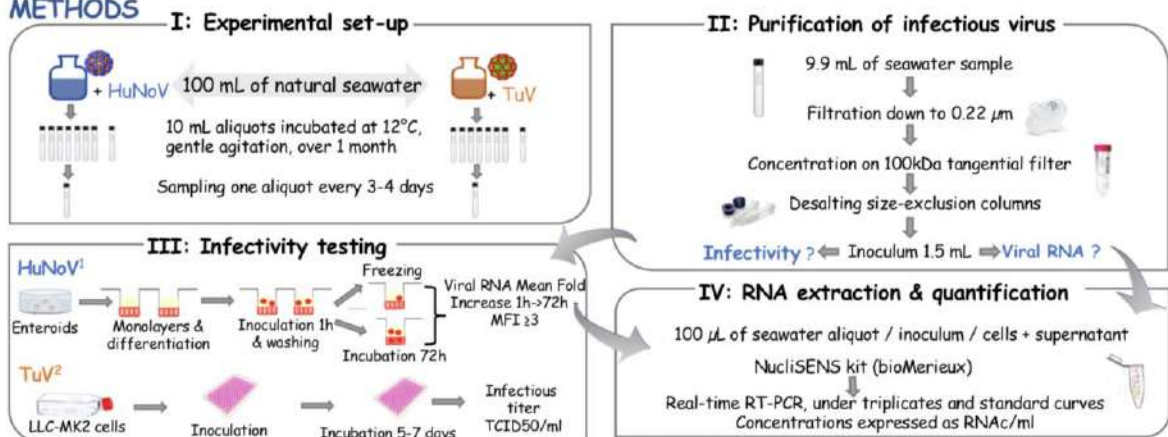
INTRODUCTION

- Human norovirus (HuNoV) may contaminate surface waters, coastal areas, and shellfish, causing foodborne outbreaks worldwide.
- As contamination of environmental samples is low, only a sensitive method such as quantitative PCR allows detection and estimates of virus concentration. However, this technique does not provide information about virus infectivity.
- We previously used Tulane virus (TuV) to model HuNoV environmental stability. The new HuNoV culture model based on human intestinal enteroids¹ now allows direct assessment of infectious HuNoV in environmental matrices such as seawater.

AIMS

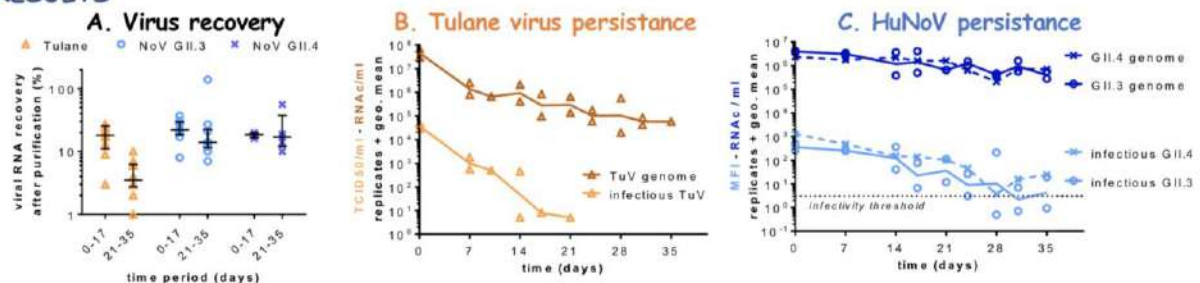
- 1) To evaluate the use of the *in vitro* cultivation system based on human intestinal enteroids to estimate the persistence of infectious HuNoV in seawater.
- 2) To compare huNoV infectivity with the detection of infectious TuV.

METHODS



- Two HuNoV strains were selected based on their capacity to be amplified in the enteroid model (1) and also on their implication in shellfish borne outbreaks for GII.3 (exp. 1 & 2) and their importance in human population for GII.4 (exp. 2 only),
- The Tulane virus was selected as a surrogate due to shared biology with HuNoV (2), and included in both experiments.

RESULTS



- Using a method optimized for TuV³, the recovery of viral RNA following purification (A) was even more efficient for both HuNoV strains, but tended to decrease with time for TuV.
- TuV presented a fast initial decrease in RNA, and infectious virus was below the quantification threshold after 14-21 days (B).
- For HuNoV, RNA levels showed a reproducible, slow decrease. Infectious HuNoV was recovered throughout the 2nd experiment and up to day 24 for the 1st (GII.3 only) as shown by efficient replication on human enteroid cells (viral RNA mean fold increase [MFI] ≥3 between 1h and 72h post-inoculation) (C).

DISCUSSION

- The observed decreases in TuV RNAc and TCID50/ml are comparable to previous data obtained in oyster digestive tissues³.
- HuNoV genome is very stable for both strains with less than one Log₁₀ decrease in RNAc/ml over one month.
- The two HuNoV strains remained infectious in seawater for 24 to 35 days based on viral replication in the enteroid cells. We observed a decrease in the MFI during replication testing that could reflect a partial degradation.
- Other experiments are needed to confirm these preliminary results, and other NoV strains need to be evaluated.

Ref: (1) Ettayebi et al. 2016, Science, 353: 1387-93; (2) Farkas 2015, J. Gen. Virol., 96:1504-14; (3) Polo et al. 2018, Front. Microbiol. 9:716.

Funding : Grant from the French Ministry of Food (DGAL).

Julien Schaeffer, Marion Desdouits, Alban Besnard, Sofia Strubbia, Soizick F. Le Guyader

Ifremer, Laboratoire de Microbiologie, LSEM/SG2M, Nantes France

Contact : julien.schaeffer@ifremer.fr ; marion.desdouits@ifremer.fr ; soizick.le.guyader@ifremer.fr

Introduction

- L'approche métagénomique utilisant le séquençage haut débit permet de décrire l'ensemble des acides nucléiques présents dans un échantillon.
- L'application de cette méthodologie à l'analyse des eaux usées brutes semble séduisante pour identifier les différents virus entériques circulant dans la population.
- Identifier des petits génomes ARN dans ce type d'échantillons particulièrement chargés en micro-organismes de diverses origines nécessite une préparation de l'échantillon afin de sélectionner les particules d'intérêt.

Echantillons

Des échantillons d'eau usée brute prélevés dans 7 stations d'épuration de l'ouest de la France entre 2006 et 2018 et conservés à -20°C ont été sélectionnés.

Méthode

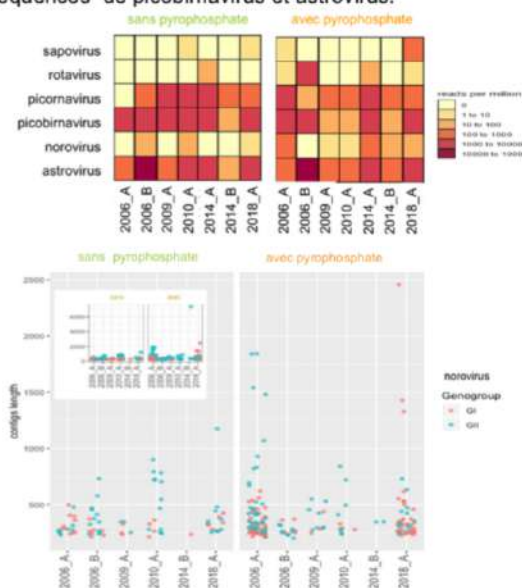
Pour favoriser la récupération des virus à ARN:

- ✓ Utilisation du pyrophosphate de sodium, détergent favorisant l'éluion des particules virales avant l'étape de concentration par le PEG,
- ✓ Elimination des ADNs par traitements enzymatiques (☒).



Résultats

- 30 à 90% des contigs n'ont pu être identifiés et 2 à 60% correspondent à des séquences non virales, principalement bactériennes.
- Pour la fraction virale, l'abondance varie entre 3,5 et 44% des reads, et l'utilisation du pyrophosphate semble avoir peu d'influence.
- Les contigs les plus abondants correspondent à des séquences de picobirnavirus et astrovirus.



L'analyse des contigs assignés comme norovirus montre un apport positif du pyrophosphate avec des contigs significativement plus longs, y compris un génome complet (insert).

Conclusion

- Les principales familles de virus entériques humains ont été identifiées dans les échantillons analysés.
- Obtention d'un génome complet de norovirus (2014_B) grâce à l'utilisation du pyrophosphate de sodium.
- Malgré l'optimisation de la méthode, les contigs restent globalement courts (< 500 b) d'où l'intérêt de poursuivre l'optimisation dans la préparation de l'échantillon ainsi que l'étape d'assemblage du pipeline bio-informatique.

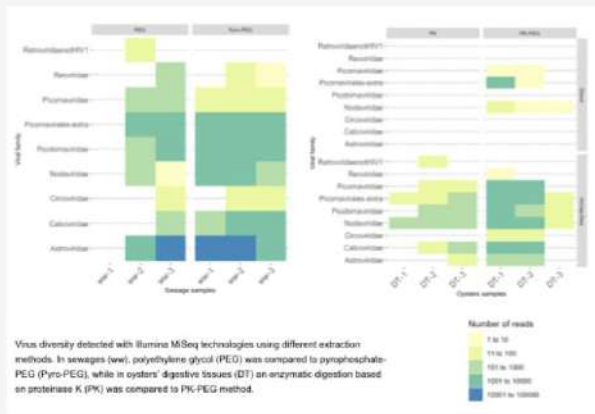
Metagenomic approach to study norovirus diversity in sewage and oyster samples

Sofia Strubbia¹, Julien Schaeffer¹, My V.T. Phan², Matthew Cotten², Bas B. Oude Munnik², David F. Nieuwenhuijse², Marion Koopmans², Soizick F. Le Guyader¹

¹ Ifremer, Laboratoire de Microbiologie, LSEM/SG2M, Nantes France

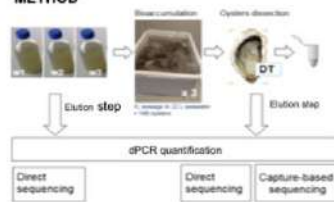
² Department of Viroscience, Erasmus Medical Centre, Rotterdam, The Netherlands

Sewages contain a large diversity of pathogenic and non-pathogenic microorganisms, reflecting what is circulating in the population. Storm overflow, rainfall or untreated sewages may contaminate the marine environment thus oysters farmed in coastal areas. Norovirus (NoV), is a small (7.2kb), ssRNA non-enveloped virus known as the major cause of human gastroenteritis worldwide. It is characterized by its diversity and resistance in the environment. Illumina MiSeq technologies were used to study NoV diversity in the environment and to evaluate the NoV selection performed by oysters during filter feeding activity.



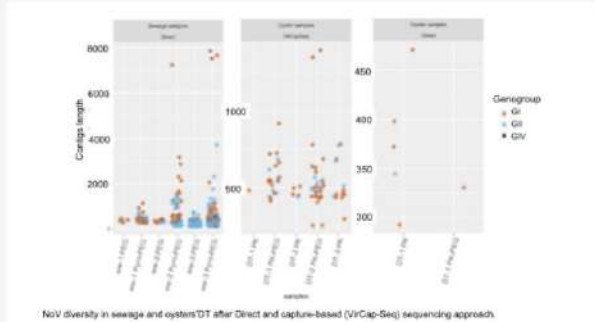
Virus diversity detected with Illumina MiSeq technologies using different extraction methods. In sewage (sw), polyethylene glycol (PEG) was compared to pyrophosphate-PEG (Pyro-PEG), while in oysters' digestive tissues (DT) an enzymatic digestion based on proteinase K (PK) was compared to PK-PEG method.

METHOD



Three raw sewage samples were collected and each sewage sample was used to contaminate a tank containing seawater and oysters (*Crassostrea gigas*) during 24 hours. Different elution protocols were used to recover NoV from samples and Illumina MiSeq technology was used to study NoV diversity in sewage and bioaccumulated oysters comparing agnostic and capture-based sequencing approach.

The kit VirCap-Seq (Roche) was used to concentrate NoV from oysters' samples. For metagenomics analysis, SLIM pipeline (Cotten et al, 2014) was applied on the raw reads and coverage and relative abundance of the constructed contigs were estimated by a mapping (Bowtie2). NoV genotyping tool v2.0 (Kroneman et al, 2011) was then used for identification of the NoV strains. Six whole genomes of NoV GI were obtained from waste water samples (strains GI.P1-GI.1, GI.P2-GI.2 and GI.P4-GI.4). For NoV GII, the longest sequence detected was GII.P7-GII.6 (5050nt).



NoV diversity in sewage and oyster DT after Direct and capture-based (VirCap-Seq) sequencing approach.

Direct deep sequencing from oysters' DT showed a small amount of NoV reads with lengths ranging from 299 to 471bp.

For low contaminated samples capture-based approach is essential to allow metagenomics analysis, but still method developments are needed to increase the recovery of viral particles. Considering the variability of NoV, the characterization of long contigs is important for strain identification.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 643476

Etude de la diversité des norovirus dans les eaux usées par la métagénomique

Sofia Strubbia, Julien Schaeffer, Alban Besnard, Soizick F. Le Guyader

Ifremer, Laboratoire de Microbiologie, LSEM/SG2M, Nantes France

Les eaux usées contiennent une large diversité de microorganismes, reflétant les divers microbiotes de la population. Pendant la période hivernale, des événements météorologiques exceptionnels peuvent entraîner des déversements d'eaux usées non traitées (débordement, rupture de canalisation...) directement dans l'environnement.

Les norovirus (NoVs) sont des petits virus non enveloppés, possédant un génome à ARN positif. Ces virus constituent l'agent majeur des gastroentérites aiguës hivernales chez l'Homme et sont présents en quantité importante dans les rejets pendant cette période. Les NoVs sont caractérisés par une importante variabilité génomique. Les souches majoritaires responsables de maladie chez l'Homme sont les génogroupes GI et GII. Leur résistance dans l'environnement leur permet de persister pendant plusieurs semaines. Une conséquence est la possible contamination des aliments, et la diversité génomique complique l'analyse des souches présentes.

Le séquençage métagénomique direct, en décrivant tous les acides nucléiques présents dans un échantillon, constitue une approche utile pour apprécier la diversité génomique.

METHODE



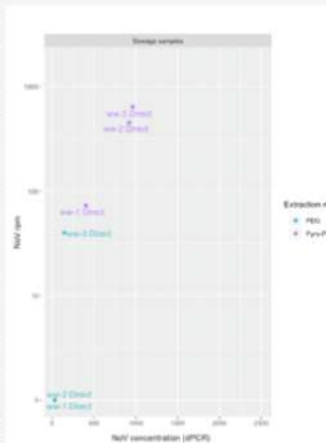
Trois échantillons d'eaux usées brutes ont suivi deux protocoles d'éluion virale : acidification directe (bleu-PEG), ou addition de pyrophosphate (Bisseux et al., 2014) et sonication (violet- Pyro-PEG) avant précipitation par le polyéthylène glycol (PEG). Plusieurs étapes ont ensuite été appliquées pour éliminer les bactéries, ou les acides nucléiques libres avant extraction des ARN par le kit NucliSENS (bioMérieux). Après quantification par digitale RT-PCR (Polo et al., 2016), les bibliothèques ont été préparées et le séquençage réalisé sur un Illumina MiSeq. Pour l'analyse métagénomique, le pipeline *SLIM* (Cotten et al., 2014) a été appliqué aux données brutes (reads) et la couverture et l'abondance relative des contigs construits ont été estimées par *Bowtie2* (mapping). L'outil *NoV genotyping tool v2.0* (Kroneman et al., 2011) a ensuite été utilisé pour l'identification des souches de NoV.

DISCUSSION

Les eaux usées traitées avec la méthode Pyro-PEG ont montré un nombre plus élevé de reads en relation avec des concentrations plus élevées. Obtenir de longs fragments est un point important dans l'identification des souches de NoV présentes considérant la variabilité génomique de ces virus et leur capacité à créer des recombinants. L'approche métagénomique montre son intérêt pour en étudier l'épidémiologie moléculaire et évaluer l'impact des traitements des eaux usées sur l'élimination sélective de certaines souches.

REFERENCES

- Bisseux et al., *Eurosurveillance*. 2018;23(7):1-11..
- Cotten et al., *PLoS ONE*. 2014. doi:10.1371/journal.pone.0093269.
- Kroneman et al., *Archives of Virology*. 2013;158(10):2059-2068.
- Polo et al. *Emerging and Infectious Disease*. 2016; 22: 2189-2191



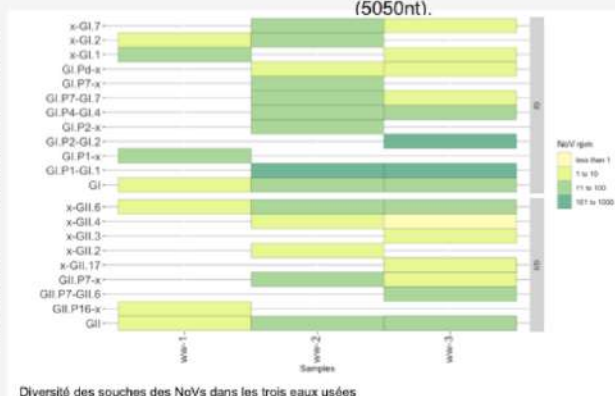
Relation entre nombre des reads par million (rpm) obtenus et concentration des NoV en copies génomes par mL quantifié par digitale RT-PCR (iPCR).

RESULTATS

Un nombre plus élevé de contigs NoV ont été identifiés avec la méthode Pyro-PEG (27, 67 et 81 contigs pour les échantillons ww-1, ww-2 et ww-3, respectivement), par rapport à la méthode PEG (5, 14 et 306 contigs respectivement), en accord avec les concentrations plus élevées obtenues avec cette méthode.

Les séquences courtes obtenues avec la méthode PEG (entre 128 et 613 nt) n'ont pas permis l'identification des souches.

La méthode Pyro-PEG a permis le séquençage de six génomes complets de NoV GI (souches GI.P1-GI.1, GI.P2-GI.2 et GI.P4-GI.4). Pour les NoV GII, la séquence la plus longue détectée était GII.P7-GII.6 (5050nt).



Diversité des souches des NoVs dans les trois eaux usées

Metagenomics for food safety

Claudia Wylezich¹, Patrick Zitzow¹, Leonie Kauer¹, Martina Zoli¹, Ariane Belka¹, Anne Pohlmann¹, Sofia Strubbia², Soizick Le Guyader², Martin Beer¹ & Dirk Höper¹

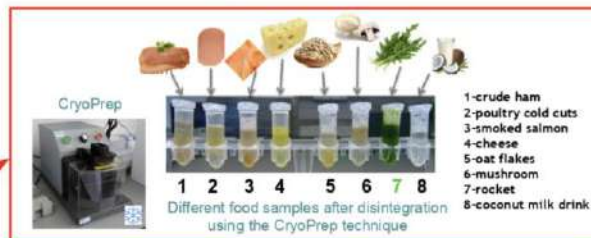
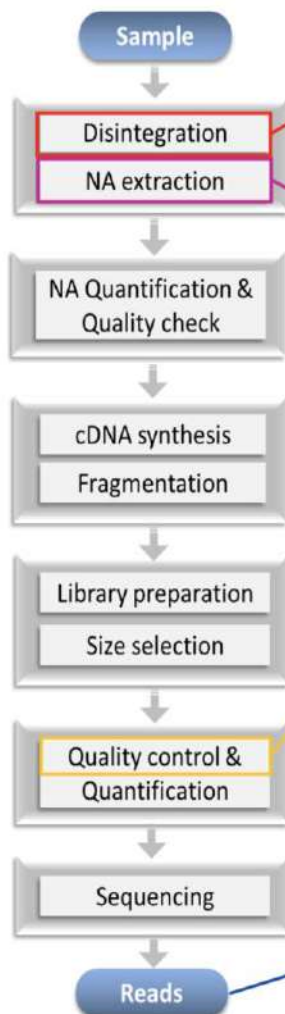
¹Friedrich-Loeffler-Institut (FLI), Institute of Diagnostic Virology, Greifswald-Insel Riems, Germany.

²IFREMER, Laboratoire de Microbiologie-LNR, Département RBE, Nantes - France

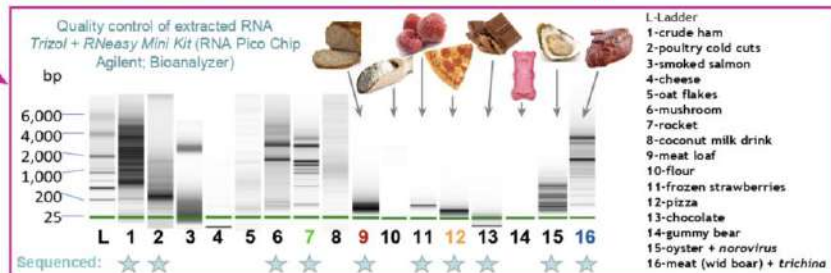
Within WP2, sample-processing pipelines including pathogen inactivation, nucleic acid (NA) extraction, and subsequent processing until sequencing were improved and intensively tested for different matrices. Laboratory Operating Procedures (LOPs) were developed for standardized protocols providing best results – regarding quality & quantity of extracted NA and sequence reads. We focused on food

matrices during the second year of COMPARE since they were very seldom handled and no protocols were available. We used the FLI in-house metagenomics workflow for testing conventional food of animal or plant origin, both processed and unprocessed, and, in addition, contaminated foods. Results for homogenization, RNA extraction, library preparation and sequencing outcome of food products are shown.

Overview of sample processing workflow

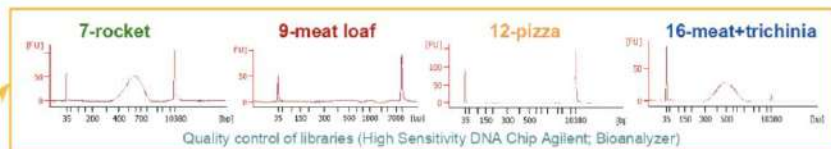


The results show a good performance of disintegration and extraction methods for unprocessed food items like rocket, mushrooms, oysters (lanes 6, 7, 15, 16 on the RNA Pico Chip) whereas in processed foods, content and integrity of RNA seems to be inherently very low (lanes 2, 3, 4, 8, 9, 12, 13 and 14).

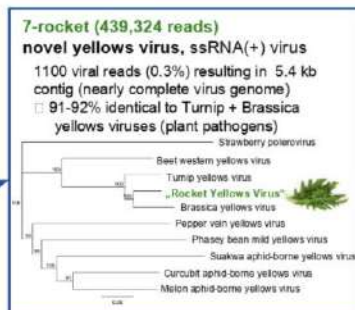


RNA extraction of special matrices like fatty (cheese, lane 4), fiber-containing (oat-flakes, lane 5) and samples of low pH (strawberries, lane 11) is challenging and possibly needs pre-treatments, like defatting or neutralization.

Another difficult sample type is material changing its physical condition under the influence of liquid nitrogen (gummy bears, lane 14, turned out to become glass-like within liquid nitrogen resulting in sharp-edged shivers destroying the Covaris tubes).



□ We were able to detect viruses and parasites in different data sets, including processed foods:

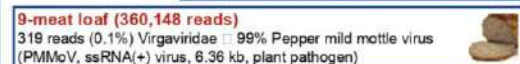


12-Pizza* recipe
Ingredients for dough
Flour (wheat, 125,475 reads *Triticum aestivum*)
Yeast (132,661 reads, *S. cerevisiae*)
Rape oil (24 reads Brassicaceae)
Bean flour (52 reads Fabaceae)

Ingredients for topping
Ham (159 reads *Sus scrofa*)
Ground beef (8,579 reads Bovidae)
Mushrooms** (84,151 reads *Agaricus bisporus*)
Tomato (29 reads *Solanum* spp.)
Onion (25 reads *Allium cepa*)
Oregano (18 reads *Origanum vulgare*)

*incl. 44 reads (0.01%) *Heterobosidion mitovirus 1* (dsRNA virus, also found in sample 6, mushroom)

**427,509 reads



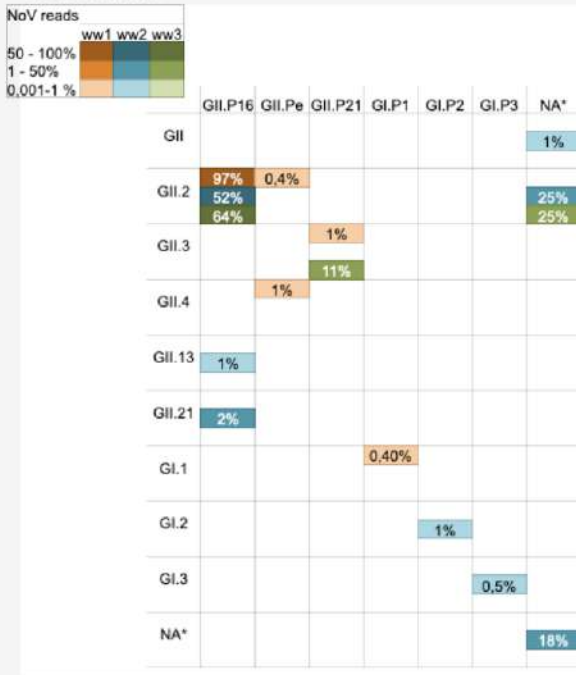
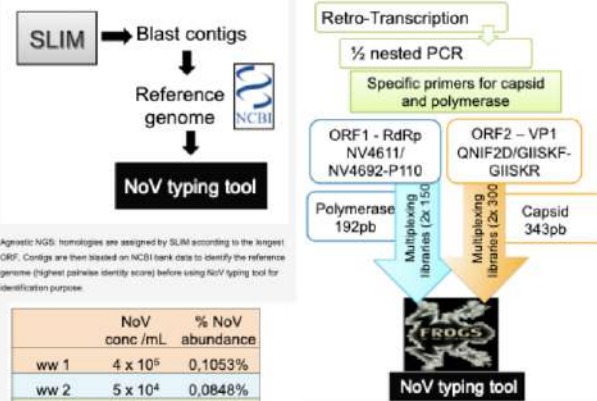
Metagenomics and metabarcoding approaches to evaluate norovirus diversity in wastewater samples

Sofia Strubbia¹, Julien Schaeffer¹, Marion Desdouits¹, My V.T. Phan², Matthew Cotten², Marion Koopmans², Soizick F. Le Guyader¹

¹ Ifremer, Laboratoire de Microbiologie, LSEM/SG2M, Nantes France

² Department of Viroscience, Erasmus Medical Centre, Rotterdam, The Netherlands

Noroviruses (NoV) are known as the major cause of human gastroenteritis worldwide. During the winter gastroenteritis season a large number of strains co-circulate in the population and thus are excreted in wastewaters. We compared agnostic next generation sequencing (NGS) and specifically primed amplicon sequencing (metabarcoding) on three wastewater samples collected from December 2016 to February 2017. Our goals are to describe NoV diversity and to identify the major strains of NoV circulating in the environment.



Agnostic NGS: homologies are assigned by SLIM according to the longest ORF. Contigs are then blasted on NCBI bank data to identify the reference genome (highest pairwise identity score) before using NoV typing tool for identification purpose.

	NoV conc./mL	% NoV abundance
ww 1	4 x 10 ⁵	0,1053%
ww 2	5 x 10 ⁴	0,0848%
ww 3	4 x 10 ³	0,02%

NoV quantification by qRT-PCR (genomic capsid/cp) compared to the total abundance of NoV reads in wastewater samples.

Metabarcoding approach: involves deep sequencing of a portion of the viral genomes amplified by PCR using degenerate primers recognizing a whole genogroup. The genome region targeted during the amplification codify for the viral capsid and for the polymerase for identifying.

Combining metagenomics (double-chart) and metabarcoding data (based on the left and polymerase on the top) for the three wastewater samples (ww1, ww2, ww3). Sequences are ranked according NoV reads abundance and identified by a hit scale. *NA: Not assigned.

Match between metagenomics and metabarcoding on capsid sequences (NoV GII.2) but not for polymerase sequences. NGS reported low abundance of NoV reads and showed consistent relation between NoV concentration and total abundance of NoV reads. Metabarcoding is a feasible approach even with low NoV concentrations, although results may be biased by the use of primers and multiple cycles PCR and no links between capsid and polymerase data are possible. Method developments are still needed to improve the recovery and to concentrate NoV particles from sewage samples.

Titre : Norovirus et huître, infectiosité et approche génomique

Mots clés : Norovirus, huîtres, eaux usées, metagénomique, infectiosité

Résumé : Cette thèse a été réalisée dans le cadre du projet Européen H2020 COMPARE qui a pour ambition d'améliorer et de faciliter la détection et la prise de décision face à la survenue d'épidémies chez l'Homme ou l'animal en utilisant des nouveaux outils de séquençage génomique.

Mes recherches ont ciblé les norovirus, connus comme les principaux agents de gastroentérite virale humaine. Leur variabilité génétique élevée et les fréquents événements de recombinaisons associés à une forte résistance dans l'environnement contribuent à la survenue régulière de ces épidémies de gastroentérites hivernales. Les huîtres élevées dans des zones côtières contaminées peuvent accumuler les norovirus pendant leur activité de filtration. L'utilisation des nouvelles techniques de séquençage haut débit et d'infectiosité sur entéroïdes, nous ont permis de progresser dans

la compréhension des mécanismes de survie et de dissémination des souches de norovirus. La préparation de l'échantillon (eaux usées ou coquillages), est primordiale pour concentrer et purifier les norovirus avant d'appliquer les méthodes de metagénomique. Les méthodes développées et optimisées dans cette thèse pour ces matrices, nous ont permis de caractériser des génomes complets de norovirus. Disposer de méthode sensible permettant d'identifier les norovirus mais également les autres virus entériques humains, émergents ou re-émergents permettra dans le futur de limiter leur transmission. L'utilisation des cellules souches intestinales nous a permis de démontrer, pour la première fois, la persistance des norovirus dans l'eau de mer.

Title : Norovirus and oysters, infectivity and genomic approach

Keywords : Norivurs, oysters, sewage, metagenomics, infectivity

Abstract : This thesis is part of the H2020 COMPARE project, which aims to accelerate the detection and decision-making of humans and animals outbreaks through the use of new genomic sequencing tools.

My research focused on noroviruses, known as the main agents of human viral gastroenteritis. Their great genetic variability, the recurrence of new recombination and the high resistance in the environment contribute to the regular occurrence of winter gastroenteritis outbreaks. Oysters farmed in contaminated coastal areas may accumulate noroviruses during their filtration activity. The use of new high throughput sequencing techniques and the novel cell culture approach to study norovirus infectivity has enabled us to progress in

understanding the mechanisms of survival and dissemination of norovirus strains.

Sample preparation (sewage or shellfish), is essential to concentrate and purify noroviruses before applying metagenomic methods. The methods developed and optimized in this thesis for these matrices, allowed us to characterize complete genomes of norovirus. Having a sensitive method for identifying noroviruses but also other human, emerging or re-emerging, enteric viruses will help in the future to limit their transmission.

The use of intestinal stem cells has allowed us to demonstrate, for the first time, the persistence of noroviruses in seawater.