International Journal of Food Microbiology 323 (2020) 108588

Contents lists available at ScienceDirect

International Journal of Food Microbiology

journal homepage: www.elsevier.com/locate/ijfoodmicro

Metagenomic to evaluate norovirus genomic diversity in oysters: Impact on hexamer selection and targeted capture-based enrichment

Sofia Strubbia, Julien Schaeffer, Alban Besnard, Candice Wacrenier, Cécile Le Mennec, Pascal Garry, Marion Desdouits, Françoise S. Le Guyader*

Ifremer, Laboratoire de Microbiologie LSEM-SG2M, France

ARTICLE INFO

Keywords:
Norovirus
Oysters
Metagenomic
Human enteric viruses

ABSTRACT

Human virus transmission through food consumption has been identified since many years and the international trade increases the risk of dissemination of viral pathogens. The development of metagenomic approach holds many promises for the surveillance of viruses in food and water. This work aimed to analyze norovirus diversity and to evaluate strain-dependent accumulation patterns in three oyster types by using a metagenomic approach. Different hexamer sets to prime cDNA were evaluated before capture-based approach to enhance virus reads recovery during deep sequencing. The study includes the use of technical replicates of artificially contaminated oysters and the analysis of multiple negatives controls. Results showed a clear impact of the hexamer set used for cDNA synthesis. A set of In-house designed (I-HD) hexamers, selected to lower mollusk amplification, gave promising results in terms of viral reads abundancy. However, the best correlation between C_T values, thus concentrations, and number of reads was observed using random hexamers. Random hexamers also provided the highest numbers of reads and allowed the identification of sequence of different human enteric viruses. Regarding human norovirus, different genogroups and genotypes were identified among contigs longer than 500 bp. Two full genomes and six sequences longer than 3600 bases were obtained allowing a precise strain identification. The use of technical triplicates was found valuable to increase the chances to sequence viral strains present at low concentrations. Analyzing viral contamination in shellfish samples is quite challenging, however this work demonstrates that the recovery of full genome or long contigs, allowing clear identification of viral strains is possible.

1. Introduction

Human virus transmission through food consumption has been identified since many years and preventative measures have been undertaken in many countries. Nonetheless, these remain imperfect, and the international trade increases the risk of dissemination of viral pathogens (Li et al., 2018). The main sources of food contamination with viruses are infected food handlers, zoonotic transmission (essentially for hepatitis E virus) or contact with sewage-contaminated waters during the production process (Bosch et al., 2018). For the two first cases, a test, targeting the virus excreted by the food handler or by the animal, may directly identify the contaminant in the foodstuff. However, when dealing with food contaminated during production, the identification of which Human enteric viruses are present may be more challenging. Indeed, a variety of human enteric viruses are excreted by the human population and may contaminate the environment mainly through sewage discharge (Metcalf et al., 1995). With the increasing demands of water for agriculture or human activities, wastewater re-use will be necessary, increasing the risk for environmental virus pollution, and subsequent viral transmission through contaminated food (Sano et al., 2016). Being able to identify the virus contaminating a food sample will be of primary importance for risk management and consumer safety (Nieuwenhuijse and Koopmans, 2017), and to follow regulations stipulating that only food without any Human pathogens can be marketed (EC 178/2002). However, the difficult question to link the genome detection and the presence of infectious particles that will constitute another challenging issue to be solved (Ettayebi et al., 2016). Indeed, Human enteric virus contamination in food is based on real-time RT-PCR, method allowing a quantitative approach by comparing the C_T value obtained with a sample to a standard curve. For now, only this method represent limit of detection (LOD) values compatible with food contamination.

Bivalve molluscan shellfish such as oysters are consumed worldwide and considered as a source of important nutrient (Venugopal and

* Corresponding author at: Ifremer, Laboratoire de Microbiologie, LSEM-SG2M, Rue de l'Ile d'Yeu, BP 21105, 44311 Nantes cedex 03 Nantes, France.
E-mail address: soizick.le.guyader@ifremer.fr (F.S. Le Guyader).

https://doi.org/10.1016/j.ijfoodmicro.2020.108588
Received 15 November 2019; Received in revised form 10 March 2020; Accepted 11 March 2020
Available online 12 March 2020

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

Metagenomics refer to the description of all nucleic acid sequences present in a sample allowing the description of bacteria, archaea, eukaryotes or viruses (Giu et al., 2019; Kallies et al., 2019). However, despite the abundance of viruses on the earth, the virome is still poorly described (Rosani et al., 2019). As a catch-all, agnostic and sensitive approach potentially allowing the detection and identification of the whole viral genomic material, metagenomic sequencing holds many promises for the surveillance of viruses in food and water (Nieuwenhuijse and Koopmans, 2017). Nevertheless, several issues remain, owing to both the diversity of potential viral contaminants and the very low level of contamination in complex food matrices (Nieuwenhuijse and Koopmans, 2017). Indeed, viral metagenomics is generally performed by removing as much host, matrix and bacterial nucleic acids as possible followed by nuclease treatment to remove free nucleic acids. These steps are of primary importance to identify human enteric viruses, most of them having a small RNA genome, among the vast amount of DNA derived from the host or the food matrix. Besides, most viruses of interest bearing an RNA genome require a conversion to cDNA. Finally, enrichment or amplification steps are also frequently needed to increase the recovery of viral sequences (Wylie et al., 2018). An approach combines cDNA synthesis and enrichment in virus sequences by using primers selected to avoid reverse-transcription of the host ribosomal RNA while allowing the priming of mammalian viruses (Endoh et al., 2005). Such non-ribosomal primers have been successfully applied to the sequencing of mammalian viruses in clinical samples (Cotten et al., 2014a, 2014b), and could contribute to the detection of human viruses in food.

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

2. Materials and methods

2.1. Oyster batches

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

Immediately after arrival in the laboratory, *Escherichia coli*, total microbiota and total marine bacteria were counted in shellfish flesh as described previously. Briefly, 15 oysters of each batch were homogenized (1:10) with Peptone Salt Solution (Oxoid, France) and further two one-Log dilutions were performed. Each of these dilutions were then inoculated on Plate Count Agar (Oxoid) and Marine agar (FisherScientific, France) on duplicate and incubated for 72 h at 30 °C and 20 °C respectively (Weiner et al., 1985). *Escherichia coli* were quantified according to the standard NF V08–106 (AFNOR, 2010).

2.2. Bioaccumulation

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

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

2.3. Quality controls

Non-bioaccumulated oysters, seawater, and UltraPure™ DNase/RNase-Free Distilled Water (TechnoFisher, France) used during the experiment were verified for the no detection of NoV contamination and included in deep-sequencing analysis. All these controls were analyzed once.

After bioaccumulation all three oyster samples were first analyzed using the ISO method (ISO 15216-1, 2017). After verification of the extraction efficiency and lack of inhibitors, the quantification was performed to verify the efficacy of the bioaccumulation.

2.4. Samples preparation

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

DT aliquots (2 g) were incubated with 2 mL of proteinase K (Thermo Fisher Scientific) for 15 min at 37 °C and 15 min at 60 °C. After sonication (Bandelin UD 2200 with cup-horn adaptor) for 3 min at maximum power (cycles of 1 min with 1 min on ice), the mixture was centrifugated for 5 min at 3000 ×*g*. Then the supernatant was mixed with 10 mM sodium pyrophosphate (Bisseux et al., 2018) and incubated 40 min at room temperature under gentle agitation. After 20 min of centrifugation at 8000 ×*g*, the supernatant (3 mL) was recovered and 1.5 mL of a PEG 24% (wt/vol)-sodium chloride (1.2 M) solution was added prior incubation for 1 h at 4 °C. After centrifugation for 20 min at 8000 ×*g*, the pellet was resuspended in 1 mL of glycine buffer (0.05 M) pH 9.

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

2.5. NoV quantification/detection

Each extract was analyzed by real-time RT-PCR using primers and probe targeting the conserved region between ORF1 and ORF2 for HuNoV GI and GII, using a one-step real-time RT-PCR kit (Invitrogen) and an M × 3000P quantitative PCR (qPCR) system (Agilent Technologies, France) (Le Guyader et al., 2009).

2.6. In-house designed hexamer set

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

2.7. Library preparation and capture-based deep sequencing

Technical triplicates were performed for each batch. Two sets of hexamers, random hexamers (New England Biolabs (NEB), USA) and I-HD hexamers were used to prepare the cDNA from the RNA extracted from the three oyster batches, the sewage sample as well as from all controls (seawater, control oysters, and UltraPure™ water). Additionally, RNA from C.gigas-1 batch was also reverse-transcribed using the Endoh hexamer set (Endoh et al., 2005). The reverse transcription was performed using the enzyme SuperScript II (ThermoFisher) and 5 µL were used to perform a real-time HuNoV GII PCR using primers and probe as described above. A total of 33 cDNA was obtained. After completion of the cDNA, they were organized into three pools for the next step, considering a uniform distribution of technical replicates (Fig. 1). Each pool contained seven cDNA prepared from a bioaccumulated oyster batch, with an even distribution of the hexamer set, two negative oyster cDNA and two libraries prepared from one water sample (either sewage, or seawater or UltraPure™ water sample) (Fig. 1). After double strand DNA synthesis using second strand reaction buffer from NEBNext Ultra RNA Library prep, a physical fragmentation (Ultrasonicator M220, Covaris) was performed for 3 min. To reach the advised quantities preconized by the manufacturer, ten amplification cycles were added to the control sample extracts. Libraries were prepared using the kit KAPA Hyper Prep (Roche), adapter ligation was performed and viral sequences were enriched using VirCapSeq-VERT capture panel (Roche) (Wylie et al., 2015). The capture was done simultaneously on the 3 pools.

Sequencing was performed on an Illumina NextSeq 500 using High output kit V2 (Illumina) to generate 2 × 150 bp reads. Run is validated if the number of passing filters is > 70%.

2.8. Bioinformatic pipeline

Metagenomics bioinformatics pipeline included quality trimming, *de novo* assembly, and sequence alignment and annotation. Illumina adapters were removed from raw reads before trimming from 3′ to 5′ to reach a Phred score above 25 (TrimGalore v.0.4.1). Trimmed reads were *de novo* assembled using SPAdes v.3.12.0 to obtained contigs (Bankevich et al., 2012). The first step of SPAdes pipeline was error correction (BayesHammer) with default setting, then assembly was performed using de Bruijn graph algorithm and a set of k-mer lengths of 21, 33, 55, 77. To ensure a reliable identification, only contigs longer than 500 nt were identified using BLAST v.2.6.0, and only sequences with identity score above 85% were considered during data analysis. The coverages and abundances of selected contigs were further evaluated using Bowtie 2 v2.3.0. (Langmead and Salzberg, 2012). Contigs originating from oyster DT sequenced in triplicate were analyzed separately and sequences identified as belonging to *Caliviridae* family were genotyped using HuNoV genotyping tool v.2.0 and nominated using the new taxonomical classification (Kroneman et al., 2011; Chhabra et al., 2019). The high amount of reads produced from sewage samples required a normalization step prior assembling using bbmap v.38.22 to reach a coverage of 40 (Bushnell et al., 2017).

2.9. Statistical analysis

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

3. Results

3.1. Oyster's microbiological analysis and controls

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

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

HuNoV concentrations in the sewage sample used for the bioaccumulations were 6.7×10^4 and 2.9×10^4 cRNA/µL for GI and GII respectively. After bioaccumulations, the three oyster batches reached similar NoV contamination levels, as measured by rRT-PCR following the ISO method, with C_T values ranging from 29.8 to 32.7 for both genogroups, corresponding to concentrations from 9.7×10^2 RNAc/g of DT for *O. edulis* batch to 2×10^3 RNAc/g of DT for both *C. gigas* batches. For all negative control oyster batches (non-bioaccumulated oysters) and the seawater used for the bioaccumulation, NoV RNA levels were under the limit of detection of the method (124 and 196 RNAc/g of DT for HuNoV GI and GII respectively).

3.2. Theoretical calculation for the three hexamer sets

The three hexamer sets used in this study, Endoh, Random and I-HD were evaluated for their theoretical probability to recognize a panel of Human enteric virus sequences, as well as mollusk and rat genomes among eukaryote sequences (Fig. 2). The random set presented the same probability (around 23%) to match with all sequences selected for our evaluation. Endoh and I-HD sets presented a lower probability to recognize rat sequences than the random hexamer set. As expected the I-HD set did not match with shellfish sequences, while the Endoh set recognized shellfish sequences with a probability around 12 to 15%. Regarding virus sequences, the Endoh set displayed a large variability

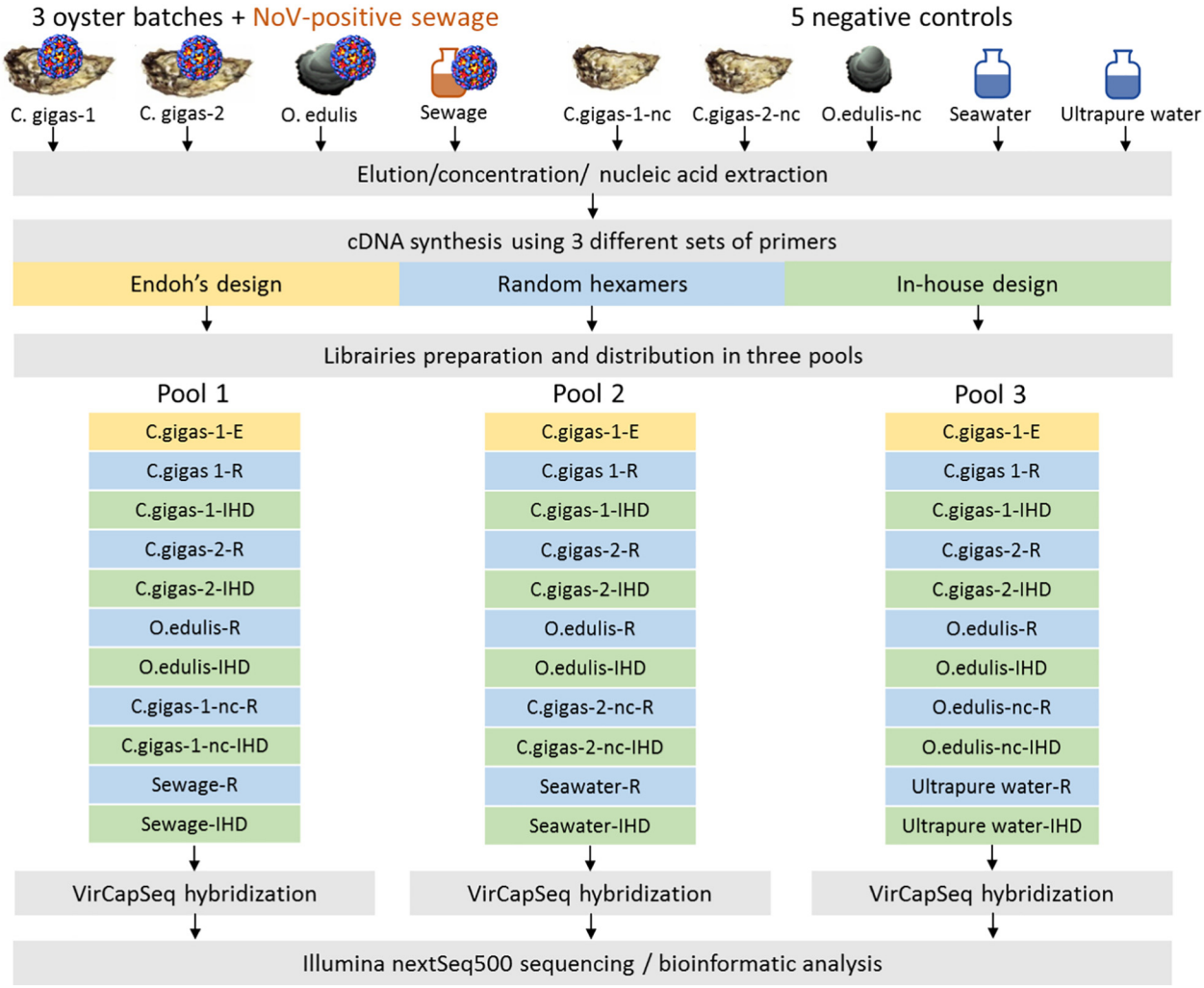

Fig. 1. Workflow of samples and libraries preparation and pool organization prior to hybridization for capture-based deep sequencing.
cDNA were prepared using Endoh (E, yellow), Random (R, blue) or *In-house* designed (I-HD, green) hexamer sets. Negative control (nc) were included in the different pool. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

with a higher probability to recognize picobirnavirus and rotavirus sequences compared to HuNoV. The I-HD set presented less variability in the probability to match with the different human enteric virus sequences selected, with a slightly higher probability to recognize HuNoV sequences than the Endoh set. Here, we aimed at evaluating the performance of our I-HD set to recover human enteric virus sequences in oyster extracts, by comparing them to random primers on all samples. The Endoh set, which is optimized for mammalian virus in rodent cells, was used to prepare triplicate libraries for only one sample (*C. gigas*-1), as a control.

3.3. Run outputs

A total of 432,023,640 raw reads were produced during the NextSeq500 run, 39% of these reads corresponded to the two libraries prepared with the sewage sample. This high number of reads (167,864,533) obtained from the sewage sample was normalized with BBNorm which induced a loss of 96% of reads, leading to read counts comparable to the other samples. Quality trimming was applied to all samples and eliminated 2% of reads of the sewage sample, and between 1% and 3% of reads from oyster samples (including contaminated oysters and negative controls). The two water controls provided high number of reads presumably due to the additional amplification cycles performed during the library preparation to obtain the same amount of DNA. After trimming the number of reads, the seawater control was lowered by 15% while for the UltraPure water control, the loss represented around 6% of the initial reads.

3.4. Host abundance and viral reads

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

The highest amounts of reads attributed to viruses were obtained from the sewage libraries (69 and 63% out of total reads with I-HD and random hexamers, respectively (Table 1). Recovering such a huge amount of reads form one library was not expected, and draws attention on the preparation of mixes into different pools as it may favor one sample and overshadows the others. This also showed the importance of technical triplicates. A low number of mollusk reads was obtained from

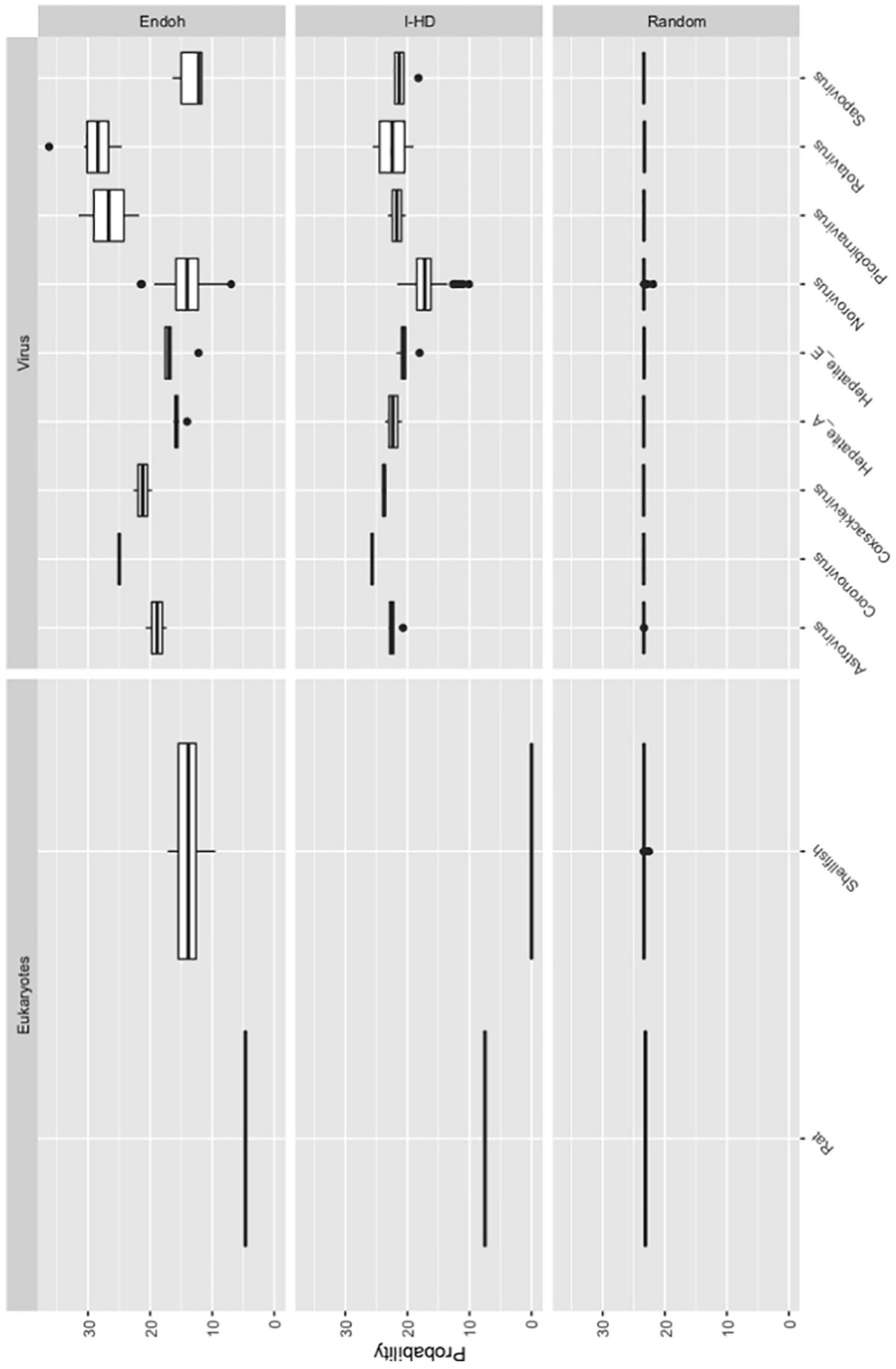

Fig. 2. Frequencies of priming sites in Eukaryote or virus genomes for three sets of hexamers.
The probability was calculated as the sum of probabilities of all non-ribosomal hexanucleotides in sense and complementary sequences to selected genomic sequences.

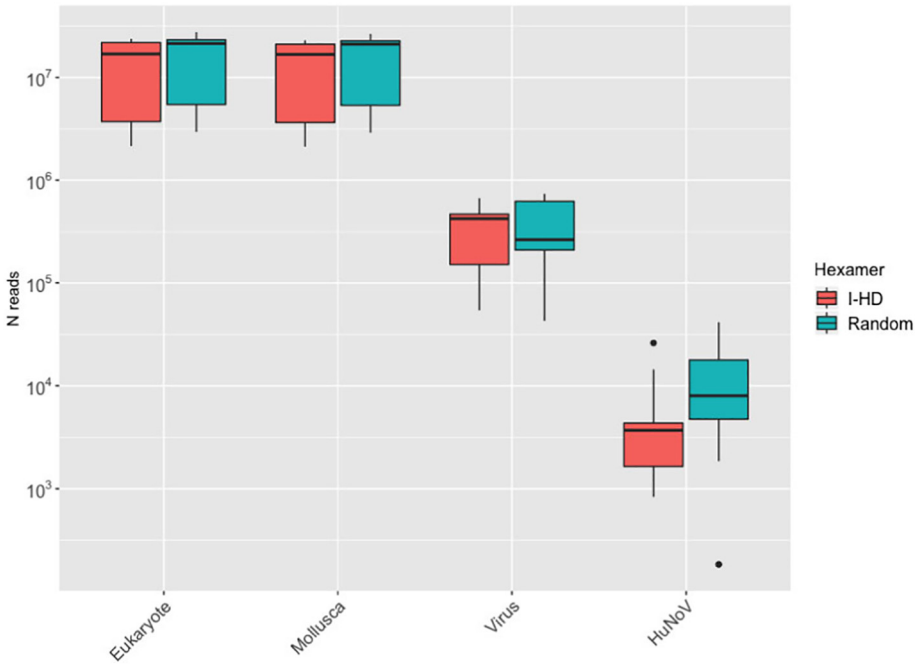

Fig. 3. Reads distribution obtained from bioaccumulated oyster libraries prepared using the I-HD (red) and random (green) sets. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

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	NoV GII
			Trimmed	Eukaryote	Mollusk	Virus	Reads	C_T values[a]
C. gigas-1	I-HD	1	4	3	3	0.05	1650	32.2 ± 0.2
		2	27	23	23	0.5	26,084	32.0 ± 0.3
		3	25	22	22	0.3	3020	32.2 ± 0.3
	Random	1	6	5	5	0.04	1854	30.6 ± 0.3
		2	30	27	26	0.2	4743	30.6 ± 0.2
		3	25	23	23	0.2	7299	30.7 ± 0.1
	Endoh	1	3	3	3	0.03	86	32.4 ± 0.6
		2	191	17	16	0.2	8887	32.2 ± 0.2
		3	30	28	27	0.2	5402	32.0 ± 0.5
C. gigas-2	I-HD	1	4	3	3	0.1	1601	32.7 ± 0.2
		2	17	15	14	0.4	833	32.5 ± 0.3
		3	25	22	21	0.4	4350	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	3955	34.8 ± 3.7
		2	22	17	17	0.7	14,454	34.8 ± 0.6
		3	22	18	18	0.5	3696	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	8011	30.3 ± 0.1
C. gigas-1 nc[b]	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[b]	I-HD	2	20	12	12	2	2	No C_T
	Random	2	18	11	11	1	0	No C_T
O. edulis nc[b]	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[c]	0.1	0.003	1	159,838	32.2 ± 0.4
	Random	1	6[c]	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

[a] NoV C_T values expressed as the mean value of three PCR replicates and the standard deviation.
[b] nc: negative control, I-HD: In House designed hexamers.
[c] Number of reads after normalization with bbnorm.

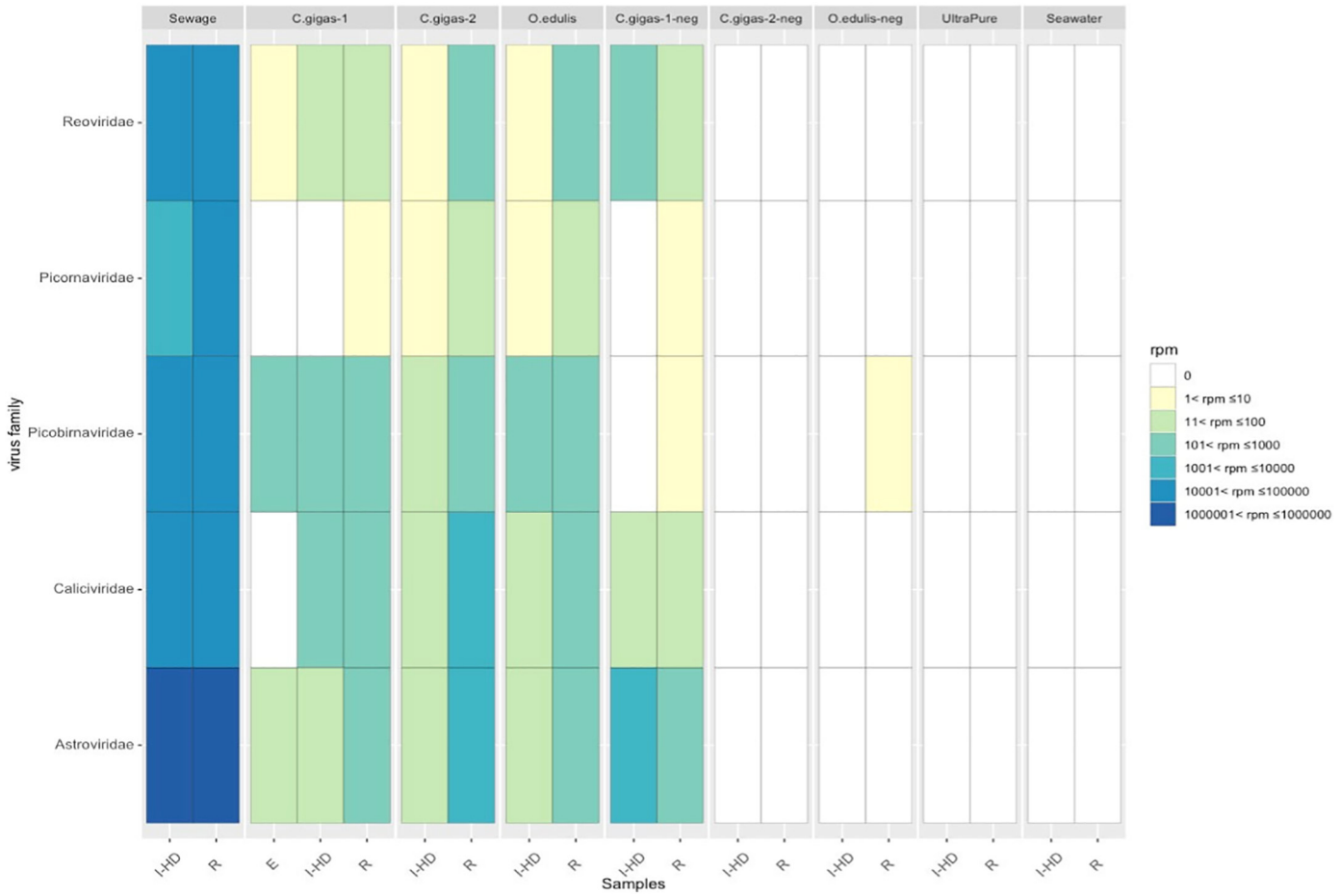

Fig. 4. Human enteric virus associated reads identified according to the different sets of hexamers used for the RT. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the seawater libraries (Table 1). This is natural seawater that has been in contact with some mollusks prior sampling. It was more surprising to obtain a high number of reads from the UltraPure water control. This was probably due to non-specific amplification which may have occurred during the additional amplification cycles performed to reach enough DNA concentration during the library preparation, or a tag jump, or due to a possible laboratory contamination during the preparation step.

3.5. Virus diversity in samples

Sequences of five viral families (*Reoviridae*, *Picornaviridae*, *Picobirnaviridae*, *Caliciviridae* and *Astroviridae)* were identified in the sewage sample and in the three bioaccumulated oyster batches (Fig. 4). In the sewage sample, the two hexamer sets gave similar numbers of reads, with the highest read numbers blasting to the *Astroviridae* family. The same five families were identified in the three oyster batches with the two hexamer sets. The *Picornaviridae* sequences were the less frequently identified sequences (ranging between 0 and 100 rpm) while *Astroviridae* and *Picobirnaviridae* families were the most represented (between 100 and 10,000 rpm). The library prepared using the Endoh set, provided the lowest read numbers for all five families. The random set provided the highest read numbers for all five families in agreement with total viral read numbers obtained (Fig. 4). The negative controls included in this study provided no significant read numbers for the five virus families except the *C. gigas*-1 negative control. This batch bought to a producer could have been contaminated in the producing area, or it could be a laboratory contamination during the sample preparation or the library preparation, as this negative control was in the pool with the most contaminated sample (sewage).

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

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

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

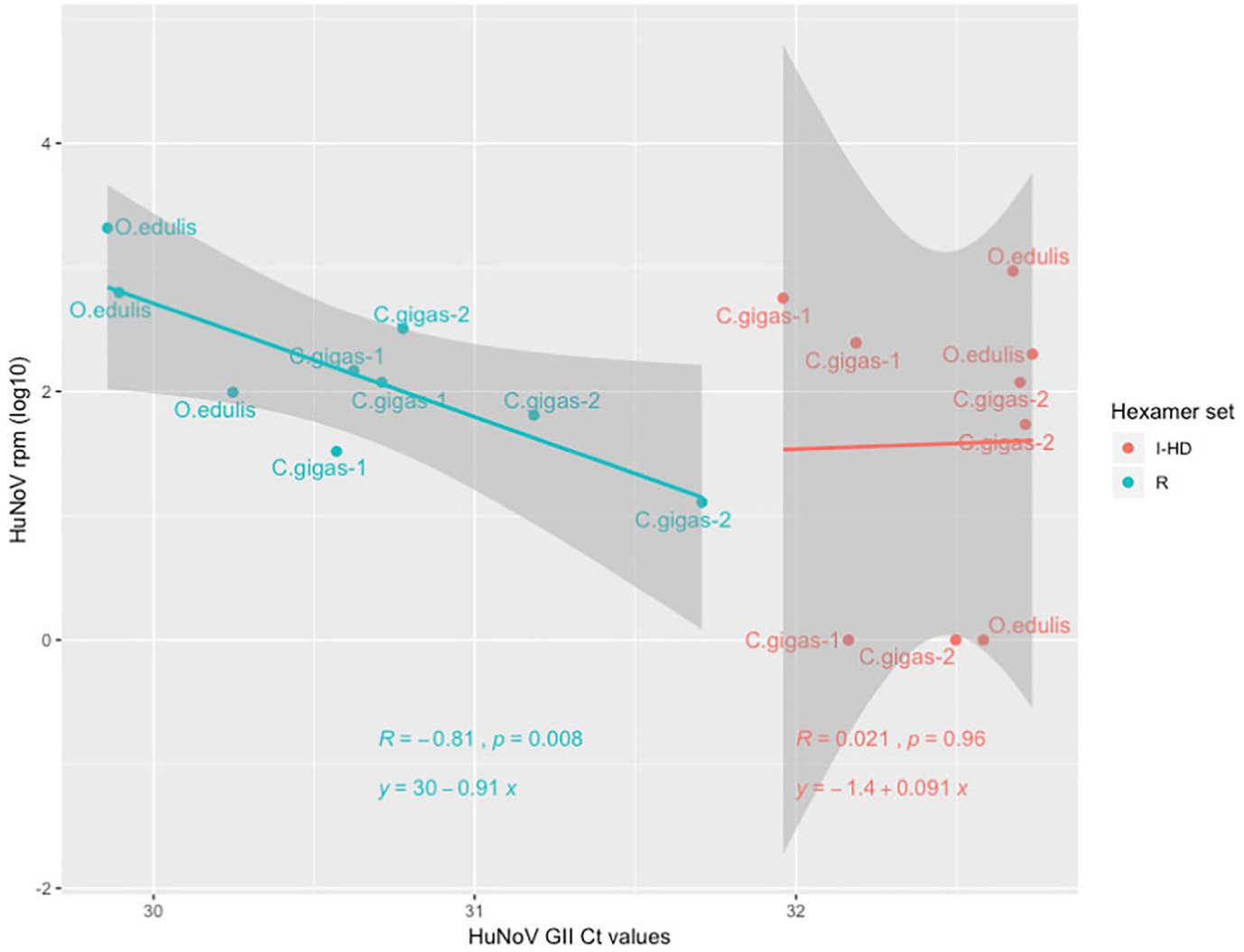

Fig. 5. Relationship between HuNoV concentrations and read expressed in reads per million (rpm), according to the hexamer sets used for cDNA.
Red dots: In-house designed (I-HD) set, blue dots: Random (R) hexamer set, The gray halo represents the range of variability between samples. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.7. Contigs length and HuNoV strain identifications

HuNoV contig length is an important criterion to select an effective method as longer contigs allow a better strain identification. Reads obtained using the Endoh set for cDNA priming, generated HuNoV contigs shorter than 500 nt, and thus were not considered for further analysis (Fig. 6). The random hexamers set allowed the recovery of the longest HuNoV contigs, for all samples except two contigs obtained from sewage sample. In this sample, the identified contig of 3211 nt obtained with I-HD was similar to a 2872 nt contig obtained with the random set. The longest HuNoV contigs were obtained from *O. edulis* libraries, which yielded two whole HuNoV genomes (Fig. 6) (Table 2). From oyster samples, the longest HuNoV contigs obtained using the I-HD set were of 1436, 899 and 1056 nt from *C. gigas*-1, *C. gigas*-2 and *O. edulis* batches respectively. Considering all samples, the I-HD set allowed the identification of 150 and 194 HuNoV GI and GII contigs, while the random hexamers set allowed the identification of 208 and 284 NoV GI and GII contigs. Indeed, when the contig lengths were too short (< 500 nt), strain identification was not possible, as observed with libraries built using Endoh set.

3.8. HuNoV strain diversity

Only contigs obtained from libraries prepared using the random hexamer set were considered for HuNoV genotyping using the online Norovirus typing tool 2.0 (Kroneman et al., 2011). For two oyster batches (C.gigas-2 and *O. edulis*), one library did not provide any contig meeting the criteria set up for the analysis (at least 500 nt length and blast identity score over 85%). For the same two batches, one of the three libraries provided few informative contigs that only allowed genogroup identification (Fig. 7). For the *O. edulis* batch, the contigs obtained for the third replicate were distributed among five HuNoV GII strains in the ORF1 and four in the ORF2 typing region and one GI (both ORF1 and 2). Two complete genomes (one GI and one GII) were identified in one library from the *O. edulis* batch and a third one was almost complete. In this same library, a sequence covering more than half of the genome was also identified (Table 2). For the C.gigas-2 batch most of the contigs obtained fell into a non-typing region of HuNoV GII ORF1 and only few contigs allowed the identification of a GII.[P16] and a GI.[P4] strains. Four sequences covering more than half of the genome were detected, three HuNoV GI and one GII strains. Five other strains longer than 1000 nt could also be identified either in the ORF1 or the ORF2 region (Table 2). Libraries obtained from the C.gigas-1 batch provided contigs for all three replicates. A GI.[P1] sequence was

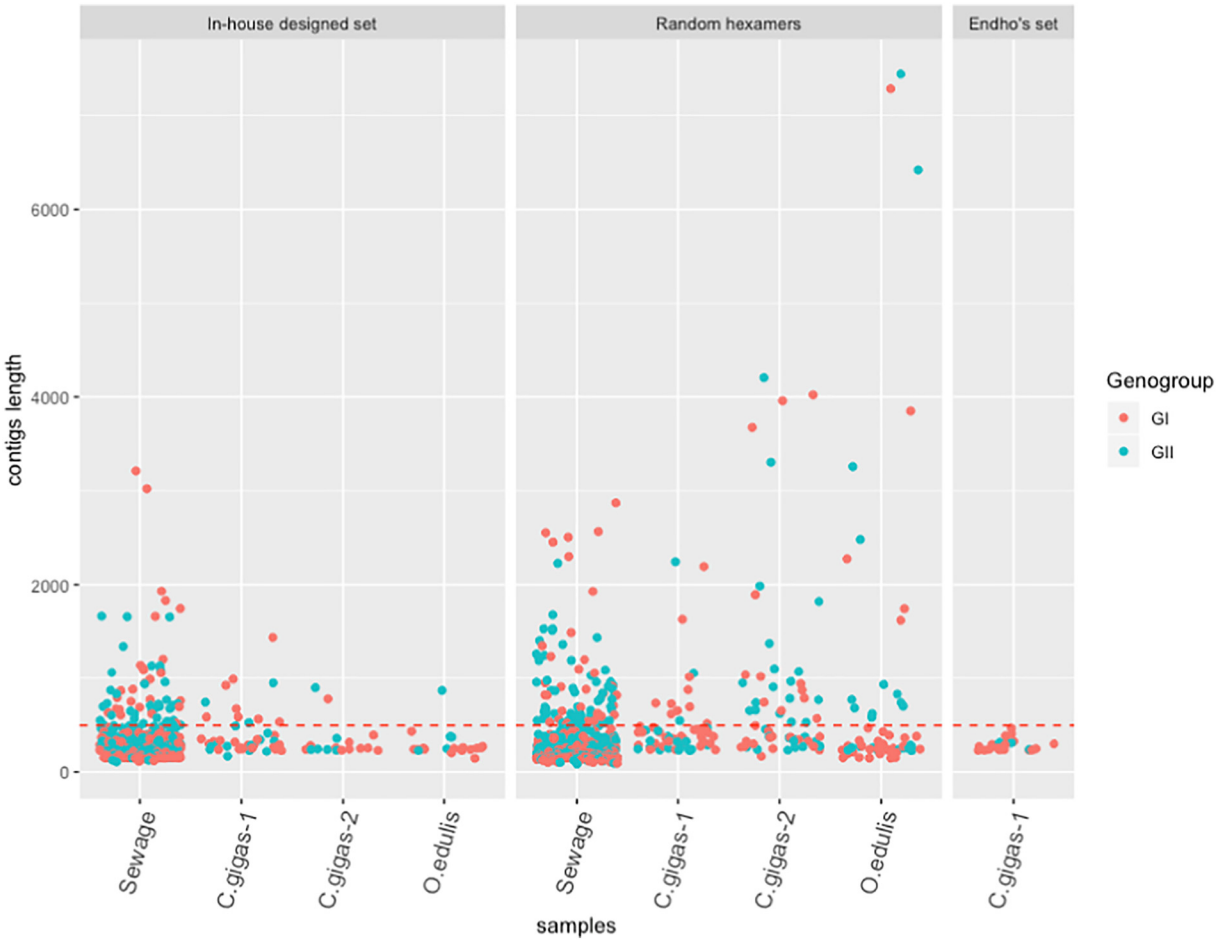

Fig. 6. Scatter plot of jittered data showing the length of contigs obtained using the three hexamer sets.
Genogroup I (GI) (red dots) and GII (blue dots) obtained for the sewage sample and bioaccumulated oyster batches are reported. The red dashed line represents the threshold contigs length fixed for data analysis (500 nt). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

observed in two replicates while the other one only provided genogroup information. GI.[P2], GII.[P16] and GII.[P17] strains were also identified in one of the replicates. For the ORF2 region, a GI.1sequence was identified in the same two replicates while the other one allowed only genogroup typing (Fig. 7).

In the sewage sample, contigs targeting non-typing region of the GII ORF1 and ORF2 were the most prevalent. Thirteen contigs were obtained from the sewage sample, but the longest was < 3000 nt and only two allowed genotyping of both the ORF1 and 2 (one for each genogroup) (Table 2).

All the strains detected in the sewage sample were detected at least once in one of the oyster batches, while the GI.2[P2] was detected only from oyster samples (full genome in *O. edulis*, and half genome in *C. gigas*-2). Considering the variability observed among the library triplicates, this approach is important to evaluate the diversity of viral strains that may be present in a sample. This may be of particular interest for low contaminated samples such as oyster samples.

4. Discussion

This study aimed to provide a method to identify HuNoV and other Human enteric viruses among the virome of oysters DT. While nowadays HuNoV constitute the known pathogen most frequently implicated in oyster related outbreaks, other viruses such as hepatitis A virus or rotavirus have also been responsible for clinical cases following raw oyster consumption, which highlights the interest of their detection (Bellou et al., 2013; Boxman et al., 2016). Oysters usually get contaminated by filtering seawater contaminated with human sewage, and thus accumulate a large diversity of human enteric viruses. Such diversity and the low copy number of the viral pathogens in contaminated oysters complicate their identification, and the containment and mitigation of foodborne outbreaks (McLeod et al., 2017). Metagenomic approaches, describing all nucleic acid present in a sample, appear as a promising tool to identify the different viruses that may be present without discriminating criteria.

The first challenge was to recover HuNoV from oyster tissues. HuNoV, and other enteric viruses, are mainly concentrated in oyster DT where they may bind specific ligands (such as histo-blood group antigen-like glycans, sialic acids) or be retained through non-specific ionic interactions (Di Girolamo et al., 1977; Le Guyader et al., 2012; McLeod et al., 2009). Targeting these tissues improve the method sensitivity and has now been validated worldwide (Atmar et al., 1995; *ISO 15216-1*, 2017; Lowther et al., 2019). We used an enzymatic treatment to elute viruses from the DT, similar to the ISO-15216-1 method) and we further improved the viral particle recovery using a sonication step and a chemical elution using sodium pyrophosphate before PEG concentration. Sonication was found useful for HuNoV cultivation in enteroids and it may help to disrupt vesicles in which HuNoV may be trapped (Ettayebi et al., 2016; Santiana et al., 2018). Alternatively, it may destroy bacteria to which the virus may bind (Miura et al., 2013). The combined used of sonication and sodium pyrophosphate was found the best method to recover viruses from sewage sludge (Brown et al., 2015).

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

Sample	Node	Reads	Length	ORF 1	ORF 2	Ref. sequence	Identity
Sewage	189	2993	2872	–	GI.1	MH638229.1	97.08
	237	4435	2564	GI.P1	–	MH638228.1	98.28
	294	5557	2299	–	GI.1	MH638229.1	99.76
	601	3380	1677	GII.P12	–	MG601447.1	98.20
	727	2939	1528	GII	–	MH218642.1	97.86
	832	1124	1435	GII	–	MF421538.1	99.03
	879	3050	1401	–	GII.3	MG601451.1	98.36
	971	5316	1346	GI.P4	GI.4	LN854563.1	97.70
	1128	2987	1244	GII	–	MH218642.1	98.12
	1147	1314	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	97.46
	38	241	2192	GI.P1	GI.4	LN854563.1	87.79
	156	1082	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	2341	4023	GI.P4	GI.4	LN854563.1	96.69
	17	1385	3960	GI.P1	GI.1	MH638228.1	98.84
	20	2296	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	2858	7445	GII.6[P7]		MH218642.1	98.06
	4	2261	7288	GI.2[P2]		LN854563.1	99.69
	5	1722	6422	GII.3[P12]		MG601447.1	97.88
	25	1594	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	1487	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.

Considering the limited knowledge on the association of HuNoV or other human enteric virus with bacteria, no filtration step was applied after the PEG concentration to prevent any loss of viral particles (Amarasiri and Sano, 2019, Conceicao-Neto et al., 2015).

Another challenge faced by metagenomics is the depletion of host or background nucleic acids that can account for up to 99% of the reads. This issue is encountered in clinical samples either for ethical issues or to improve the sensitivity (Giu et al., 2019, Oechslin et al., 2018,), as well as in food samples (Bartsch et al., 2018; Wylezich et al., 2018). The use of non-ribosomal hexanucleotides that favor RNA virus reverse transcription was found useful for species-independent detection of viruses in infected cells (Endoh et al., 2005). Following the same rational, we selected a combination of hexamers with the aim to reduce the binding to oyster and mussels rRNA. This indeed reduced the host background by about 10% compared to random hexamers. However, these I-HD hexamers were less successful to detect our target virus sequences than the random hexamer set, which may be due to the reduced number of sequences from HuNoV and other Human enteric viruses that were used to validate this selection. Thus, to increase the sensitivity and selectivity of our approach towards Human enteric viruses, the future development of this work will be to complete our hexamer selection by extending to more HuNoV strains including new variants, and to other Human enteric virus sequences. Indeed, beside the first step of sample preparation and nucleic acid purification this reverse transcription is the key step for a successful deep sequencing. The control performed by real-time PCR showed that depending on the hexamer set used, a difference of up to 3 C_T can be observed corresponding to a 10 times difference in HuNoV concentration. Such gain is important when looking for low contamination levels as usually found in marketed oysters or in oysters implicated in foodborne outbreaks (EFSA, 2019; McLeod et al., 2017; Polo et al., 2016). In all samples, the use of random hexamers led to a more efficient transcription of HuNoV sequences and to longer contigs allowing a better HuNoV strain identification compared to the two other sets. Indeed, norovirus genotyping requires contigs falling into specific regions of the genome at the end of ORF1 and the beginning of ORF2, which is less likely when reads are scarce and contigs, short. While the Endoh set was applied to only one sample, all contigs recovered were short (< 500 nt) and did not allow HuNoV identification. The I-HD set looks promising as shown by results obtained with the sewage sample, and the reduce mollusk background reads. To fully exploit the potential of metagenomics applied to low contaminated samples, quality controls and good laboratory practices are essential. RNA is sensitive to degradation and may be easily lost during the different steps of the protocol. However, the main challenge is to prevent contamination as already observed for many other molecular detection methods. Despite all precautions (dedicated rooms, filter tips…) we cannot exclude that identification of virus contigs in one of the negative control oyster batch and the UltraPure water resulted from contamination during the process, probably during the library preparation or the flow cell amplification (Schnell et al., 2015). The other negative controls (waters and oysters) presented few HuNoV reads that may be due to PCR errors occurring during the additional amplification cycles needed to reach the 10 ng/μL requested to satisfy the library kit instructions. The high number of reads recovered from the random-set primed library for the UltraPure water was unexpected, and among the virus reads no contig matching with Human enteric viruses was identified. None of the contigs detected in our negative controls meet our quality criteria of 500 nt with 85% Blast score (applied for all enteric viruses), and thus they were not considered for data analysis. Considering the huge sequencing capacity of high throughput metagenomic sequencing which is able to detect contaminant from reagents, from past or other experiments undertaken in the laboratory, the inclusion of negative controls that follow all the experiment process is essential (Brown et al., 2016; Nieuwenhuijse and Koopmans, 2017; Wylie et al., 2018).

In a previous study, a reference sample that we inserted in three separated runs provided similar HuNoV diversity, but with some variation in contigs lengths (Strubbia et al., 2019). In the present study, technical triplicates were used to evaluate the impact of library preparation. Such approach was found valuable to minimize potential bias in aquarium water virome analysis (Kim et al., 2017). Our study confirmed the interest of such replicates as NoV diversity varied from no contig to several sequences assigned to different genotypes, depending on the library considered for one given sample. Replicate libraries may be recommended in the future for low contaminated sample as it enhances the probability to identify rare sequences. When dealing with food samples this is important for further risk analysis. Finding two NoV sequences of < 200 base pairs in a sample clearly identified as being responsible of an outbreak may be interpreted as a positive argument to destroy the food in question, but it may also be a challenging issue for risk managers in absence of any epidemiological data (Bartsch et al., 2018). Providing solid arguments of repeatable detection in technical or biological replicates may help to increase confidence in this highly sensitive method.

Another valuable option to increase method sensitivity towards Human enteric viruses is to enrich the nucleic acid extract using a targeted sequence capture (Brown et al., 2016; Wylie et al., 2018). As

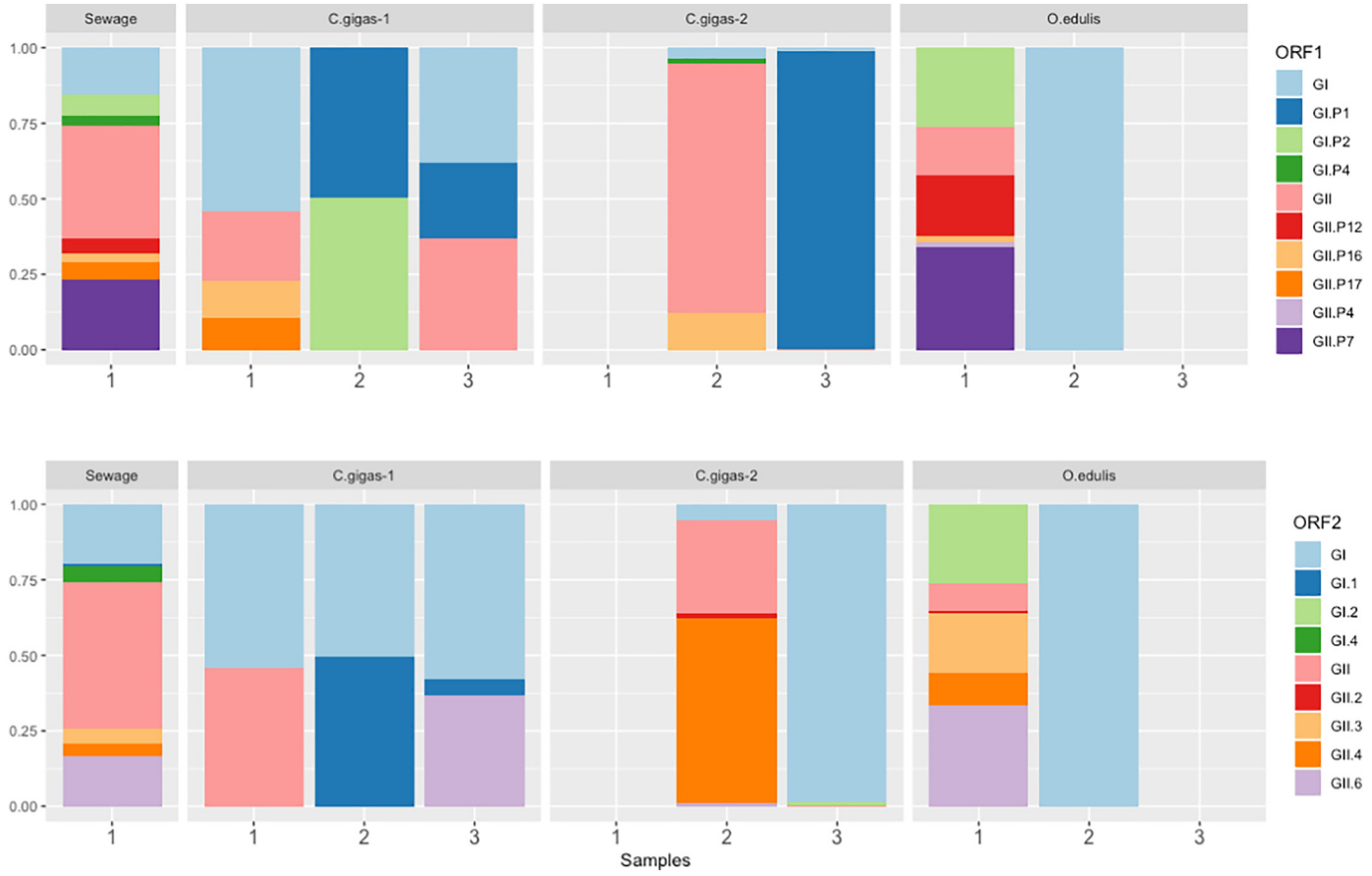

Fig. 7. HuNoV ORF1 and ORF2 identified in sewage and oyster samples using random hexamers during library preparation.
The three technical replicates are reported for each oyster batch, with numbers 1,2, 3 corresponding to the pool. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

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

In this study, two full HuNoV genome sequences were identified in one of the oyster sample contaminated with a thousand copies of HuNoV GI or GII per g of DT. These promising results demonstrate the possibilities to improve the method sensitivity and its applicability to even less contaminated oyster samples. Considering the variability of HuNoV and their high frequency of genome recombination, it is important to achieve the recovery of long contigs that allows strain identification both in the ORF1 and ORF2 typing regions (Chhabra et al., 2019). HuNoV from different genotypes differ in their ability to cause disease thus a proper identification will help risk manager to take adequate decisions, and for this it is crucial to include adequate controls to be confident in results obtained (Atmar et al., 2018; Parra et al., 2017). For food outbreak analysis, the strain identification in the two ORFs will be the only suitable way to identify transmission pathways (Nieuwenhuijse and Koopmans, 2017; Nooij et al., 2018). Over the three oyster batches, 28 contigs longer than 1000 nt were obtained, half of them providing the ORF1-ORF2 junction region, showing that the method proposed here may help to investigate shellfish-borne outbreaks.

In human or in environmental matrices, the microbiome may play a role for HuNoV selection or persistence (Amarasiri and Sano, 2019, Monedero et al., 2018, Walker and Baldridge, 2019). In this study we compared two batches of *Crassostrea gigas* oysters: one that has never been exposed to human fecal contamination and one bought from a local producer. This producer was selected in a clean area to avoid pre-existing HuNoV or Human enteric virus contamination, but still some differences were observed in bacteria concentrations. Nonetheless, no difference either in terms of HuNoV bioaccumulation or in terms of

strains identification was found between the two batches. No difference either was observed with *Ostrea edulis*, except that this batch was the one that provided the longest contigs for HuNoV identification.

5. Conclusions

Environmental samples such as oyster tissues are complex matrix generally characterized by low virus concentration. We show here that our metagenomics approach allows the sequencing of NoV full genomes from contaminated oysters samples, and can be further optimized. Combined with the development of infectivity test to characterize infectious viruses will constitute a step forward for shellfish contamination analysis (Ettayebi et al., 2016). The scientific interest on developing an approach able to reduce the host background is high and need to be further investigated. An optimization of the hexamer set we designed can be considered extending the number and the diversity of the reference sequences used for hexamer selection. Furthermore, the use of triplicate libraries is recommended to increase the chances to detect viruses present in low concentrations. Reliable data are also ensured by the use of negative controls which should always be included in a metagenomic run.

Data availability

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

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We are grateful to the sewage treatment plant staff for the sewage sample, and to our colleagues from the Marine Mollusks Experimental platform (PMM Bouin, Ifremer) for providing the two batches of oysters. This study was supported by the European Community's Horizon 2020 - Research and Innovation Framework Programme under the COMPARE project, grant agreement N°643476. SS was supported by a grant from the IFREMER Direction Scientifique. This work benefited from equipment and services from iGenSeq core facility, at ICM (Institut Cerveau Moelle, Paris, France).

References

AFNOR, 2010. NF V08-106. Dénombrement des *Escherichia coli* dans les coquillages vivants - Technique indirecte par impédancemétrie directe.
Amarasiri, M., Sano, D., 2019. Specific interactions between human norovirus and environmental matrices: effects on the virus ecology. Viruses 11, 224.
Atmar, R.L., Neill, 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. Appl. Environ. Microbiol. 61, 3014–3018.
Atmar, R.L., Ramani, S., Estes, M.K., 2018. Human noroviruses: recent advances in a 50-year history. Curr. Opin. Infect. Dis. 31, 422–432.
Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikiv, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Pribelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comp. Biol. 19, 455–477.
Bartsch, C., Hiper, D., Made, D., Johne, 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.
Bellou, M., Kokkinos, P., Vantarakis, A., 2013. Shellfish-borne viral outbreaks: a systematic review. Food Environ. Virol. 5, 13–23.
Bisseux, M., Colombet, J., Mirand, A., Roque-Alfonso, A.-M., Abranavel, F., Izopet, J., Archimbaud, C., Peigue-Lafeuille, H., Debroas, D., Bailly, J.-L., Henquell, C., 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. Euro Surveill 23 (7).
Bosch, A., Gkogka, E., Le Guyader, F.S., Loisy-Hamon, F., Lee, A., van Lieshout, L., Marthi, B., Myrmel, M., Sansom, A., Schultz, A.-C., Winkler, A., Zuber, S., Phister, T., 2018. Foodborne viruses: detection, risk assessment, and control options in food processing. Int. J. Food Microbiol. 285, 110–128.
Boxman, I., Verhoef, L., Vennema, H., Ngui, S., Friesema, I.H., Whiteside, C., Lees, D., Koopmans, M., 2016. International linkage of two food-borne hepatitis A clusters through traceback of mussels, the Netherlands, 2012. Euro Surveill 21 (3).
Brown, M.R., Camezuli, S., Davenport, R.J., Petelenz-Kurdziel, E., Ovreas, L., Curtis, T.P., 2015. Flow cytometric quantification of viruses in activated sludge. Wat. Res. 68, 414–422.
Brown, J.R., Roy, S., Ruis, C., Romero, E.Y., Shah, D., Willimas, R., Breuer, J., 2016. Norovirus whole-genome sequencing by sure select target enrichment: a robust and sensitive method. J. Virol. 54, 2530–2537.
Bushnell, B., Rood, J., Singer, E., 2017. BBMerge-accurate paired shotgun read merging via overlap. PLoS One 12, e0185056.
Chhabra, P., de Graaf, M., Parra, G.I., Chan, M.C.-W., Green, K., Martella, V., Wang, Q., White, P.A., Katayama, K., Vennema, H., Koopmans, M.P.G., Vinje, J., 2019. Updated classification of norovirus genogroups and genotypes. J. Gen. Virol. online first. https://doi.org/10.10999/jgv.0.001318.
Conceicao-Neto, N., Zeller, M., Lefrere, H., de Bruyn, P., Beller, L., Deboutte, W., Yinda, C.K., Lavigne, R., Maes, P., Van Ranst, M., Heylen, E., Matthijnssens, J., 2015. Molecular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis. Sci. Rep. 5, 16532.
Cotten, M., Munnink, B.O., Canuti, M., Deijs, M., Watson, S.J., Kellam, P., van der Hoek, L., 2014a. Full genome virus detection in fecal samples using sensitive nucleic acid preparartion, deep sequencing, and a novel iterative sequence classification algorithm. PLoS One 9, e93269.
Cotten, M., Petrova, V., Phan, M.V.T., Rabaa, M., Watson, S.J., Ong, S.H., Kellam, P., 2014b. Deep sequencing of norovirus genomes defines evolutionary patterns in an urban tropical setting. J. Virol. 88, 11056–11069.
Di Girolamo, R., Liston, J., Matches, J., 1977. Ionic binding, the mechanism of viral uptake by shellfish mucus. Appl. Environ. Microbiol. 33, 19–25.
EFSA, 2019. Analysis of the European baseline survey of norovirus in oysters. EFSA J. 17, 5762.
Ettayebi, K., Crawford, S.E., Murakami, K., Broughman, J.R., Karandikar, U., Tenge, V.R., Neill, F.H., Blutt, S.E., Zeng, X.-L., Qu, L., Kou, B., Opekum, A.R., Burrin, D., Graham, D.Y., Ramani, S., Atmar, R.L., Estes, M.K., 2016. Replication of human noroviruses in stem cell-derived human enteroids. Science 353, 1387–1393.
Giu, W., Miller, S., Chiu, C.Y., 2019. Clinical metagenomic next-generation sequencing for pathogen detection. Annu. Rev. Pathol. Mech. Dis. 14, 317–336.
Glockner, F.O., Yilmaz, P., Quast, C., Gerken, J., Beccati, A., Ciuprina, A., Bruns, G., Yarza, P., Peplies, J., Westram, R., Ludwig, W., 2017. 25 years of serving the community with ribosomal RNA gene reference databases and tools. J. Biotechnol. 261, 169–176.
ISO 15216-1, 2017. Microbiology of Food and Animal Feed - Horizontal Method for Determination of Hepatitis A Virus and Norovirus in Food Using Real-time RT-PCR, Part 1: Method for Quantification.
Kallies, R., Holzer, M., Toscan, R.B., da Rocha, U.N., Anders, J., Marz, M., Chatzinotas, A., 2019. Evaluation of sequencing library preparation protocols for viral metagenomic analysis from pristine asquifer groundwaters. Viruses 11, 484.
Kim, Y., van Bonn, W., Aw, T.G., Rose, J.B., 2017. Aquarium viromes: viromes of human-managed aquatic systems. Front. Microbiol. 8, 1231.
Kittigul, L., Panjangampatthana, A., Rupprom, K., Pombubpa, K., 2014. Genetic diversity of rotavirus strains circulating in environmental water and bivalve shellfish in Thailand. Int. J. Environ. Res. Public Health 11, 1299–1311.
Kroneman, A., Vennema, H., Deforche, K., Avoort, H.V.D., Penaranda, S., 2011. An automated genotyping tool for enteroviruses and noroviruses. J. Clin. Virol. 51, 121–125.
Langmead, B., Salzberg, S., 2012. Fast gapped-read aligment with Bowtie 2. Nat. Methods 9, 357–359.
Le Guyader, F.S., Le Saux, J.-C., Ambert-Balay, K., Krol, J., Serais, O., Parnaudeau, S., Giraudon, H., Delmas, G., Pommepuy, M., Pothier, P., Atmar, R.L., 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.
Le Guyader, F., Parnaudeau, S., Schaeffer, J., Bosch, A., Loisy, F., Pommepuy, M., Atmar, R.L., 2009. Detection and quantification of noroviruses in shellfish. Appl. Environ. Microbiol. 74, 618–624.
Le Guyader, F.S., Atmar, R.L., Le Pendu, J., 2012. Transmission of viruses through shellfish: when specific ligands come into play. Cur. Op. Virol. 2, 103–110.
Li, M., Baker, C.A., Danyluk, M.D., Belanger, Ph., Boelaert, F., Cressey, P., Gheorghe, M., Polkinghorne, B., Toyofuku, H., Havelaar, A.H., 2018. Identification of biological hazards in produce consumed in industrialized countries: a review. J. Food Prot. 81, 1171–1186.
Lowther, J.A., Bosch, A., Butot, S., Ollivier, J., Mäde, D., Rutjes, S.A., Hardouin, G., Lombard, B., in't Veld, P., Leclercq, A., 2019. Validation of EN ISO method 15216 - part1 - quantification of hepatitis A virus and norovirus in food matrices. Int. J. Food Microbiol. 288, 82–90.
McLeod, C., Hay, B., Grant, C., Greening, G., Day, D., 2009. Localization of norovirus and poliovirus in Pacific oysters. J. Appl. Microbiol. 106, 1220–1230.
McLeod, C., Polo, D., Le Saux, J.-C., Le Guyader, F.S., 2017. Depuration and relaying: a

review on potential removal of norovirus from oysters. Comp. Rev. Food Sci. Food Saf. 16, 692–706.
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. Ann. Rev. Microbiol. 49, 461–487.
Miura, T., Sano, D., Suenaga, A., Yoshimura, T., Fusawa, M., Nakagomi, T., Nakagomi, O., Okabe, S., 2013. Histo-blood group antigen-like substances of human enteric bacteria as specific adsorbents for human noroviruses. J. Virol. 87, 9441–9451.
Monedero, V., Buesa, J., Rodriguez-Diaz, J., 2018. The interactions between host glycobiology, bacterial microbiota, and viruses in the gut. Viruses 10, 96.
Nieuwenhuijse, D., Koopmans, M.P.G., 2017. Metagenomic sequencing for surveillance of food- and waterborne viral diseases. Front. Microbiol. 8, 230.
Nooij, S., Schmitz, D., Vennema, H., Kroneman, A., Koopmans, M.P.G., 2018. Overview of virus metagenomic classification methods and their biological applications. Front. Microbiol. 9, 749.
Oechslin, C.P., Lenz, N., Liechti, N., Ryter, S., Agyeman, P., Bruggmann, R., Leib, S.L., Beuret, C.M., 2018. Limited correlation of shotgun metagenomics following host depletion and routine diagnostics for viruses and bacteria in low concentrated surrogate and clinical samples. Front. Cel. Inf. Microbiol. 8, 375.
Endoh, D., Mizutani, T., Kirisawa, R., Maki, Y., Saito, H., Kon, Y., Morikawa, S., Hayashi, M., 2005. Species-independent detection of RNA virus by representational difference analysis using non-ribosomal hexanucleotides for reverse transcription. Nucl. Ac. Res. 33, 1–11.
Oude Munnink, B.B., van der Hoek, L., 2016. Viruses causing gastroenteritis: the known, the new and those beyong. Viruses 8, 42.
Parra, G.I., Squires, R.B., Karangwa, C.K., Johnson, J.A., Lepore, C.J., Sosnovtev, S.V., Green, K.Y., 2017. Static and evolving norovirus genotypes: implications for epidemiology and immunity. PLoS Pathog. 13, e1006136.
Polo, D., Schaeffer, J., Fournet, N., Le Saux, J.-C., Parnaudeau, S., McLeod, C., Le Guyader, F.S., 2016. Digital PCR for quantifying norovirus in oysters implicated in outbreaks, France. Emer. Inf. Dis. 22, 2189–2191.
Polo, D., Schaeffer, J., Teunis, P., Buchet, V., Le Guyader, F.S., 2018. Infectivity and RNA persistence of a norovirus surrogate, the Tulane virus, in oysters. Front. Microbiol. 9, 716.
Rosani, U., Shapiro, M., Venier, P., Allam, B., 2019. A needle in a haystack: tracing bivalve-associated viruses in high-throughput transcriptomic data. Viruses 11, 205.
Sano, D., Amarasiri, M., Hata, A., Watanabe, T., Katayama, H., 2016. Risk managment of viral infectious diseases in watewater reclamation. Environ. Int. 91, 220–229.
Santiana, M., Ghosh, S., Ho, B.A., Rajasekaran, V., Du, W.-L., Mutsafi, Y., De Jesus-Diaz, D.A., Sosnovtsev, S.V., Levenson, E.A., Parra, G.I., Takvorian, P.M., Cali, A., Bleck, C., Vlasova, A.N., saif, L.J., Patton, J.T., Lopalco, P., Corcelli, A., green, K.Y., Altan-Bonnet, N., 2018. Vesicle-cloaked virus clusters are optimal units for inter-organismal viral transmission. Cell Host Microbe 24, 208–220.
Schnell, I.B., Bohmann, K., Gilbert, T.P., 2015. Tag jumps illumintaed-reduing sequence-to-sample misindentifications in metabarcoding studies. Mol. Ecology Res. 15, 1289–1303.
Strubbia, S., Schaeffer, J., Oude Munnink, B.B., Besnard, A., Phan, M.V.T., Nieuwenhuijse, D.F., de Graaf, M., Schapendonk, C.M.E., Wacrenier, C., Cotton, M., Koopmans, M., Le Guyader, F.S., 2019. Metavirome sequencing to evaluate norovirus diversity in sewage and realated bioaccumulated oysters. Front. Microbiol. 10, 2394.
Venugopal, V., Gopakumar, K., 2018. Shellfish: nutritive value, health benefits, and consumer safety. Comp. Rev. Food Sci. Food Saf. 16, 1219–1242.
Walker, F.C., Baldridge, M.T., 2019. Interaction between noroviruses, the host, and the microbiota. Cur. Op. Virol. 37, 1–9.
Weiner, R.M., Segall, A.M., Colwell, R.R., 1985. Characterization of a marine bacterium associated with *Crassostrea virginica* (the eastern oyster). Appl. Environ. Microbiol. 49, 83–90.
Wylezich, C., Papa, A., Beer, M., Hoper, D., 2018. A versatile sample processing workflow for metagenomic pathogen detection. Nature Scien. Rep. 8, 13108.
Wylie, T.N., Wylie, K.M., Herter, B.N., Storch, G.A., 2015. Enhanced virome sequencing using targeted sequence capture. Genome Res. 25, 1910–1920.
Wylie, K.M., Wylie, T., Buller, R., Herter, B., Cannella, M.T., Storch, G.A., 2018. Detection of viruses in clinical samples by use of metagenomic sequencing and targeted sequence capture. J. Clin. Microbiol. 56, e01123-01118.
Yu, Y., Cai, H., Hu, L., Lei, R., Pan, Y., Yan, S., Wang, Y., 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.