OXFORD

BIOLOGY Methods & Protocols

Biology Methods and Protocols, 2024, bpae053

https://doi.org/10.1093/biomethods/bpae053
Advance Access Publication Date: 27 July 2024
Methods Article

Assessing RNA integrity by digital RT-PCR: Influence of extraction, storage, and matrices

Sebastien Wurtzer [1,*], Mathilde Duvivier[1], Heberte Accrombessi[1], Morgane Levert[1,2], Elise Richard[1] Laurent Moulin[1,3]

[1]Research & Development Department, Eau de Paris. DRDQE, FR-9400, France
[2]Paris Sorbonne Universite, CNRS, EPHE, UMR 7619 Metis, e-LTER Zone Atelier Seine, F-75005, Paris, France
[3]Obepine SIG, Paris, FR-75000, France

*Correspondence address. Research & Development Department, 33 avenue Jean Jaurès, FR-94200 Ivry-sur-Seine, France. Tel: +33 145154226; E-mail: sebastien.wurtzer@eaudeparis.fr

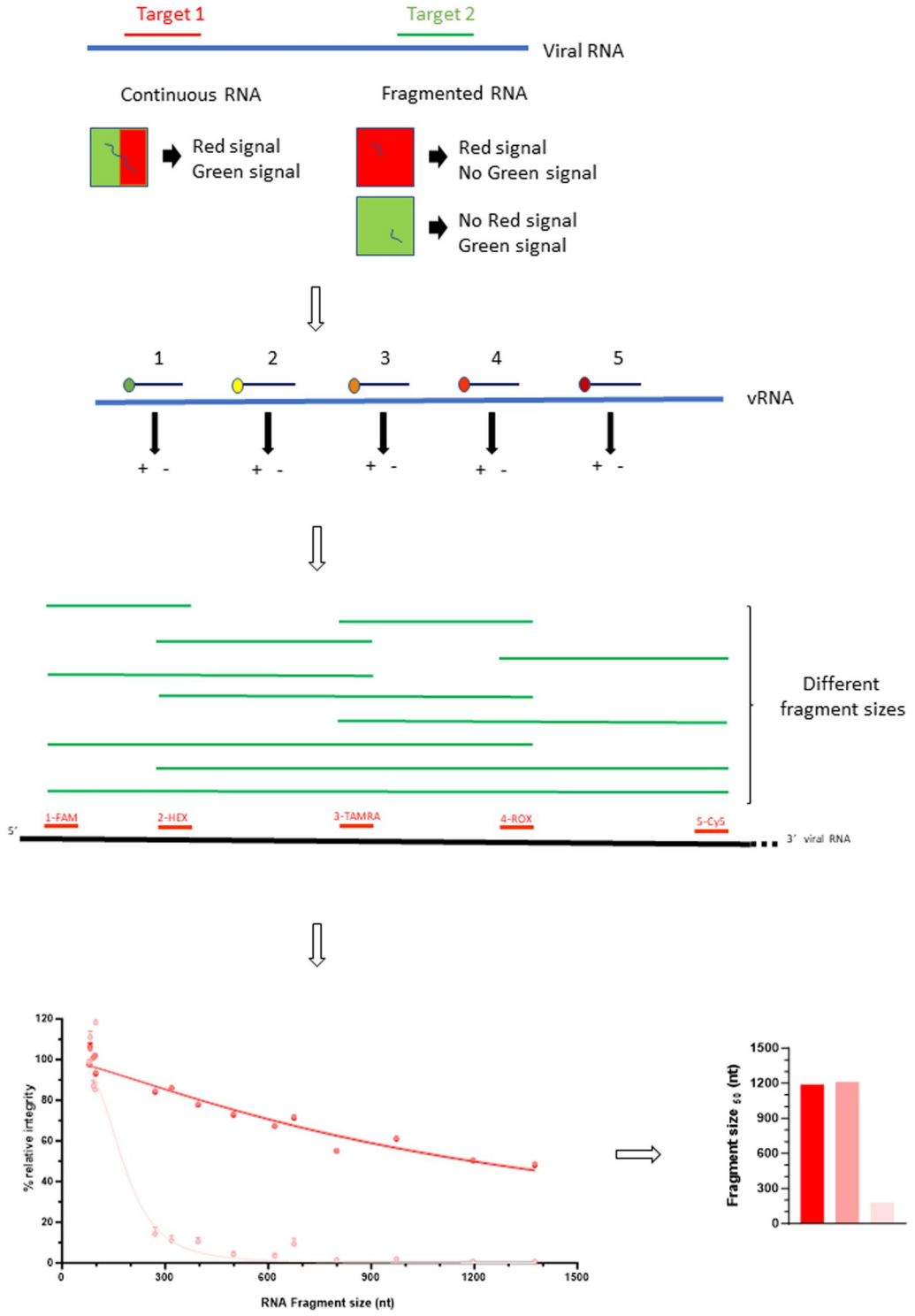
Abstract

The development of high-throughput sequencing has greatly improved our knowledge of microbial diversity in aquatic environments and its evolution in highly diverse ecosystems. Relevant microbial diversity description based on high-throughput sequencing relies on the good quality of the nucleic acid recovered. Indeed, long genetic fragments are more informative for identifying mutation combinations that characterize variants or species in complex samples. This study describes a new analytical method based on digital Polymerase Chain Reaction (PCR) partitioning technology for assessing the fragmentation of nucleic acid and more specifically viral RNA. This method allows us to overcome limits associated with hydrolysis probe-based assay by focusing on the distance between different amplicons, and not, as usual, on the size of amplicons. RNA integrity can thus be determined as a new fragmentation index, the so-called Fragment size 50. The application of this method has provided information on issues that are inherent in environmental analyses, such as the storage impact of raw samples or extracted RNA, extraction methods, and the nature of the sample on the integrity of viral RNA. Finally, the estimation of fragment size by digital PCR (dPCR) showed a very strong similarity with the fragment size sequenced using Oxford Nanopore Technology. In addition to enabling objective improvements in analytical methods, this approach could become a systematic quality control prior to any long-read sequencing, avoiding insufficiently productive sequencing runs or biases in the representativeness of sequenced fragments.

Keywords: RNA integrity; fragmentation; wastewater; digital PCR; long-read sequencing

Received: 19 January 2024. **Revised:** 15 July 2024. **Editorial decision:** 16 July 2024. **Accepted:** 26 July 2024

Graphical Abstract



Introduction

Genomic analysis methods for environmental monitoring, especially in wastewater have gained considerable interest in recent years. Targeted molecular approaches, mainly based on quantitative PCR or digital PCR, enable the detection and quantification of microorganism genomes excreted from biological fluids of infected individuals (stool, urine, or saliva) and finally discharged into the sewage system. Collective epidemiological surveillance has demonstrated its relevance in tracking the global epidemic dynamic through its various waves of infection [1, 2], anticipating viral resurgence at an early stage [3], assessing the repercussions of certain health policies such as confinements or curfews [4], and even dating the introduction of emerging viruses into a territory as well as their spread through the population [5, 6]. Moreover, most RNA viruses quickly evolve, producing a diversity of quasi-species and emerging variants [7, 8]. The monitoring of

known or cryptic variants is an epidemiological surveillance strategy that is complementary to clinical genomic surveillance [9, 10]. Such monitoring conducted using PCR-based methods enables the tracking of temporal trends in global virus circulation, and eventually some known variants [1, 2]. However, the great viral diversity and the redundancy of mutations shared by most of the circulating variants make it challenging to estimate the prevalence and geospatial distribution of these lineages. This in turn limits our ability to grasp the variant spreading in the population with targeted molecular tools.

Genomic monitoring in wastewater by sequencing is a relevant, but particularly complex, strategy for tracing variants. Indeed, wastewater sample, unlike clinical samples, contains a multitude of viral genomes. Analysis of the Influenza Segment 1 consensus sequence was enough for determining subtypes [11]. For severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), some studies attempted to analyze consensus sequences without being able to reform contigs. Lineage assignment with tools, such as pangolin, designed for clinical samples containing mainly one dominant variant, do not allow to estimate the relative abundances of multiple SARS-CoV-2 lineages [12, 13]. Others analyzed the relative abundance of each mutation and associated certain mutations with variants, as with targeted PCR detection, but with a greater diversity of targets [14]. Some recent promising studies of raw data analysis proposed analytical workflows based on deconvolution algorithms to estimate relative abundances within a complex mixture [15, 16].

Most whole-genome sequencing in wastewater is based on multiplex overlapping short amplicons [17, 18]. This strategy, user-friendly, implied fragmentation of genomic information. In samples comprising several different variants, as in wastewater, this greatly complicates downstream bioinformatics analysis and the information that can be extracted from the sequencing data [19]. In this context, long-read sequencing enables the sequencing of much longer DNA fragments than traditional short-read sequencing techniques and offers several advantages. It allows to improve the detection of complex structural variants by spanning large structural variations, such as insertions, deletions, inversions, and translocations, which are often missed or incorrectly assembled with short reads. Long reads can also resolve complex regions of the genome, including repetitive sequences, facilitating more accurate *de novo* genome assembly [20]. Long reads can finally provide more comprehensive and accurate information on microbial communities, by improving taxonomic resolution to species and genotype levels [21, 22]. Furthermore, due to the complex composition of environmental samples, a large proportion of the virus particles are partially degraded by the time they are detected [23]. It is difficult to say whether this degradation occurs in the matrix itself or whether it stems from the methods used in the analytical workflow (i.e. concentration, extraction, or even storage). This degradation affects the viral particle structure designed to protect the genome but could also have an impact on nucleic acids. Such fragmentation of the genetic material could allow genomic quantification through amplification of short fragments by RT-qPCR or RT-dPCR but would be more problematic for sequencing of viral genome long fragments.

There is a lack of tools available for the assessment of nucleic acid integrity, especially in the case of RNA. Whether the approach is qualitative (e.g. denaturing agarose gel) or quantitative (e.g. microfluidic-based device using capillary gel electrophoresis) [24], RNA integrity is estimated based on a ratio between eukaryotic or prokaryotic ribosomal RNAs [25]. These methods also require large quantities of biological material. These tools are therefore not suitable for the specific estimation of viral RNA quality, particularly in highly complex environmental samples (low concentration of the targeted nucleic acid, a mix of several microorganisms, etc.). Few recent studies have taken advantage of the multiplexing capabilities of qPCR or dPCR to estimate DNA integrity. DNA fragmentation was assessed by using a single forward primer and several reverse primers, thus generating fragments of varying sizes. DNA integrity was calculated as a comparison between longer and shorter DNA fragments [26, 27]. Another study sought to identify tumor markers by simultaneously detecting amplicons of different sizes within the same region [28] or different regions of the genome [29].

The aim of this work was to develop a molecular tool, based on digital PCR technology, for assessment of targeted nucleic acid integrity. The proof-of-concept was achieved by targeting the viral RNA of coxsackievirus B3, a choice based especially on several factors. First, it is a nonenveloped enteric virus involved in gastroenteritis, which makes it a suitable candidate for analysis in wastewater samples. Second, it replicates rapidly in cell culture. Third, it produces large quantities of genetic material, which is beneficial for the sensitivity of the assay. It is therefore particularly well-suited to, and resistant in, an environment such as wastewater. RNA integrity was then analyzed to identify limiting steps, especially by comparing nucleic acid extraction techniques, assessing the effect of the wastewater matrix, and the impact of sample storage.

Materiel and methods

Virus stock preparation

Coxsackievirus B3, strain Nancy (CV-B3, #ATCC VR-30) was cultivated on confluent monolayer cultures of Buffalo Green Monkey kidney (BGMK, #ATCC PTA-4594) cells. Cells were grown and infected in Dulbecco's Modified Eagle's Medium (DMEM) high glucose (DUTSCHER, FRANCE), supplemented with 2% fetal bovine serum (DUTSCHER, FRANCE), and nonessential amino acids (DUTSCHER, FRANCE). The supernatant was collected 48 h post-infection and clarified by centrifugation at 2,000*g* for 15 min.

Wastewater sampling

The effect of the water matrix on the quality of the extracted viral RNA was tested. While there are probably many causes of viral RNA degradation in an environmental water sample, raw wastewater is one of the matrices with the shortest viral persistence. The presence of many micro-organisms encourages predation and the presence of numerous chemical pollutants such as detergents is likely to affect the sample stability. Fresh raw wastewaters were sampled from the sewers of the city of Paris. Among the various samples collected, two were selected based on different turbidity values measured using the TU5200 instrument (HACH, FRANCE) expressed in the Nephelometric Turbidity unit (NTU). Raw wastewater samples were spiked using coxsackievirus B3 supernatant and incubated for 24 h with slow stirring to allow interactions with the organic matter.

Viral RNA extraction

Viral RNA was extracted and purified using two different methods. The first method, hereafter referred to as the manual method, was based on the QIAamp viral RNA mini kit (#52904, QIAGEN, FRANCE) according to the manufacturer's recommendations. The second one, referred to as the automated method, used the QIAsymphony DSP Virus/Pathogen Kit (QIAGEN, France)

on the QIAsymphony extractor (QIAGEN) according to the manufacturer's recommendations. All experimentations were done with four independent replicates. For all the conditions tested (type of matrix, extraction method, or storage conditions), only 200 µl of raw sample were analyzed to avoid a potential impact of the sample concentration processes on the viral RNA integrity and possible inhibitions as well. These aspects were not analyzed in this study.

Molecular amplification by RT-dPCR

A nonoverlapping 5-plex assay was designed using AlleleID software version 7 (Premier Biosoft) to amplify five different regions of CV-B3. Primers and probes were reported in Table 1. Briefly, CV-B3 genomes were quantified by digital one-step RT-PCR using QIAcuity instrument (QIAGEN, FRANCE). Multiplex reactions were performed using the QIAcuity One-Step Advanced Viral RT-PCR mastermix (QIAGEN, FRANCE) with 400 nM of each primer and 200 nM of each probe in 8.5k—96-well nanoplates. The thermal profile included a reverse transcription done at 50°C for 40 min, followed by activation step at 95°C for 2 min. Then amplification was done by 45 cycles of incubation at 95°C for 15 s and 58°C for 1 min. The threshold was automatically set. Quantification results were exported in .csv format using the current results tab, whereas raw partition distribution was exported in .csv format using the multiple occupancy tab.

Determination of RI

The amplification results were detection (+) or absence (–) for each amplicon within each partition. Multiple occupancy files reported the number of partitions for each combination of amplicons. In the 5-plex assay, there were 2^5 possible combinations of results. Some of these corresponded to continuous detection of adjacent fragments, and thus to potential non-fragmented regions. Combinations showing discontinuous fragment detection corresponded to random amplicon colocalization events within the same partition. Thus, the average number of positive partitions resulting from random colocalization events (referred to as rdAPP, for random discontinuous average positive partitions) had to be estimated. This random event occurrence was to be estimated according to whether there were 2 ($rdAPP_2$), 3 ($rdAPP_3$), or 4 ($rdAPP_4$) amplicons within a partition. For example, $rdAPP_2$ was the sum of positive partitions for two amplicons, resulting from colocalization events of discontinuous amplification products, divided by the number of combinations including these two amplicons and resulting from random distribution.

$$\mathrm{rdAPP} = \frac{\sum \mathrm{rdPP}}{rC} \quad (1)$$

"rdPP" (random discontinuous positive partitions) was the positive partitions corresponding to discontinuous detection and then random colocalization of amplicons, and "rC" (random combination) were the combinations of discontinuous amplicon detection.

Genome positive partitions (gPP) corresponded to the average of the total positive partitions (tPP) per amplicon. We assumed that the minimum fragment size corresponded to the average length of the amplicon.

$$\mathrm{gPP} = \frac{\sum_{i=1}^{N} tPPi}{N} \quad (2)$$

"tPP" was the total positive partitions for each amplified region, and 'N' was the number of amplicons determined by a fluorescent dye, that is the number of amplified regions.

Relative integrity (RI) of each fragment was estimated by the ratio of the positive partitions of a continuous detection combination, from which the average number of positive partitions resulting from random colocalization was deduced, to the maximum RI. 'fPP' (fragment positive partitions) was the positive partitions corresponding to a fragment size. rdPP deduced ($rdAPP_2$, $rdAPP_3$, or $rdAPP_4$) depended on the positivity combination (2, 3, or 4 amplicons). For example, if a fragment is determined by the continuous positivity of three amplicons, rdPP calculated for three amplicons were deduced.

$$\mathrm{RI} = \frac{\sum(\mathrm{fPP}) - \mathrm{rdAPP}}{\mathrm{gPP}} \quad (3)$$

All calculations were performed using Microsoft Excel software.

Estimation of fragment size 50 index

Based on the RI of each fragment estimated in four independent replicates, a decay curve of RI as a function of RNA fragment size was estimated using a variable-slope dose–response model.

$$Y = 100/\left(1 + (FS_{50}/X)^{Slope}\right) \quad (4)$$

Y was the RI; X was RNA fragment size, and "FS_{50}" was Fragment $size_{50}$, that is the estimated median fragment size of the distribution.

Table 1. List of oligonucleotides of the different molecular assays

Position	Name	Forward primer sequence	Position	Name	Reverse primer sequence	Name	Probe sequence	Dye labeling	Quencher
486	CVB3_set1_F	ATGCGATCACATCATCAC	565	CVB3_set1_R	GACACTAATGCTAAAAGTGTAC	CVB3_set1_P	ATGCGATCATCCTGAATTGGTCCA	FAM	BHQ-1
889	CVB3_set2_F	GTGGAAAGTTTTGTACAGATTA	985	CVB3_set2_R	GGTAGCGAAAGGAAAGTC	CVB3_set2_P	TGTCTGTCTCATCGCCACTGAATC	HEX	BHQ-1
1780	CVB3_set3_F	GTCTCTGAACTTCTCATTCC	1862	CVB3_set3_R	ATGCCAAGGAGCTAGTAG	CVB3_set3_P	CAGTTCTAAGTTGGTGCCGTCCT	TAMRA	BHQ-2
666	CVB3_set9_F	GATGGGAGTTGCACAAGTAGTC	757	CVB3_set9_R	TGTCTGGTTTGCTTGCCTAAA	CVB3_set9_P	TGTGCGTGTATCCAAGCTTCTCAAGT	Cy5	QXL 670
1187	CVB3_set10_F	CGTCCTCCAGTTTCATTGGT	1285	CVB3_set10_R	CACACACGTGGATGAGTACAT	CVB3_set10_P	TGATATCTAGGGTGGCTAGTTGGCCT	ROX	QXL 610

Viral RNA sequencing

First-strand cDNA synthesis from RNA was carried out using SuperScript IV VILO Master Mix (ThermoFisher Scientific, France) including random primers. RNA from four replicates have been pooled, and libraries have been prepared using Ligation sequencing gDNA—Native Barcoding kit V14 (Oxford Nanopore, England) according to manufacturer recommendations. Sequencing was performed on a MinION Mk1c device using the R10.4.1 flow cell (Oxford Nanopore). Deconvolution of barcoded sequencing data was done by MinKNOW. Total reads were merged in a single fastq file per barcode using "concatenate multiple datasets" tool (Galaxy version 1.1.4). Viral reads were isolated from concatenated reads by mapping against CV-B3 ATCC VR 30 reference sequence using Bowtie 2 (Galaxy version 2.5.3). Aligned reads were analyzed using NanoPlot (Galaxy version 1.42.0), and viral reads were counted according to 50 bp windows using Filter sequences by length (Galaxy version 1.2). Reads smaller than 50 bp were not considered.

Statistical analysis and modeling of RI decay curves were modeled using GraphPad Prism v10.1 software.

Results

Principle, proof-of-concept, and experimental validation

Principle

dPCR is a technology in which a PCR reaction mixture is split and evenly distributed into thousands of nanoliter reactors containing none, one or a few copies of target DNA. Each partition is analyzed after end-point PCR amplification to determine the presence or absence of fluorescence. Then, a statistical estimation of target molecules present in the sample is calculated. This approach, which is increasingly used for quantitative environmental monitoring, offers further prospects (Fig. 1A).

In multiplexed reactions, different segments of the viral genome are amplified independently. The various nucleic acid fragments are randomly distributed in the different partitions. If the amplicons are located on different fragments, this will lead to the generation of positive partitions for one or more of the amplicons (Fig. 1B). On the other hand, if the amplicons correspond to regions located on the same strand, then the fluorescent signals generated by these amplicons should colocalize within the same partition. Some of the partitions generating signals from adjacent regions may also result from random colocalization of independent fragments (Fig. 1C). The dPCR principle is based on limit dilution to have as many unique events per partition as possible, thus limiting the random colocalization events of distinct segments amplified within the same partition. Combinations of simultaneous amplification of adjacent regions, also called continuous detection, provide information on fragment size whose integrity is preserved (Fig. 1D). On the opposite, combinations of simultaneous amplification of nonadjacent regions (or discontinuous detection) provide information on the proportion of random colocalization events. The minimum detectable fragment size is that of a single amplicon and corresponds to maximum measurable integrity. The ratio of positive partitions relative to fragment size to tPP of single amplicon determines a RI for each fragment size.

Proof-of-concept using double-stranded DNA

To carry out the proof-of-concept of this methodology, 5-plex molecular assay, consisting of five nonoverlapping targets specific to coxsackievirus B3 strain Nancy, was designed (Fig. 2A). The assay consisted of five amplifications spread over 1376 nt of viral RNA. This assay generated combinations of amplifications covering RNA lengths ranging from 271 (i.e. the two closest targets) to 1376 nt (e.g. the two farthest targets) with a median length of all PCR combinations of 647 nt) (Fig. 2B). In parallel, a 2500-bp double-stranded DNA (between positions 261 and 2760 of the genome sequence), produced by *in vitro* synthesis, was inserted into the pMA-RQ cloning vector (Fig. 3A).

As differences in PCR efficiency between the different amplifications could induce significant biases in the assessment of RI, viral genome quantification efficiency was systematically assessed for the different amplified regions. The different parts of the viral genome were similarly quantified across the genome. These results underlined the similar amplification efficiencies independently of the amplified genome target (Fig. 3B). The RI estimated for each genome length showed values close to 100% regardless of the number of plasmid copies used per reaction (Fig. 3C).

To validate the proof-of-concept, this plasmid was fragmented using restriction endonucleases. Fragmentation of the plasmid by BamHI, with two cut sites at positions 377 and 1457, generated two fragments of 1080 bp (containing the amplicons 1, 2, 9, and 10) and 3871 bp (containing only the amplicon 3). Double enzymatic digestion with BamHI and SacI, the latter also having two cut sites at positions 114 and 1574, generated four fragments: 117 bp (not detectable as containing no amplicon), 343 bp (containing the amplicon 10), 737 bp (containing the amplicons 1, 2, and 9), and 3754 bp (containing only the amplicon 3) (Fig. 3A).

Quantification of the various genomic fragments showed comparable values for each amplicon, irrespective of the fragmentation state of the genomic target, including in circular form (Supplementary Fig. S1). Using this assay, 93% of valid partitions were positive for all five amplicons (1, 2, 3, 9, and 10) in the digested plasmid. For the plasmid cut by BamHI, 46% of partitions were positive for amplicon 3 only, and 46% for amplicons 1, 2, 9, and 10 within the same partitions. Finally, for the doubly digested plasmid, 33% of valid partitions were positive for amplicon 3 only, 30% for amplicon 10 only, and 33% for amplicons 1, 2, and 9. The proportions of positive partitions were in line with the number of detectable fragments and the positioning of the various amplicons on this genomic target (Fig. 3D). A schematic diagram of fragment distribution and associated fluorescence illustrates the raw results obtained (Supplementary Fig. S2).

Experimental validation on virus RNA

Impact of RNA concentration on the RNA RI estimation

A combination of amplicons within the same partition indicates either the presence of a preserved RNA fragment serving as a template for these amplifications or the presence of several distinct fragments within the same partition. In the latter case, the higher the RNA concentration, the greater the probability that several distinct amplicons colocalized within the same partition, leading to overestimation of the RI. CV-B3 was replicated on BGMK cells. Total nucleic acids from a fresh viral production were extracted using an automated method and four independent replicates were serially diluted. Each dilution was quantified using RT-dPCR. Quantification was linear within the range of dilutions 1/200 and 1/102 400 and was comparable between the different amplicons for each dilution. The 100-fold dilutions showed too few negative partitions to correctly quantify and were considered out of range (Supplementary Fig. S3).

The RI of RNA fragments was estimated based on this molecular assay. As expected, the lower the RNA concentration, the lower the estimated RI, highlighting RI overestimation resulting from

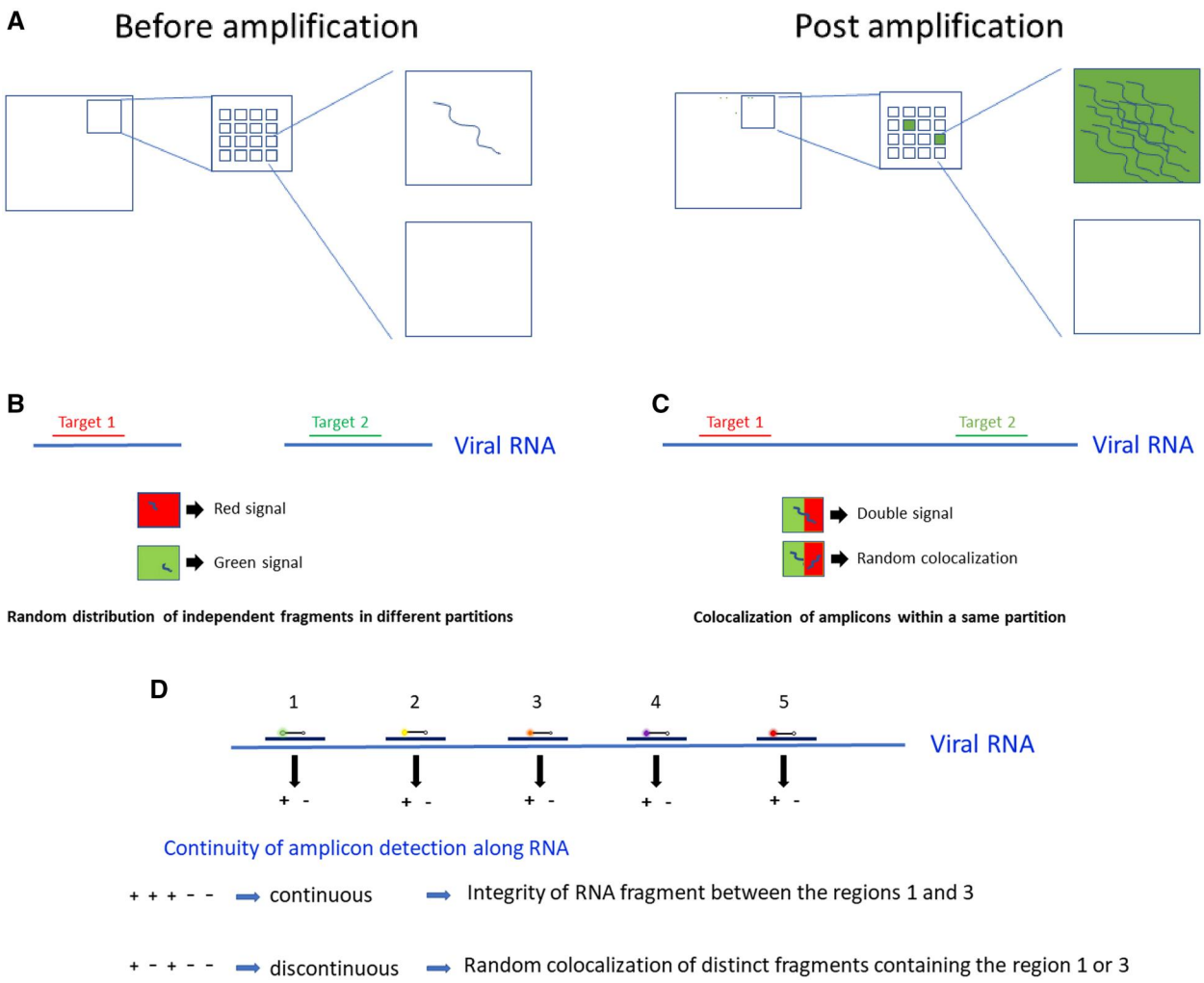


Figure 1. Principle of RNA integrity assessment. (A) RT-dPCR enables the amplification of RNA fragments isolated within partitions. The resulting quantification is based on statistical counting according to Poisson distribution. (B) In the case where independent RNA fragments are amplified in the same multiplex reaction, random distribution within different partitions would lead to positive partitions for one of the amplicons only. (C) If the amplified regions are located on the same RNA fragment, then the partitions containing this fragment would be double positive. Some of the double-positive partitions could also result from random colocalization of independent fragments within the same partition. In this case, the proportion of random events in the totality of double-positive partitions has to be estimated. (D) In a 5-plex reaction, there are 25 combinations of amplicons. When detection is continuous along the RNA fragment within the same partition, it is therefore likely that this partition contains a unique, conserved fragment of the region covered by these amplicons. Conversely, when detection is discontinuous along the RNA genome within the same partition, this corresponds to a random colocalization of two distinct fragments. The various continuous combinations enable variable RNA lengths to be determined. Random colocalization events are necessarily subtracted to determine the most likely number of partitions that can contain a preserved fragment. The number of positive partitions for each continuous combination, compared with the total number of RNA copies, provides information on the RI of each RNA fragment size. Modeling the decay curve of RI as a function of RNA fragment length enables to estimate a new fragmentation index called FS50, corresponding to the minimum size of 50% of the target RNA fragments in the sample.

random colocalization. However, from dilution 1/6400 onwards, the estimate of RI remained constant up to dilution 1/102 400. By drawing the integrity decay curve as a function of RNA fragment size, we were able to define a Fragment size$_{50}$ (FS_{50}). This new fragmentation index corresponded to the median size of 50% of RNA fragments. On these RNA extracts, FS_{50} reached a stable median value (383 ± 31 nt) from dilution 1/6400, suggesting a negligible impact of events resulting from random colocalization. Based on RNA quantifications, the RNA quantity range for correct RI estimate was between 561 and 26 copies/µl (Fig. 4).

Comparison of RNA fragment size estimation and sequencing read size distribution

RNA extracted from new fresh viral supernatant using an automated method was reverse-transcribed, and cDNA was sequenced using Oxford Nanopore Technology (ONT). The distribution of viral read size (Fig. 5A), mapping on CV-B3 sequence, was compared to the estimation of RNA fragment sizes provided by RT-dPCR (Fig. 5B). Genomic coverage was 100% with 28× depth. The median size of viral sequencing reads was 428 bp, while FS_{50} was 416 nt. The proportion of RNA fragment sizes and the fragmentation index were estimated using the present dPCR-based tool.

Application of viral RNA integrity assessment for environmental purposes

Different viral productions were carried out to question what could impact the integrity of the purified viral RNA genome, especially in environmental analysis context.

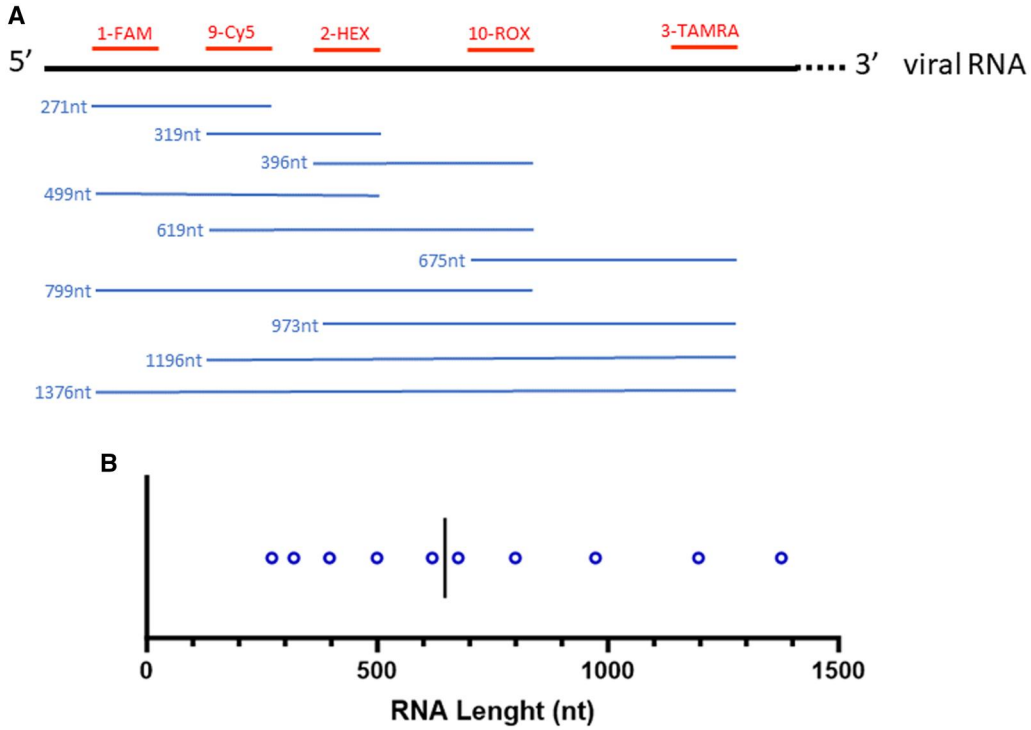


Figure 2. Design of nonoverlapping 5-plex assay and RNA fragment lengths expected. (A) 5-plex molecular assay has been designed based on coxsackievirus B3 (#ATCC-VR-30) sequence. The assay included amplicons noted 1, 2, 3, 9, and 10. The lines, below the viral RNA, represent the minimum sizes of regions of the viral genome that can be estimated by the different combinations of continuous detection of amplicons within the same partition. (B) Distribution of RNA fragment lengths. The assay generated combinations of amplifications covering RNA lengths ranging from 271 to 1376 nt, centered on a median length of 647 nt. Each circle represents a minimum genome length detectable by a combination of continuous detection of amplicons within the same partition and indicated in Fig. 2A. The horizontal bar represents the median value of this distribution.

Impact of nucleic acid extraction method on viral RNA integrity

Based on extraction methods available in the laboratory, a comparison was made between a kit dedicated to the extraction of viral RNA using a silica column (hereinafter referred to as the manual method) and an automated method for the extraction of total nucleic acids using magnetic beads (subsequently called automated method). A new viral production was carried out and the nucleic acids were extracted from viral supernatant using both methods. Quantification of viral RNA showed a significantly better yield with the manual method than with the automated one (Mann–Whitney test, $P = 0.0079$). Average concentrations were 384 ± 29 copies/µl for the manual method and 111 ± 13 copies/µl for the automated method considering all genome regions (Supplementary Fig. S4).

The RI of RNA extracted with the manual method was very close to that of RNA from automated method. FS_{50} was 431 nt for manually extracted RNA and 416 nt for RNA automatically extracted (Fig. 6A). FS_{50} values were compared to the median size of sequencing reads and showed very close results to the experimental error [median sequencing read size: 480 bp, 95% IC (453–507) for manual extraction; 478 bp, 95% IC (433–533) for automated extraction].

Impact of raw sample storage on viral RNA integrity

Sample storage conditions, or even pretreatment to ensure innocuity of hazardous samples to handle, are a key point in wastewater-based epidemiology. Fresh viral supernatant was stored for 4 h at 4°C, frozen at −80°C or subjected to heat-inactivation usually used to inactivate micro-organisms (60°C for 30 min). Viral RNA was extracted using both methods. The results obtained using the manual extraction are presented (Fig. 6B). Average RNA concentrations were 384 ± 29 copies/µl, 420 ± 29 copies/µl, and 13 ± 2 copies/µl for samples stored at 4°C, those thawed at −80°C, and those heat-inactivated, respectively (Supplementary Fig. S5). While heat exposure induced a significant loss of viral RNA detectable (Kruskal–Wallis test, $P = 0.026$) and fragmentation of viral RNA (FS_{50} = 186 nt), in contrast to freeze–thawed samples (FS_{50} = 412 nt) and those stored at 4°C (FS_{50} = 431 nt) (Fig. 6B). The results agreed with median sequencing read sizes [(4°C): 480 bp, 95% IC (453–507); (−80°C): 450 bp, 95% IC (405–506)]. Results from heat-inactivated samples were indicated but not relevant with regard to too few viral reads (Supplementary Fig. S6) and then large 95% IC range [341 bp, 95% IC (206–863)]. Results from samples extracted using automated method are presented in Supplementary Figs S5, S7, and S8.

Impact of extracted RNA storage conditions on RNA integrity

As molecular analyses of environmental samples are frequently retrospectively carried out, the storage of extracted nucleic acids is also essential for wastewater-based epidemiology. Fresh viral supernatant replicates were extracted using both methods, then immediately amplified or stored for 1 week at −20°C prior to amplification. Quantification of extracted viral RNA was not significantly affected by RNA freezing–thawing (Supplementary Fig. S9). However, RNA freezing–thawing resulted in moderate reduction of RI (FS_{50}= 431 nt and FS_{50}= 312 nt from fresh and frozen RNA extracted with manual kit, respectively), confirmed by sequencing [median sequencing reads: 480 bp, 95% IC (421–445) from fresh RNA and 390 bp, 95% IC (367–415) from frozen RNA]

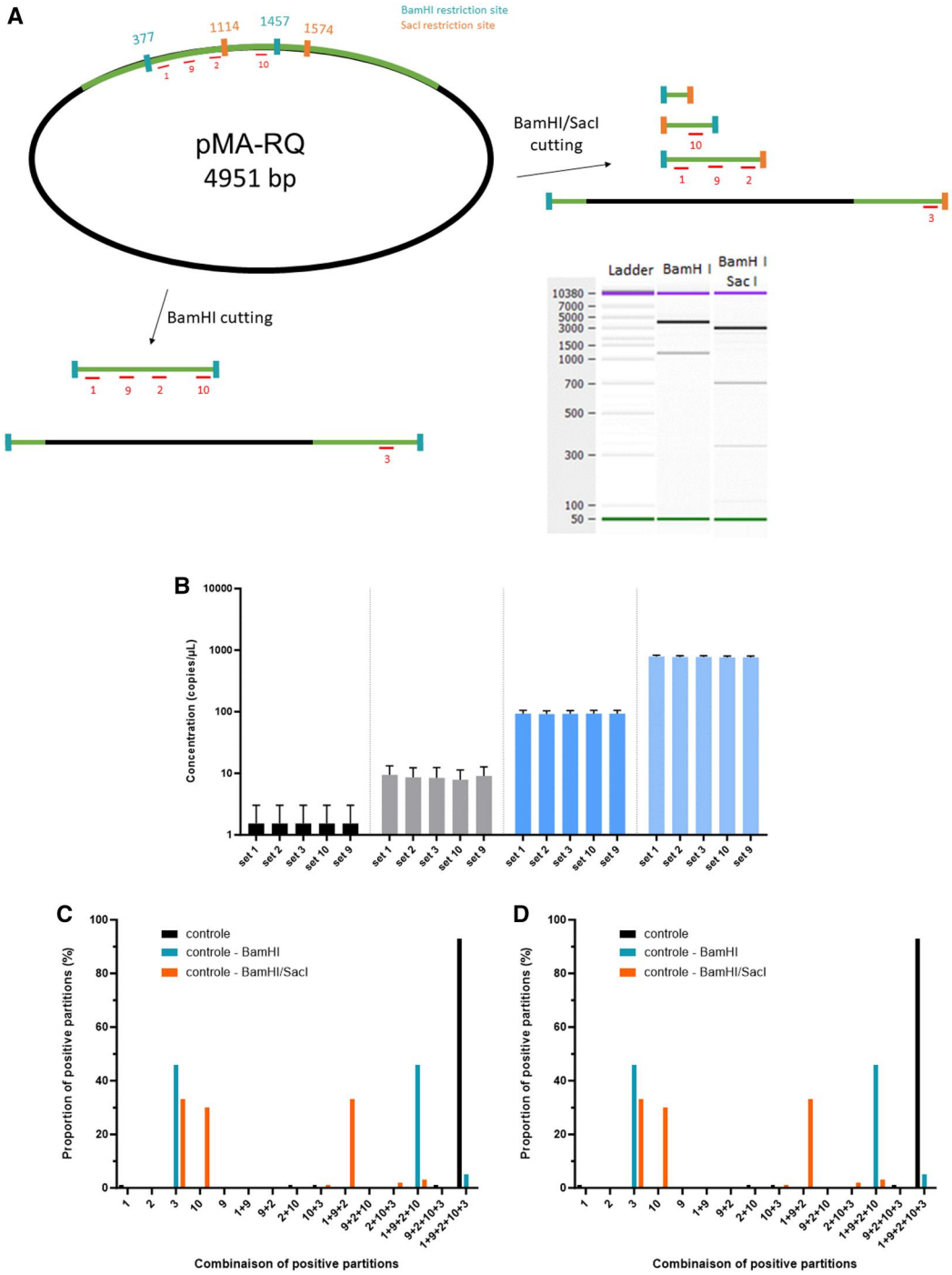


Figure 3. Validation of the proof-of-concept. (A) Restriction map of control plasmid. The latter was constructed by cloning of 2500 bp of coxsackievirus B3 sequence into pMA-RQ vector. This control has been fragmented using restriction endonucleases (BamHI only or double digestion BamHI/SacI). The positions of amplicon were indicated on each fragment, and enzymatic restriction was verified by capillary electrophoresis. BamHI restriction site is reported in positions 377 and 1457, and SacI restriction site in positions 1114 and 1574. The numbered lines indicates the position of the different amplicons. (B) Quantification of cloned target using the assay. All the amplicons were equally quantified for each dilution, confirming similar PCR efficiencies for the different amplified targets. The different colors represent 10 times serial dilutions of the plasmid ranging between 1000 copies/µl and 1 copies/µl of reaction. (C) RI in function of fragment size for the plasmid. The dots corresponded to the RI of each fragment identified by the continuous detection of amplicon combinations. (D) Molecular detection of each fragment within partitions. Linearized plasmid was amplified with the molecular assay. The proportion of positive partitions among all positive partitions was reported according to the different combinations of amplicons within the same partition. The bars of the chart indicated the proportion of positive partitions according to whether these partitions were positive for an amplicon only (i.e. a single region of the genome) or for a combination of amplicons (several regions of the genome). According to Fig. 3A, when the plasmid was fragmented by BamHI only, two fragments were detected. One fragment contained amplicons 1, 2, 9, and 10, while the second was positive only for amplicon 3. As the enzymatic digestion resulted in two fragments, each fragment represented approximately 50% of the total detectable fragments. In the case of a double digestion by BamHI and SacI, four fragments were generated. Three of the four fragments were detectable. One fragment tested positive for the amplicons 1, 2, and 9, a second fragment was positive only for the amplicon 3, and the third detectable fragment was positive for the amplicon 10 only. As the double enzymatic digestion resulted in three different detectable fragments, each fragment accounted for a third of all detectable fragments.

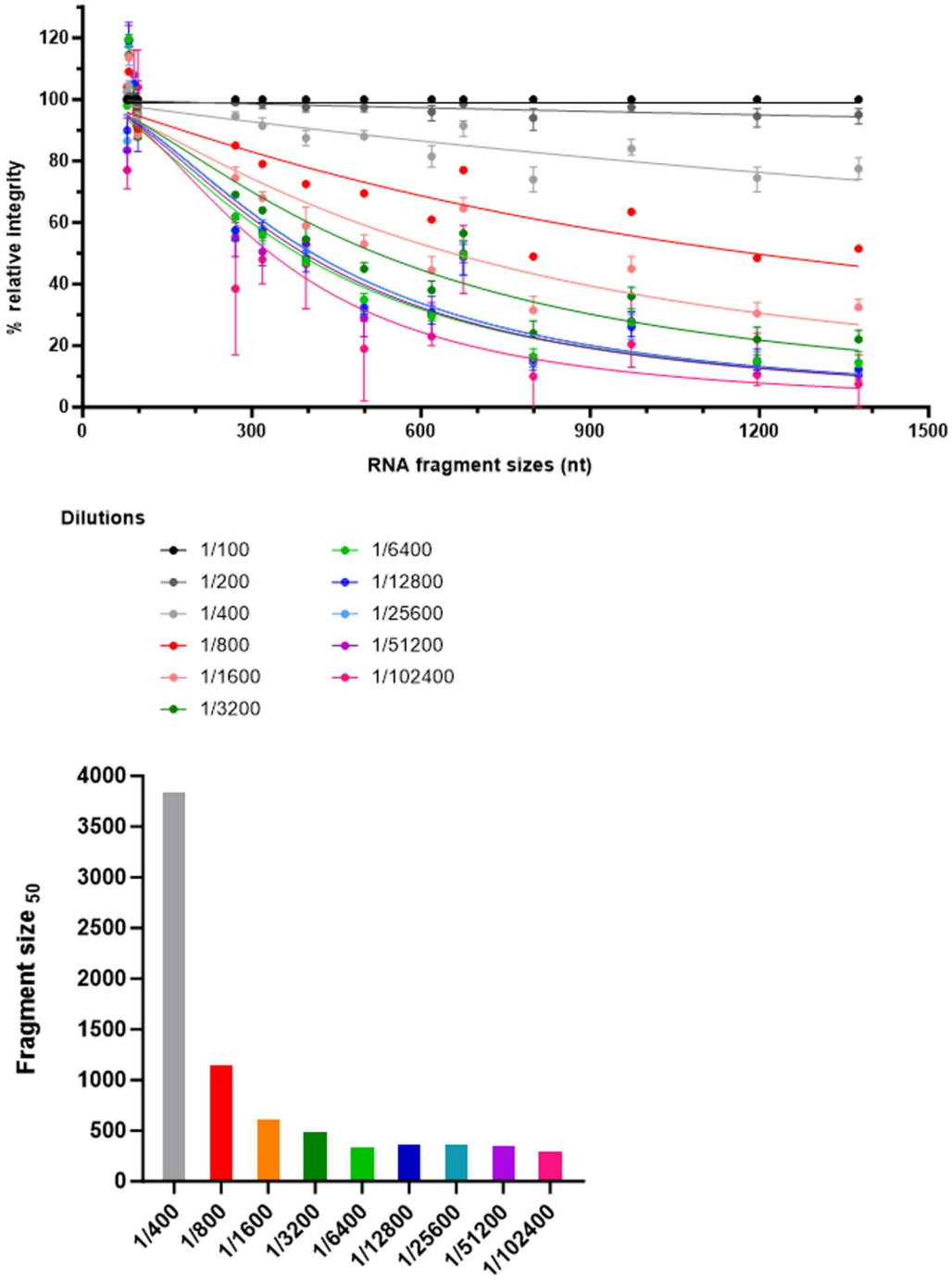


Figure 4. Impact of RNA concentration on the relevance of RNA RI estimation. The higher the RNA concentration, the greater the probability that two independent amplicons colocalized within the same partition, leading to overestimation of the RI. RI as a function of RNA fragment size was drawn and fragmentation index (FS50) was estimated for each RNA concentration.

(Fig 6C and Supplementary Fig. S10). Results from RNA extracted using automated method were provided in Supplementary Figs S9, S11, and S12.

Impact of sample matrix on viral RNA integrity

Environmental matrix has also been investigated on viral RNA integrity. New fresh viral supernatant was collected and used to spike raw wastewater having high (822 NTU) or low (23 NTU) turbidity. RNA was extracted using both methods and then amplified.

RNA concentration was significantly lower with automated method (control: 24 ± 3 copies/µl) than with manual method (control: 73 ± 9 copies/µl) (Kruskal–Wallis test, $P = 0.035$). RNA concentrations from wastewater samples were not significantly different from the control sample (Kruskal–Wallis test, $P > 0.05$) (Supplementary Figs S13A and S13B). RNA RI was not notably affected by the nature of the water matrix in these conditions (Fig. 7). For manual extraction method, FS_{50} was estimated at 555 nt, 482 nt, and 503 nt for control, low turbidity wastewater, and high turbidity wastewater samples, respectively. Similar

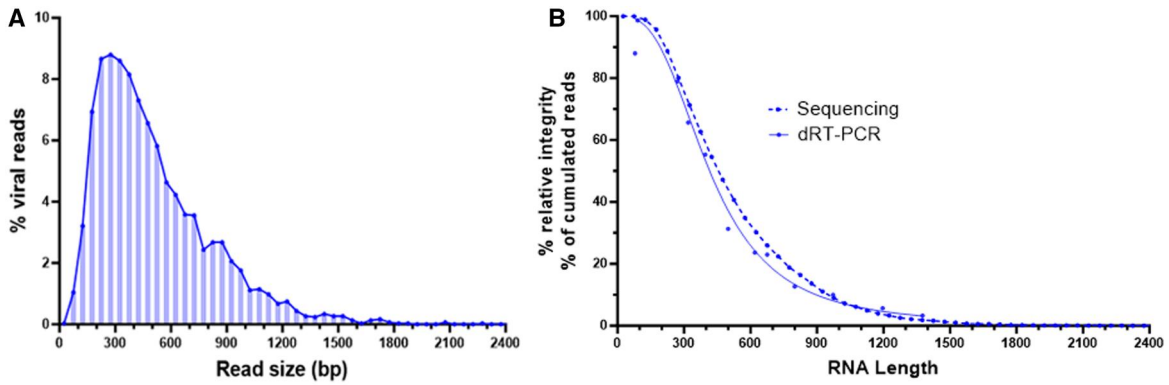


Figure 5. Comparison of RNA fragment size estimation and sequencing read size distribution. RNA has been extracted from four viral supernatant replicates using automated method. RI of fresh RNA has been estimated. At the same time, a pool of the four extracted RNA has been sequenced using ONT. (A) Distribution of sequencing read sizes grouped by 50-bp windows. (B) Comparison between cumulated percentages of sequencing reads and relative RNA integrity estimated by dRT-PCR.

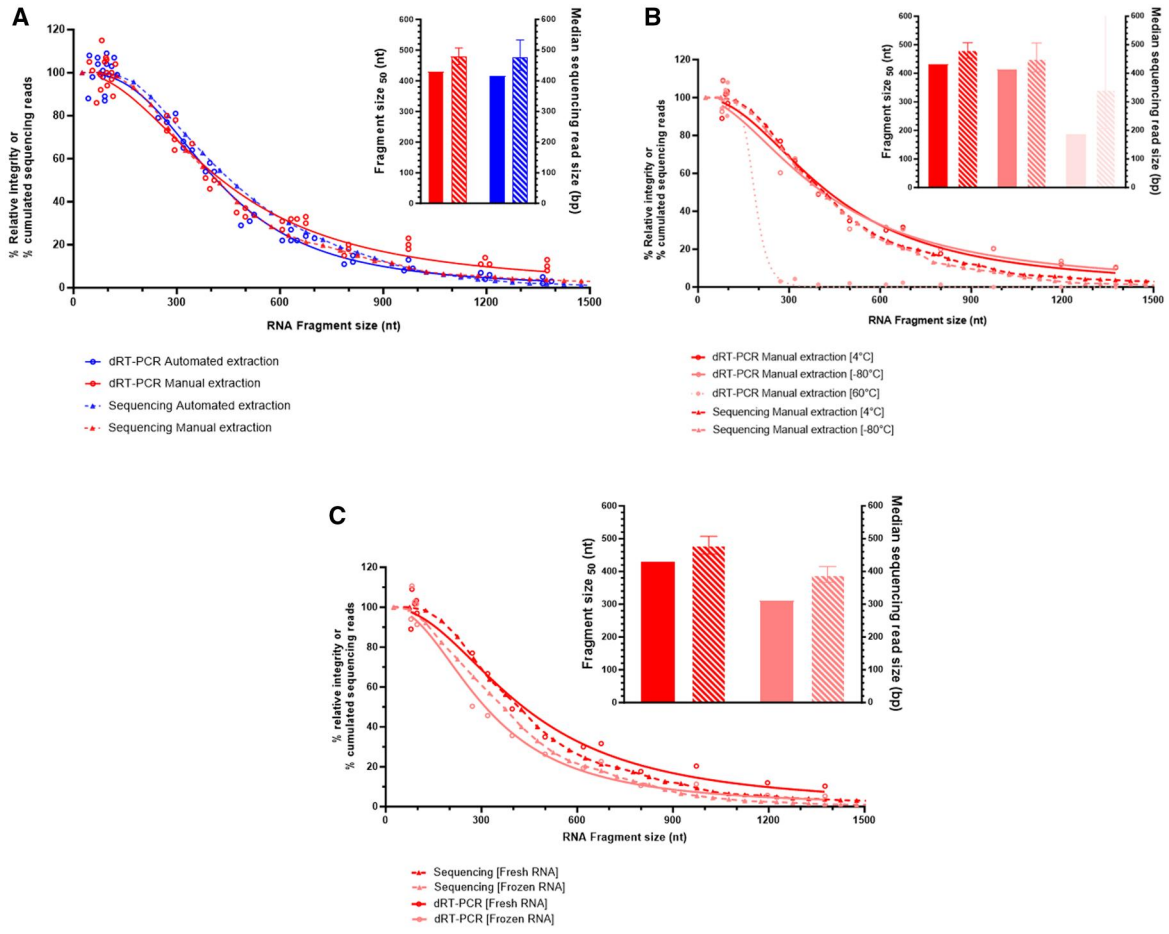


Figure 6. Application of viral RNA integrity assessment for environmental purposes. (A) Impact of nucleic acid extraction method on viral RNA integrity. Four replicates have been extracted using manual method or automated method. RNA has been amplified and quantified. RI (solid curves) is represented as a function of RNA fragment size determined using dPCR-based assay. FS50 index (solid bar charts) was deducted from integrity decay curves for each extraction method. The percentage of cumulated sequencing reads depending on their size (grouped and sorted by 50-bp windows) was also drawn (dotted curves). The median size of sequencing reads (hatched bar charts) has been calculated and compared to the FS50. The error bars on chart represented the 95% IC. (B) Impact of raw sample storage on viral RNA integrity. Four replicates have been stored at 4°C or −80°C for 4 h or inactivated by heat exposure (60°C for 30 min) and then extracted using manual method. RI (solid curves) is represented as a function of RNA fragment size determined using dPCR-based assay. FS50 index (solid bar charts) was deducted from integrity decay curves for each storage condition. The percentage of cumulated sequencing reads depending on their size (grouped and sorted by 50-bp windows) was also drawn (dotted curves). The median size of sequencing reads (hatched bar charts) has been calculated and compared to the FS50. The error bars on chart represented the 95% IC. Too few reads were obtained to draw a curve of cumulated sequencing reads from heated sample (Supplementary Fig. S5). (C) Impact of RNA storage on viral RNA integrity. Four replicates have been extracted using both methods and then used or stored at −20°C for 7 days before use. RNA has been amplified and quantified. RI (solid curves) is represented as a function of RNA fragment size determined using dPCR-based assay. FS50 index (solid bar charts) was deducted from integrity decay curves for each storage condition. The percentage of cumulated sequencing reads depending on their size (grouped and sorted by 50-bp windows) was also drawn (dotted curves). The median size of sequencing reads (hatched bar charts) has been calculated and compared to the FS50. The error bars on chart represented the 95% IC.

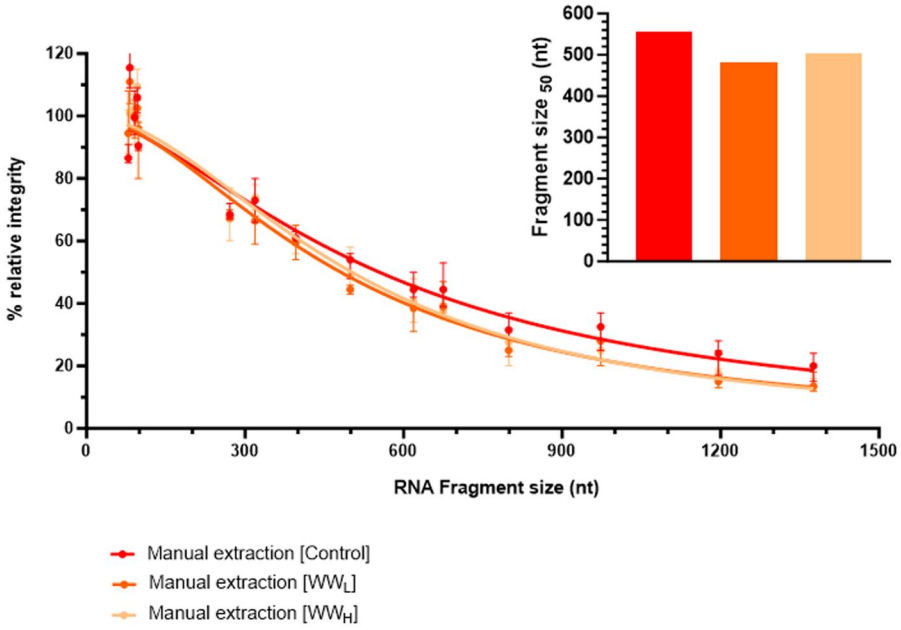


Figure 7. Impact of hydric matrix on viral RNA integrity. Four replicates of each sample have been spiked with coxsackievirus B3 and then extracted using manual method. RNA has been amplified and quantified. Wastewater samples differed by their turbidity: high (noted WWH) or low (noted WWL). RI is represented as a function of RNA fragment size determined. FS50 index was deducted from integrity decay curves for each sample.

Table 2. Summary of tested conditions that could impact the quantity and the quality of extracted viral RNA.

Process	Tested conditions	Impact on	
		RNA Quantity	RNA Quality
RNA extraction	Manual silica membrane/Automated magnetic beads	Moderate	Low
Raw sample	Freezing at −80°C	Low	Low
	Heat inactivation 60°C	High	High
Extracted RNA	Freezing at −20°C	Low	Moderate
Sample matrix	Wastewater	Low	Low

results based on automated RNA extraction were provided in Supplementary Fig. S14.

Discussion

The development of new genomic analysis technologies such as New Generation Sequencing (NGS) has greatly improved our knowledge of microbial diversity in water environments. However, many sequencing methods are limited by the size of sequenced fragments. For example, the read size from bacterial 16S ribosomal DNA sequencing generally affects the accuracy of taxonomic assignment of microorganisms (for review, Zhang *et al.* [30]). In the case of SARS-CoV-2 sequencing in wastewater, even though recent methods for deconvolving reads from complex samples have appeared [15, 16], the possibility of having as much genetic information as possible on a single fragment would provide greater confidence in predicting circulating variants. Low and/or variable quality of extracted viral RNA has a major impact on epidemiological and/or health surveillance. Characterization of nucleic acid quality is rarely considered in method optimization. Few approaches have been developed to assess RNA integrity in a targeted manner. In addition to biases linked to frequent inhibition by environmental matrices, there are biases in PCR efficiency resulting from the amplification gradient of long-range PCR. The shorter the fragment to be amplified, the more efficient the amplification reaction. On the other hand, the ability of thermal cyclers to discriminate between different fluorescent signals favors multiplexed amplification, with each fluorescent signal associated with an amplicon of varying size. This approach is technically constrained by amplicon size limits, compatible with probe-based detection. In this study, we described a new method for specifically assessing the quality of viral RNA. This approach took advantage of both the multiplexing capabilities of instrument and the PCR reaction partitioning provided by dPCR technology. The PCR efficiency of qPCR assay is highly dependent on amplicon length. To overcome the poor PCR efficiency of long amplicon, this method focuses on the distance between amplified regions. This assay allowed fragments of up to 1371 nt to be considered. The principle was to analyze a very wide distribution of RNA fragment sizes (up to 10 in this proof-of-concept) and to model the RI decay curve as a function of fragment size. As each sample was analyzed in four replicates, this method demonstrated its repeatability and enabled to define a new robust indicator of RNA integrity. This new fragmentation indicator, so-called FS50, provided direct information on the minimum size of 50% of extracted RNA (i.e. the median value of RNA sizes).

Any molecular analysis of environmental samples necessarily involves several steps that aimed at concentrating the microorganisms of interest present in the sample and extracting their nucleic acids. Various treatments can impact the quantity and

quality of extracted nucleic acids, such as the need to inactivate potentially highly pathogenic microorganisms present in samples, or to neutralize environmental components (such as natural organic matter) that tend to largely inhibit extraction and amplification steps. The molecular tool developed has been applied to address most of concerns associated with environmental analysis and results are summarized in Table 2. The study revealed some disparities in RNA recovery yield between different extraction methods, but no difference in RNA quality between two technologically different extraction methods. A manual method, dedicated to the purification of RNA by retention on a silica membrane, was compared to an automated method for the purification of total nucleic acids (DNA and RNA indifferently) using silica-coated magnetic beads. Contrary results have been reported [31] but in their study, the authors focused on viruses associated with suspended solids. The clogging of membranes by this organic matter could probably be at the origin of lower RNA recovery yield using membrane-based methods, as already reported [32]. By studying storage conditions, freezing raw samples at −80°C did not seem to affect the quality of extracted RNA, whereas storage of extracted RNA at −20°C was accompanied by an increased RNA. Previous study reported moderate sequencing coverage decrease of SARS-CoV-2 after RNA freezing in agreement with our results. However, contrary to our report on CV-B3, a strong decrease in SARS-CoV-2 RNA sequencing coverage was reported after raw wastewater sample freezing [33]. Wastewater is a complex matrix, in which SARS-CoV-2, and probably enveloped viruses in general, is rapidly degraded, contrary to coxsackievirus B3, a non-enveloped enteric virus commonly implied in human gastroenteritis. The presence of detergents, for example, can impact the lipid bilayer, degrading infectivity and promoting degradation of the viral genome. The nature of viral genome protection could have a different impact on this degradation, depending on whether the genome is protected by a capsid like nonenveloped viruses, or by a corecapsid like Chikungunya virus or Zika virus which are enveloped viruses, or conversely by a ribonucleoprotein complex like coronavirus or influenza virus. In this context, we did not report any notable impact of wastewater matrix on CV-B3 RNA integrity.

To reduce the risk of operator exposure to potentially pathogenic microorganisms, it is generally recommended to thermally inactivate these microorganisms prior to handling. This study showed that this significantly reduced RNA recovery, according to previous studies targeting SARS-CoV-2 [31, 34], but above all that, it resulted in higher RNA fragmentation. This pretreatment step therefore appeared compatible with qualitative molecular detection on highly contaminated samples, as only a very small fragment is amplified, although this may reduce the sensitivity of the analysis. On the other hand, it should be discouraged for downstream applications such as long-read sequencing. The impact of microorganism concentration methods (filtration, precipitation, or centrifugation) prior to nucleic acid extraction could also influence genome integrity and should be assessed.

The estimated proportions of viral RNA sizes were compared with the size distribution of sequencing reads of this same RNA. Estimation of the FS_{50} index was remarkably like the median size of the sequencing reads, thus confirming the relevance of the tools developed. However, the RT-dPCR estimate tended to underestimate the RNA fragment size compared with the sequencing reads. This method is based on the distance between amplified regions. The greater the distance between two amplicons, less accurate the estimate of fragment size. Indeed, in partitions positive for the first two amplicons, the region is intact at least up to the 3'-end of the second amplicon. However, this region could also be conserved up to the 5'-end of the third amplicon. Thus, this approach provides information on the minimum size of RNA fragments, which would explain this slight underestimation of RNA size. However, considering 10 different sizes and modeling the integrity decay curve probably limits this uncertainty that ranged between 8% and 35% considering all the comparison results available from this study.

However, we have identified several factors that could affect the results presented. First, ONT sequencing of viral RNA necessarily involved a prior reverse transcription step, which we hoped would be as conservative as possible in terms of fragment size. This crucial step was done using random hexamer and a highly processive and thermostable reverse transcriptase. However, incomplete conversion of RNA into cDNA was not able to be ruled out, but this did not seem to be the case in our results in view of the similarity of the results obtained by dPCR and sequencing. Direct RNA sequencing kit exists but it is based on polyadenylated fragments. If the 3'-end of the CVB3 genome is naturally polyadenylated, analysis of internal genome fragments would have required a poly(A)-tailing step, also creating uncertainty [35]. Second, RI decay curve was generally incomplete, and settings of this curve did not include any constraints to ensure that the curve tends toward 0, even for full-length genome size. It is possible that the presence of long fragments could impact modeling of the curve and might reduce the accuracy of the extrapolation. Including amplicon for generating full-length fragments might specifically respond to such issues. Another bias could come from the impossibility of distinguishing the random colocalization events of five amplicons from the truly intact fragments carrying the five amplicons in partitions positive for the five amplicons. This probability is very low for the range of concentrations determined, and we estimated that the impact of this bias was probably marginal. Another limitation of this approach is the minimum size of amplifiable RNA fragments. This is determined by the length of the amplicons making up the assay. Finally, the technical difficulty of designing an assay with smaller inter-amplicon distances without risking partially overlapping amplification limits the representativity of small fragments.

Depending on the sequencing quality filters, nucleic acid fragmentation could lead to the analysis of only fragments, corresponding to a minority fraction of circulating genomes, not representative of genetic diversity within the sample. The assessment of nucleic acid fragmentation could thus be imposed as a quality control, a decisive prerequisite for ensuring the relevance of long-read sequencing.

In this proof-of-concept, we sought to develop an innovative analytical method to address issues associated with environmental molecular analysis. However, the measurement of nucleic acid integrity has a much wider field of application (clinical, water, soil, etc.) even with low-concentrated samples. When no simple cell culture system is available, assessing the infectivity of viruses is always very problematic. The integrity of viral particles, in particular accessibility to the viral genome, has been used as a proxy for viral infectivity. However, there are limits to the relevance of this approach, particularly when degradation only affects the viral genome without affecting the envelope or capsid [36]. This assessment of genomic integrity could be a complementary approach. In a different field, DNA fragmentation has long been associated with programmed cell death and the maintainance of genomic stability [37]. This method of assessing the fragmentation could be especially applied to cell-free DNA

[38] and make it possible to discover new tumor markers for early diagnosis of cancer and surveillance during cancer progression. This innovation could also open new perspectives in very different fields that have not been investigated in this study.

Conclusions

This study is a proof-of-concept of an innovative method for determining the RI of nucleic acids, especially viral RNA. The method takes advantage of digital PCR partitioning technology combined with multiplexing of amplifications on the same genome. Considering the distances between the different amplified regions, instead of the length of the amplicons themselves, without the PCR efficiency issues associated with long amplicons, allows us to know the size distribution of the targeted RNA. This study reports on determining a new fragmentation index relative to a targeted RNA fragmentation state and an application of this molecular tool in an environmental analysis context, notably for long-read sequencing purposes. Characterization of the fragmentation of nucleic acids extracted from a sample should become a quality control prerequisite for any long-read sequencing, to optimize the use of this NGS technology and reduce interpretation bias linked to non-representative analysis of the totality of fragments.

Author contributions

Sebastien Wurtzer (Conceptualization [lead], Data curation [lead], Formal analysis [lead], Methodology [lead], Visualization [lead], Writing—original draft [lead]), Mathilde Duvivier (Formal analysis [supporting]), Heberte Accrombessi (Formal analysis [supporting]), Morgane Levert (Writing—review & editing [equal]), Elise Richard (Writing—review & editing [equal]), and Laurent Moulin (Conceptualization [supporting], Formal analysis [supporting], Funding acquisition [lead], Writing—review & editing [lead])

Supplementary data

Supplementary data is available at *Biology Methods and Protocols* online.

Data availability

The data underlying this article will be shared on reasonable request to the corresponding author.

Conflict of interest statement. None declared.

Funding

This project was funded by Eau de Paris.

References

1. Hasing ME, Lee BE, Gao T *et al.* Wastewater surveillance monitoring of SARS-CoV-2 variants of concern and dynamics of transmission and community burden of COVID-19. *Emerg Microbes Infect* 2023;**12**:2233638.
2. Wurtzer S, Levert M, Dhenain E *et al.* From Alpha to Omicron BA.2: new digital RT-PCR approach and challenges for SARS-CoV-2 VOC monitoring and normalization of variant dynamics in wastewater. *Sci Total Environ* 2022;**848**:157740.
3. Ahmed W, Tscharke B, Bertsch PM *et al.* SARS-CoV-2 RNA monitoring in wastewater as a potential early warning system for COVID-19 transmission in the community: a temporal case study. *Sci Total Environ* 2021;**761**:144216.
4. Wurtzer S, Marechal V, Mouchel J *et al.* Evaluation of lockdown effect on SARS-CoV-2 dynamics through viral genome quantification in waste water, Greater Paris, France, 5 March to 23 April 2020. *Eurosurveillance* 2020;**25**:2000776.
5. de Jonge EF, Peterse CM, Koelewijn JM *et al.* The detection of monkeypox virus DNA in wastewater samples in the Netherlands. *Sci Total Environ* 2022;**852**:158265.
6. Wurtzer S, Levert M, Dhenain E, OBEPINE SIG *et al.* First Detection of Monkeypox Virus Genome in Sewersheds in France: the Potential of Wastewater-Based Epidemiology for Monitoring Emerging Disease. *Environ Sci Technol Lett* 2022;**9**:991–6.
7. Domingo E, Perales C. Viral quasispecies. *PLoS Genet* 2019;**15**:e1008271.
8. Singh K, Mehta D, Dumka S *et al.* Quasispecies Nature of RNA Viruses: lessons from the Past. *Vaccines (Basel)* 2023;**11**:308.
9. Gregory DA, Trujillo M, Rushford C *et al.* Genetic diversity and evolutionary convergence of cryptic SARS- CoV-2 lineages detected via wastewater sequencing. *PLoS Pathog* 2022;**18**:e1010636.
10. Karthikeyan S, Levy JI, De Hoff P *et al.* Wastewater sequencing reveals early cryptic SARS-CoV-2 variant transmission. *Nature* 2022;**609**:101–8.
11. Vo V, Harrington A, Chang C-L *et al.* Identification and genome sequencing of an influenza H3N2 variant in wastewater from elementary schools during a surge of influenza A cases in Las Vegas, Nevada. *Sci Total Environ* 2023;**872**:162058.
12. Fontenele RS, Kraberger S, Hadfield J *et al.* High-throughput sequencing of SARS-CoV-2 in wastewater provides insights into circulating variants. *Water Res* 2021;**205**:117710.
13. Gupta P, Liao S, Ezekiel M *et al.* Wastewater Genomic Surveillance Captures Early Detection of Omicron in Utah. *Microbiol Spectr* 2023;**11**:e00391-23.
14. Rios G, Lacoux C, Leclercq V *et al.* Monitoring SARS-CoV-2 variants alterations in Nice neighborhoods by wastewater nanopore sequencing. *Lancet Reg Health Eur* 2021;**10**:100202.
15. Jahn K, Dreifuss D, Topolsky I *et al.* Early detection and surveillance of SARS-CoV-2 genomic variants in wastewater using COJAC. *Nat Microbiol* 2022;**7**:1151–60.
16. Schumann V-F, de Castro Cuadrat RR, Wyler E *et al.* SARS-CoV-2 infection dynamics revealed by wastewater sequencing analysis and deconvolution. *Sci Total Environ* 2022;**853**:158931.
17. Nemudryi A, Nemudraia A, Wiegand T *et al.* Temporal Detection and Phylogenetic Assessment of SARS-CoV-2 in Municipal Wastewater. *Cell Rep Med* 2020;**1**:100098.
18. Quick J, Grubaugh ND, Pullan ST *et al.* Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc* 2017;**12**:1261–76.
19. Knyazev S, Hughes L, Skums P, Zelikovsky A. Epidemiological data analysis of viral quasispecies in the next-generation sequencing era. *Brief Bioinform* 2021;**22**:96–108.
20. Beaulaurier J, Luo E, Eppley JM *et al.* Assembly-free single-molecule sequencing recovers complete virus genomes from natural microbial communities. *Genome Res* 2020;**30**:437–46.
21. Tedersoo L, Albertsen M, Anslan S *et al.* Perspectives and benefits of high-throughput long-read sequencing in microbial ecology. *Appl Environ Microbiol* 2021;**87**:e00626-21.
22. Zaragoza-Solas A, Haro-Moreno JM, Rodriguez-Valera F *et al.* Long-read metagenomics improves the recovery of viral

diversity from complex natural marine samples. *mSystems* 2022;**7**:e00192-22.
23. Wurtzer S, Waldman P, Ferrier-Rembert A *et al.* ; OBEPINE Consortium. Several forms of SARS-CoV-2 RNA can be detected in wastewaters: implication for wastewater-based epidemiology and risk assessment. *Water Res* 2021;**198**:117183.
24. Swerdlow H, Gesteland R. Capillary gel electrophoresis for rapid, high resolution DNA sequencing. *Nucleic Acids Res* 1990; **18**:1415–9.
25. Schroeder A, Mueller O, Stocker S *et al.* The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol* 2006;**7**:3.
26. Han J, Lee JY, Bae Y-K. Application of digital PCR for assessing DNA fragmentation in cytotoxicity response. *Biochim Biophys Acta Gen Subj* 2019;**1863**:1235–42.
27. Pinzani P, Salvianti F, Zaccara S *et al.* Circulating cell-free DNA in plasma of melanoma patients: qualitative and quantitative considerations. *Clin Chim Acta* 2011;**412**:2141–5.
28. Umetani N, Kim J, Hiramatsu S *et al.* Increased integrity of free circulating dna in sera of patients with colorectal or periampullary cancer: direct quantitative PCR for ALU repeats. *Clin Chem* 2006;**52**:1062–9.
29. Didelot A, Kotsopoulos SK, Lupo A *et al.* Multiplex picoliter-droplet digital PCR for quantitative assessment of DNA integrity in clinical samples. *Clin Chem* 2013;**59**:815–23.
30. Zhang G, Wei L, Chang C *et al.* Molecular biological methods in environmental engineering. *Water Environ Res* 2016;**88**:930–53.
31. Palmer EJ, Maestre JP, Jarma D *et al.* Development of a reproducible method for monitoring SARS-CoV-2 in wastewater. *Sci Total Environ* 2021;**799**:149405.
32. Zheng X, Deng Y, Xu X *et al.* Comparison of virus concentration methods and RNA extraction methods for SARS-CoV-2 wastewater surveillance. *Sci Total Environ* 2022;**824**:153687.
33. Barbé L, Schaeffer J, Besnard A *et al.* ; OBEPINE Consortium. SARS-CoV-2 whole-genome sequencing using oxford nanopore technology for variant monitoring in wastewaters. *Front Microbiol* 2022;**13**:889811.
34. Islam G, Gedge A, Lara-Jacobo L *et al.* Pasteurization, storage conditions and viral concentration methods influence RT-qPCR detection of SARS-CoV-2 RNA in wastewater. *Sci Total Environ* 2022;**821**:153228.
35. Wongsurawat T, Jenjaroenpun P, Taylor MK *et al.* Rapid sequencing of multiple RNA viruses in their native form. *Front Microbiol* 2019;**10**:260.
36. Prevost B, Goulet M, Lucas FS *et al.* Viral persistence in surface and drinking water: suitability of PCR pre-treatment with intercalating dyes. *Water Res* 2016;**91**:68–76.
37. Yan B, Wang H, Peng Y *et al.* A unique role of the DNA fragmentation factor in maintaining genomic stability. *Proc Natl Acad Sci U S A* 2006;**103**:1504–9.
38. Cristiano S, Leal A, Phallen J *et al.* Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature* 2019; **570**:385–9.

Biology Methods and Protocols, 2024, 00, 1–14
https://doi.org/10.1093/biomethods/bpae053
Methods Article