



**HAL**  
open science

## Definition of a concentration and RNA extraction protocol for optimal whole genome sequencing of SARS-CoV-2 in wastewater (ANRS0160)

Ahlam Chaqroun, Ghina El Soufi, Zuzana Gerber, Julie Loutreul, Nicolas Cluzel, Damien Delafoy, Florian Sandron, Léo Di Jorio, Stéphanie Raffestin, Vincent Maréchal, et al.

### ► To cite this version:

Ahlam Chaqroun, Ghina El Soufi, Zuzana Gerber, Julie Loutreul, Nicolas Cluzel, et al.. Definition of a concentration and RNA extraction protocol for optimal whole genome sequencing of SARS-CoV-2 in wastewater (ANRS0160). *Science of the Total Environment*, 2024, 952, pp.175823. 10.1016/j.scitotenv.2024.175823 . hal-04764746

**HAL Id: hal-04764746**

<https://hal.univ-lorraine.fr/hal-04764746v1>

Submitted on 4 Nov 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



## Definition of a concentration and RNA extraction protocol for optimal whole genome sequencing of SARS-CoV-2 in wastewater (ANRS0160)

Ahlam Chaqroun<sup>a,i</sup>, Ghina El Soufi<sup>b,c,i</sup>, Zuzana Gerber<sup>d</sup>, Julie Loutreul<sup>e,i</sup>, Nicolas Cluzel<sup>f,i</sup>, Damien Delafoy<sup>d</sup>, Florian Sandron<sup>d</sup>, Léo Di Jorio<sup>b,c,i</sup>, Stéphanie Raffestin<sup>g,i</sup>, Vincent Maréchal<sup>h,i</sup>, Christophe Gantzer<sup>a,i</sup>, Robert Olasso<sup>d</sup>, Jean-François Deleuze<sup>d</sup>, Olivier Rohr<sup>b,c,i</sup>, Nicolas Boudaud<sup>e,i,1</sup>, Clémentine Wallet<sup>b,c,i,1</sup>, Isabelle Bertrand<sup>a,i,\*,1</sup>

<sup>a</sup> Université de Lorraine, CNRS, LCPME, F-54000 Nancy, France

<sup>b</sup> Université de Strasbourg, UPR CNRS 9002 ARN, F-67300 Schiltigheim, France

<sup>c</sup> Université de Strasbourg, IUT Louis Pasteur, F-67300 Schiltigheim, France

<sup>d</sup> CEA, Centre National de Recherche en Génomique Humaine, Université Paris-Saclay, F-91057 Evry, France

<sup>e</sup> ACTALIA, F-50000 Saint Lô, France

<sup>f</sup> Maison des Modélisations Ingénieries et Technologies (SUMMIT), Sorbonne Université, Paris 75005, France

<sup>g</sup> Institut Pasteur de la Guyane, French Guiana, Cayenne 97300, France

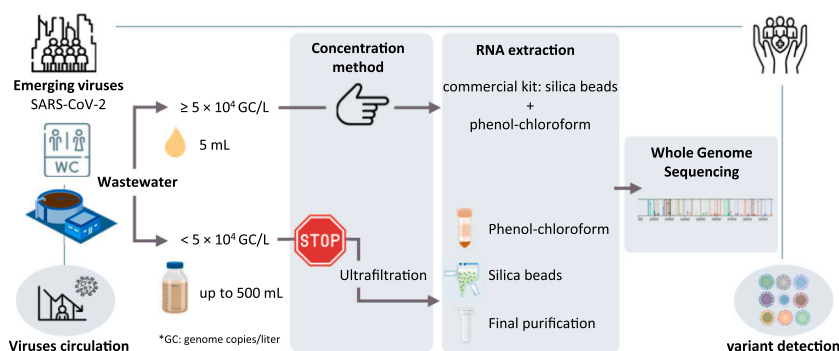
<sup>h</sup> INSERM, Centre de Recherche Saint-Antoine, Sorbonne Université, Paris 75012, France

<sup>i</sup> OBEPINE consortium, Paris, France

### HIGHLIGHTS

- Optimized protocols are required for sequencing viral RNA in influent samples.
- Ultrafiltration/silica bead RNA extraction can be used on 100–500 mL of influent.
- Above  $5 \times 10^4$  GC/L, concentration method is not required for sequencing SARS-COV-2.
- Below  $5 \times 10^4$  GC/L, an ultrafiltration method allows successful sequencing.
- Below  $4 \times 10^3$  GC/L, sequencing is feasible in half of cases from 500 mL of influent.

### GRAPHICAL ABSTRACT



### ARTICLE INFO

Editor: Qilin Wang

**Keywords:**  
Concentration methods  
SARS-CoV-2

### ABSTRACT

Monitoring the presence of RNA from emerging pathogenic viruses, such as SARS-CoV-2, in wastewater (WW) samples requires suitable methods to ensure an effective response. Genome sequencing of WW is one of the crucial methods, but it requires high-quality RNA in sufficient quantities, especially for monitoring emerging variants. Consequently, methods for viral concentration and RNA extraction from WW samples have to be

\* Corresponding author at: Université de Lorraine, CNRS, LCPME, F-54000 Nancy, France.

E-mail address: [isabelle.bertrand@univ-lorraine.fr](mailto:isabelle.bertrand@univ-lorraine.fr) (I. Bertrand).

<sup>1</sup> These authors contributed equally to this work.

<https://doi.org/10.1016/j.scitotenv.2024.175823>

Received 6 June 2024; Received in revised form 26 July 2024; Accepted 25 August 2024

Available online 26 August 2024

0048-9697/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Ultrafiltration  
Wastewater  
Whole genome sequencing

optimized before sequencing. The purpose of this study was to achieve high coverage ( $\geq 90\%$ ) and sequencing depth (at least  $\geq 200\times$ ) even for low initial RNA concentrations ( $< 10^5$  genome copies (GC)/L) in WW. A further objective was to determine the range of SARS-CoV-2 RNA concentrations that allow high-quality sequencing, and the optimal sample volume for analysis. Ultrafiltration (UF) methods were used to concentrate viral particles from large influent samples (up to 500 mL). An RNA extraction protocol using silica beads, neutral phenol-chloroform treatment, and a PCR inhibitor removal kit was chosen for its effectiveness in extracting RNA and eliminating PCR inhibitors, as well as its adaptability for use with large influent samples. Recovery rates ranged from 24 % to 63 % ( $N = 17$ ) for SARS-CoV-2 naturally present in WW samples. 200 mL WW samples can be enough for UF concentration, as they showed high quality sequencing analyses with between  $5 \times 10^4$  GC/L and  $6 \times 10^3$  GC/L. Below  $6 \times 10^3$  GC/L, high-quality sequencing was also achieved for  $\sim 40\%$  of the samples using 500 mL of WW. Sequencing analysis for variant detection was performed on 200 mL WW samples with coverage of  $>95\%$  and sequencing depth of  $>1000\times$ . Analyses revealed the predominance of variant EG.5, known as Eris (66%–100%). The use of UF methods in combination with a suitable RNA extraction protocol appear promising for sequencing enveloped viruses in WW in a context of viral emergence.

## 1. Introduction

The presence of human viruses in wastewater (WW) can indicate that these viruses are excreted by the human population whose waste is treated by the WW treatment plant (WWTP), provided that these viruses' genomes persist long enough to be detected (Carducci et al., 2020). Wastewater-based epidemiology (WBE) has emerged as an effective tool for tracking microbial circulation in populations. Through the analysis of temporal variations in viral concentrations and diversity, WBE not only helps to assess the infection rates within a population, it also enables the early detection of emerging viral strains (Polo et al., 2020). This methodology has proven particularly valuable in monitoring SARS-CoV-2, the virus responsible for COVID-19, and serves as an "early warning system" for public health authorities (Daughton, 2020; Hart and Halden, 2020; Polo et al., 2020). The WBE approach employs molecular techniques like reverse transcription-real-time quantitative PCR (RT-qPCR), RT droplet digital PCR (RT-ddPCR), RT digital PCR (RT dPCR), or high-throughput qPCR to measure viral or bacterial nucleic acid concentrations or to detect multiple targets simultaneously in wastewater samples (Ahmed et al., 2022; Kabdaşlı and Tünay, 2021; Sims and Kasprzyk-Hordern, 2020). While these methods are reliable tools for tracking emerging variants in WW, they are limited to the detection of specific mutations. Genome sequencing is essential for understanding the genetic evolution of viruses in WW and to gain comprehensive epidemiological insights. However, identifying variants of interest (VOI) or variants of concern (VOC), especially those present at low frequencies, requires high-quality sequencing (Van Poelvoorde et al., 2021). This is made challenging by various issues specific to WW sequencing, such as fragmented genomes due to the nature of the samples (Otero et al., 2022); intensive sample processing like sample homogenization during viral concentration and RNA extraction steps, especially the bead-beating procedure (Feng et al., 2023); potential PCR inhibitors that are co-purified during viral concentration; and notably, low concentrations of viral RNA (Rouchka et al., 2021). Overcoming these challenges requires specialized concentration and extraction methods tailored for handling low viral RNA concentrations effectively while maintaining genome quality.

Until now, most studies have focused on targeted sequencing in regions with key mutations (Cancela et al., 2023; Kuroiwa et al., 2023; Lin et al., 2020; Smyth et al., 2022), while whole-genome sequencing (WGS) provides more comprehensive information, enabling the identification of new mutations for a more precise evaluation of genetic evolution. Despite instances of successful WGS with coverage exceeding 90 % (Child et al., 2023; Crits-Christoph et al., 2021; Karthikeyan et al., 2022), several other studies have obtained sequencing coverage of below 50 % (Cha et al., 2023; Feng et al., 2023; Li et al., 2022; Otero et al., 2022). In most studies with successful sequencing protocols, WW samples had high viral concentrations ( $> 10^5$  GC/ $\mu$ L). These protocols, effective during periods of high viral prevalence in human populations, may be insufficient when virus circulation is low due to epidemic

fluctuations or during the early stages of variant emergence. These situations require large sample volumes to be processed, which can result in the simultaneous concentration of other substances that impede sequencing analysis (Tamás et al., 2022).

There is currently no standardized protocol for SARS-CoV-2 sequencing in WW. Various techniques are described in the literature, ranging from sample concentration to library preparation. The choice of a suitable concentration methods is crucial, because periods of low viral circulation in human populations can be an important limitation and concentrating small sample volumes of WW ( $\leq 100$  mL) may not be sufficient (Ahmed et al., 2022). The UF and polyethylene glycol (PEG) precipitation methods can be used with large volumes of WW ( $> 100$  mL) (Bofill-Mas and Rusiñol, 2020). A recent study using an UF membrane-based concentration method applied to 500 mL of influent showed recovery rates over 50 % for SARS-CoV-2 and non-enveloped F-specific RNA phages belonging to genogroup II (FRNAPH-II); it also enabled the detection of viruses at concentrations as low as 40 GC/L (El Soufi et al., 2024). However, this interesting approach requires further optimization, especially RNA extraction, to allow high-quality sequencing. Selecting the right protocol for RNA extraction is also challenging since it must maximize the recovery of nucleic acids without damaging them (Michael-Kordatou et al., 2020). Most RNA extraction methods rely on column extraction techniques, producing RNA at high concentrations while maintaining high purity. However, this kind of method is limited by the minimal size of fragments that can adhere to the column, resulting in the potential loss of a substantial portion of RNA fragments (Tamás et al., 2022). Extraction methods using magnetic beads made of silica can also be applied to WW concentrates due to their ability to effectively recover high-purity nucleic acids, including RNA. Additionally, an organic extraction method involving TRIzol™ is frequently used for RNA extraction, but residual traces of TRIzol™ or isopropanol could disturb molecular analyses (Tamás et al., 2022). Lastly, the right combination of viral concentration and RNA extraction methods is critical for effective genomic analysis. Kaya et al. (2022) have evaluated 11 virus concentration protocols based on PEG precipitation and UF for their effectiveness in recovering enveloped viruses in WW samples. These protocols were combined with four RNA extraction protocols resulting in 44 different possible approaches. They reported a wide range of recovery rates for bovine respiratory syncytial virus (BRSV) RNA, ranging from  $< 0.03\%$  to  $64.7\%$  ( $\pm 1.6\%$ ) for initial WW volumes of between 30 mL and 100 mL.

Regarding sequencing technology, Illumina combined with various sets of primers, especially ARTIC primers, is one of the most frequently used methods; it allows the coverage of almost the entire SARS-CoV-2 genome (Tamás et al., 2022). The process applied to samples upstream of sequencing significantly influences its quality and consistency (Feng et al., 2023).

Uncertainty remains in terms of the minimal RNA concentration required for achieving high-quality sequencing in WW samples. This is crucial given the challenge of sequencing small quantities of RNA,

especially in the context of viral emergence. To address this limitation, our study aims to assess various combinations of protocols for viral concentration and RNA extraction using large WW sample volumes (100 mL to 500 mL) collected from diverse WWTPs. Through this comparative analysis, we aim to identify the most effective combination that ensures optimal recovery of viral RNA genomes as well as high-quality sequencing, defined as coverage of  $\geq 90\%$  and depth of at least  $\geq 200\times$ . Additionally, we aim to determine the threshold concentrations of SARS-CoV-2 RNA that require a concentration step, thereby identifying the most appropriate volume of WW for analysis. The long-term objective of our research is to contribute to the identification of emerging viruses in WW through sequencing, providing insights to effectively address future viral emergence scenarios.

**2. Material and methods**

**2.1. Influent samples**

Influent samples from 15 urban WWTPs located in France (metropolitan France and French Guiana) were collected during 24 h periods using automated samplers. The characteristics of the WWTPs (location, treatment capacity) and the number of samples per WWTPs used for each section of the study are described in Table 1. ‘‘N’’ refers to the number of influent samples per WWTP, while ‘‘n’’ refers to the number of analyses per protocol, as some influent samples were analyzed using multiple protocols.

**2.2. Ultrafiltration concentration protocols**

Two methods based on the frontal UF approach were used to concentrate virus from the influent samples: stirred UF (SUF) and centrifugal UF (CUF) (Table 2).

The CUF method used a Centricon® Plus-70 ultrafilter with a cut-off of 100 kD (Merck Millipore) to process 100 mL or 200 mL influent samples, as described by Bertrand et al. (2021). Then, the concentrates underwent the NS3 RNA extraction protocol (CUF-P1 in Table 2).

The SUF used the Amicon® Stirred Cell 400 mL system (model 8400, UFSC40001, Merck Millipore) and Ultracel® membrane discs of regenerated cellulose with a cut-off of 10 kD (PLGC07610, Merck Millipore). The SUF method was performed as described (El Soufi et al., 2024) on 200 mL or 500 mL of influent samples. Concentrates obtained from 200 mL of influent were subjected to the NS3 RNA extraction protocol (SUF-

**Table 1**  
Influent samples and related WWTP.

WWTP	Location (region)	Capacity (inhabitants equivalent)	Sampling period	Number of samples (N)
Comparison of concentration methods (Section 3.1) and sequencing analysis (Section 3.2)				
WWTP-1	Grand Est	400,000	January–September 2023	23
WWTP-3	Nouvelle Aquitaine	135,000	January–February 2023 / September 2023	7
WWTP-10	Nouvelle Aquitaine	< 20,000	January–February 2023	1
WWTP-11	Provence -Alpes-Côte d’Azur	340,000	September–October 2023	4
WWTP-12	Normandie	550,000	August–September 2023	6
WWTP-13	Grand Est	102,000	September–October 2023	4
WWTP-14	French Guiana	30,000	August–September 2023	5
WWTP-15	French Guiana	60,000	August–September 2023	3

**Table 2**

Viral concentration methods and RNA extraction protocols applied on large volumes of influent (100 mL to 500 mL). The genome quantification methods and the number of samples for each protocol and laboratory are also given.

Protocol	CUF-P1			SUF-P2	SUF-P3	
Concentration method	Centrifugal UF			Stirred UF	Stirred UF	
Laboratory	Lab 1	Lab 2	Lab 2	Lab 1	Lab 1	Lab 3
Volume of influent (mL)	200	200	100	200	500	500
Volume of concentrate (mL)	5	5	5	5	5	5
RNA extraction method	NS3 - 10 mL of NucliSENS (bioMérieux) lysis buffer - Neutral phenol chloroform treatment - 140 µL of silica beads - OneStep PCR Inhibitor Removal kit (Zymo research)				NS4 - 10 mL of NucliSENS (bioMérieux) lysis buffer - Neutral phenol chloroform treatment (twice) - Lysate splitted into 3 fractions - 140 µL of silica beads / fraction - 300 µL of eluate - RNA Clean & Concentrator™ (Zymo research)	
Volume of extracted RNA (µL)	100	100	100	100	100	100
RNA quantification method for SARS-CoV-2	C	A	A	C	C	B
RT-qPCR kit for FRNAPH GGII	A	A	A	A	A	B
Number of samples for recovery rate determination of concentration method	N = 6	N = 3	N = 3	N = 7	N = 8	N = 5
Number of samples for RT-qPCR inhibition evaluation	N = 9	N = 3	N = 3	N = 7	N = 9	N = 5
Number of samples for sequencing analysis	N = 7	N = 17	NT	N = 17	N = 12	N = 16

A: RNA UltraSens™ One-Step Quantitative RT-PCR kit / B: Master Mix Fast Virus 1-Step TaqMan™ kit / C: One-Step RT-ddPCR Advanced kit for probes / NT: Not tested.

P2 in Table 2) and concentrates obtained from 500 mL of influent were subjected to the NS4 RNA extraction protocol (SUF-P3 in Table 2).

The NS3 and NS4 RNA extraction protocols are described below (see Section 2.3).

**2.3. RNA extraction protocols for concentrates obtained from large influent samples**

Two RNA extraction protocols, NS3 and NS4, were specifically used for concentrates obtained from 100 mL to 500 mL of influent, as described in Table 2. These protocols are quite similar to the one previously described by Bertrand et al. (2021) and represent modified versions of the NS2 protocol, which is described in Chaqroun et al., under review. They were based on the NucliSENS® kit (bioMérieux) using magnetic silica beads and on neutral phenol-chloroform (PC) treatment. For NS3, the volume of silica beads was increased from 70 µL to 140 µL. For NS4, the neutral PC treatment was applied twice, and the aqueous phase was divided into three fractions, each mixed with 140 µL of silica beads. The total eluate of 300 µL was subsequently cleaned and concentrated to 100 µL using the RNA Clean & Concentrator™ from Zymo Research. All extracted RNA samples were subsequently stored at  $-80\text{ }^\circ\text{C}$  until further analysis.

#### 2.4. Estimation of the recovery rates of ultrafiltration methods

In parallel to the UF methods applied to 100 mL to 500 mL of influent, 5 mL of unconcentrated influent was subjected to direct RNA extraction using the NS2 protocol (Chaqroun et al., under review) and served as a reference in order to calculate the viral RNA recovery rate as follows:

$$R = \frac{\frac{\text{Genome copies (GC)}}{\text{liter (L)}} \text{ of concentrated ww}}{\frac{\text{GC}}{L} \text{ of unconcentrated ww}} * 100$$

In influent samples with low prevalence of SARS-CoV-2 (below  $5 \times 10^4$  GC/L), the unconcentrated 5 mL of influent underwent the NS2 RNA extraction protocol (Chaqroun et al., under review), followed by a re-concentration step using the RNA Clean & Concentrator kit (R1016, Zymo Research), resulting in a final eluate volume of 6  $\mu$ L to be able to calculate the viral RNA recovery rate.

#### 2.5. Viral genome quantification

SARS-CoV-2 RNA quantification was performed by targeting the E gene with the PCR system developed by [Corman et al. \(2020\)](#) and using either RT-qPCR or RT-digital droplet PCR (RT-ddPCR) methods ([Table 2](#)).

In the case of RT-qPCR, two different kits from Applied Biosystems™ were used: either the RNA UltraSens™ One-Step Quantitative RT-PCR kit (Lab 1 and 2) or the Master Mix Fast Virus 1-Step TaqMan™ kit (Lab 3). For the RNA UltraSens™ One-Step Quantitative RT-PCR kit, the reaction was executed in a final volume of 20  $\mu$ L, containing 2  $\mu$ L of RNA, 1  $\mu$ M of primers, and 0.3  $\mu$ M of the probe. Amplification reactions were carried out on a BioRad CFX96™ thermal cycler with the following cycling parameters: 30 min at 50 °C for the reverse transcription step, followed by 5 min at 95 °C, and then 45 cycles of 15 s at 95 °C and 40 s at 58 °C for PCR amplification. For the Master Mix Fast Virus 1-Step TaqMan™ kit, 5  $\mu$ L of RNA was used in a final reaction volume of 10  $\mu$ L. The primer concentration was 0.4  $\mu$ M, and the probe concentration was 0.2  $\mu$ M. Amplification reactions were also conducted on a BioRad CFX96™ thermal cycler, using the following cycling parameters: 50 °C for 5 min (reverse transcription), followed by an initial denaturation cycle at 95 °C for 20 s, and subsequently, 45 cycles at 95 °C for 5 s (denaturation) and 58 °C for 40 s (annealing/extension). The quantification was performed using a standard curve based either on full-length amplicon cloned into pCR2.1 plasmid (Invitrogen, #452640) ([Wurtzer et al., 2021](#)) for laboratories 1 and 2, or a synthetic quantified SARS-CoV-2 RNA (Biorad, COV019) for laboratory 3. The Limit of Detection (LOD) and the Limit of Quantification (LOQ) of RT-qPCR assays were determined using replicates of 10-fold dilutions of SARS-CoV-2 RNA control. The LOD was defined as the lowest RNA target concentration that achieved 95 % detection success ([Borchardt et al., 2021](#); [Bustin et al., 2009](#); [Klymus et al., 2020](#)). With the RNA UltraSens™ One-Step Quantitative RT-PCR kit and synthetic SARS-CoV-2 RNA Control 2 from Twist Bioscience (ref MN908947.3), the LOD was set at 10 GC/reaction (6 replicates). Using the Master Mix Fast Virus 1-Step TaqMan™ kit and synthetic SARS-CoV-2 RNA from Bio-Rad (ref COV019), the LOD was established at 5 GC/reaction (10 replicates) ([El Soufi et al., 2024](#)). The LOQ is defined by the user since there is currently no standardized maximum coefficient variation (CV) for qPCR ([Borchardt et al., 2021](#); [Klymus et al., 2020](#)). With the RNA UltraSens™ One-Step Quantitative RT-PCR kit and synthetic SARS-CoV-2 RNA Control 2 from Twist Bioscience (ref MN908947.3), the LOQ was set at 10 GC/reaction (6 replicates) with a CV of 25 %. Using the Master Mix Fast Virus 1-Step TaqMan™ kit and synthetic SARS-CoV-2 RNA from Bio-Rad (ref COV019), the LOQ was established at 10 GC/reaction (10 replicates) with a CV of 25 %.

Regarding RT-ddPCR, the One-Step RT-ddPCR Advanced kit for probes from Bio-Rad was used. In this protocol, the reaction was conducted within a total volume of 22  $\mu$ L, which included 4  $\mu$ L of RNA, 0.9

$\mu$ M of primers, and 0.2  $\mu$ M of the probe. The samples were placed in the droplet generator with 70  $\mu$ L of generator oil to generate up to 20,000 droplets per sample. The resulting picolitre droplet emulsions (40  $\mu$ L) were transferred to a Veriti 96-Well Thermal Cycler (Applied Biosystems) and followed these cycling parameters: 1 h at 50 °C (reverse transcription) and 10 min at 95 °C, then 40 cycles of 30 s at 95 °C (denaturation) and 1 min at 58 °C (annealing/extension), followed by 10 min at 98 °C (enzyme deactivation). After amplification, the plate was transferred to the QX200™ Droplet Reader (Bio-Rad). The QuantaSoft™ Software (Bio-Rad) was then used to determine the initial concentration of each target RNA. This determination was based on counting the number of positive droplets per well using fluorescence amplitude detection and Poisson distribution modeling. For a result to be considered interpretable, at least 10,000 accepted droplets must be achieved. Additionally, the negative controls (PCR negative control and extraction negative control) should not contain any positive droplets and were also used for threshold setting ([Huggett, 2020](#)). Finally, if the sample contained one or more positive droplets, it was considered as positive. If the sample contained three or more positive droplets, it was considered quantifiable. Using synthetic SARS-CoV-2 RNA Control 2 from Twist Bioscience (ref MN908947.3), the LOD (detection criteria >95 %) and LOQ (CV of 40 %) were determined to be 6 GC/reaction (6 replicates).

The FRNAPH-II genome was quantified using an RT-qPCR assay (VTB4-Fph II set, [Wolf et al., 2010](#)). These non-enveloped phages are naturally present in influents at concentrations ranging from  $10^5$  to  $10^9$  GC/L ([Hartard et al., 2015](#); [Hata et al., 2013](#)). FRNAPH-II was used as a control for the concentration and/or extraction processes ([Section 3.1](#)). Additionally, it served as an inhibition control to assess RT-qPCR inhibition in the extracted RNA samples ([Section 3.1](#)) through RNA quantification in both undiluted extracts and 1/10 diluted extracts. For the RNA UltraSens™ One-Step Quantitative RT-PCR kit (Lab 1 and 2), identical conditions to those specified above for SARS-CoV-2 were applied. Regarding the Master Mix Fast Virus 1-Step TaqMan™ kit (Lab 3), the only difference with the protocol applied for SARS-CoV-2 was the amount of RNA used: only 2.5  $\mu$ L of RNA was employed with a final volume of 10  $\mu$ L. The amplification reaction was performed in fast cycling mode using the following conditions: 50 °C for 5 min (reverse transcription), followed by 95 °C for 20 s (reverse transcriptase inactivation/initial denaturation). Subsequently, there were 45 cycles at 95 °C for 3 s (denaturation) and 58 °C for 30 s (annealing/extension).

RT-qPCR inhibition was estimated for each sample of extracted RNA by comparing the quantification cycle (Cq) values obtained for the undiluted extract with those of the 1/10 diluted extract. If the difference between the two Cq values ( $\Delta$ Cq) reached 2.3, then the inhibition was 50 %. If this difference reached 1.3, then the inhibition was 75 %. According to [ISO 15216-1 \(2017\)](#) on the genome quantification of noroviruses and hepatitis A virus in food and bottled water matrices, up to 75 % inhibition is acceptable.

#### 2.6. The whole genome sequencing of concentrates from large volume of influent

Next-generation sequencing analyses were conducted using the Illumina COVIDSeq kit (Illumina, Inc.). In this process, 8.5  $\mu$ L of RNA, extracted from influent samples, underwent reverse transcription and subsequent enrichment for the SARS-CoV-2 genome via two multiplex PCR reactions (using ARTIC v4.1 primers) with 35 cycles each. The amplicons were then tagmented using EBLTS HT Beads, and indexed in a PCR with 7 cycles. The final libraries were pooled by volume, the pool was purified with Illumina Tune Beads and quantified using a Qubit dsDNA HS Assay kit (Thermo Fisher Scientific, Inc.). The average fragment size determined through LabChip GX analysis was 479 bp. Short-read sequencing was performed using the NovaSeq 6000 SP Reagent Kit v1.5, 300 cycles (Illumina, Inc.). To evaluate the depth and breadth of coverage, we used our in-house bioinformatic pipeline ([Gerber et al.,](#)

2022), with down-sampling to 7 M read pairs for each library and with filtering of reads <40 bp after clipping. Variants were filtered using the 'bcftools-filter' command, keeping variants with BMF < 0.05 and MF < 0.3. Strain assignment was done using Pangolin database 4.3.1 (updated November 27, 2023) (O'Toole et al., 2021).

## 2.7. Statistical analysis

Statistical analyses were conducted using GraphPad - Prism software (version 9.5.0). For independent samples, the Kruskal-Wallis test ( $n > 2$  series) was used to compare viral recovery based on the size of the influent sample. Additionally, the Mann-Whitney test ( $n = 2$  series) was used to assess the respective recovery levels of the SUF and CUF methods, as well as the differences between enveloped and non-enveloped viruses (Section 3.1), and to compare sequencing parameters (Section 3.2). For paired samples, the Friedman test ( $n > 2$  series) was used, to evaluate sequencing depth and coverage in terms of the analyzed volume (Section 3.2). Statistically significant results were indicated by  $p$ -values <0.05. The asterisks indicate significant differences (\* for  $p$ -value  $\leq 0.05$ ; \*\* for  $p$ -value  $\leq 0.01$ ; \*\*\* for  $p$ -value  $\leq 0.001$  and \*\*\*\* for  $p$ -value  $\leq 0.0001$ ), and 'ns' stands for non-significant differences.

## 2.8. Data smoothing

Two specific models were applied to smooth the data relating to the breadth of coverage and the median depth of sequencing. The first model applied to the coverage data is a mixture model combining the Gaussian distribution of measurements around a sigmoid curve and a uniform distribution. By choosing a distribution of measurements around a sigmoid, it is possible to impose the constraint that the smoothed values must lie between 0 and 100 %. Mixing this distribution with a uniform distribution can then be used to model the occurrence of outliers. The median depth was simply modeled as a Gaussian distribution of measurements around a sigmoid curve shifted by a constant factor. Detailed information about these models is available in Supplementary material.

## 3. Results

### 3.1. Evaluation of two ultrafiltration methods for large influent samples

In this paper, we evaluated the effectiveness of SUF and CUF methods for concentrating both enveloped (SARS-CoV-2) and non-enveloped (FRNAPH-II) RNA viruses from influent samples. The CUF approach was applied to influent samples of 100 mL and 200 mL and combined with the NS3 RNA extraction protocol (CUF-P1 in Table 2). The SUF approach was applied to influent samples of 200 mL (SUF-P2 in Table 2) and 500 mL (SUF-P3 in Table 2).

The NS2 protocol was the most suitable RNA extraction protocol for 5 mL of influent, and it also seemed to be the most easily adaptable to the concentrate samples obtained from large volumes of influent (100 to 500 mL) (Chaqroun et al., under review). To adapt NS2 to large volumes of influent, the volume of silica beads was increased (NS3 and NS4 RNA extraction protocols) and the concentrate was divided into three fractions before adding silica beads (NS4 RNA extraction protocol). Thus, the NS3 protocol was used for concentrates from 100 mL and 200 mL of influent (in CUF-P1 and SUF-P2), and the NS4 protocol was applied to concentrates from 500 mL of influent (in SUF-P3) (Table 2). Influent samples for this study were collected from WWTPs-1 and -3. For WWTP-1, the samples ( $N = 14$ , Table 1) were collected between January and July 2023, a period of low SARS-CoV-2 circulation in human populations. The SARS-CoV-2 RNA concentrations ranged from  $4.3 \times 10^3$  to  $4.8 \times 10^4$  GC/L. Influent samples from WWTP-3 ( $N = 3$ ) were collected during a period spanning from January to February 2023, and showed SARS-CoV-2 RNA concentrations ranging from  $1.8 \times 10^4$  to  $2.3 \times 10^5$

GC/L. At the same time as the concentration of these large volumes of influent, 5 mL of influent was used directly for RNA extraction using the NS2 protocol to assess the recovery rates of SARS-CoV-2 genome (Fig. 1A) and FRNAPH-II genome (data not shown).

The CUF-P1 protocol applied to 100 mL of influent showed recovery rates of  $27 \pm 21$  % for the SARS-CoV-2 genome (Fig. 1A) and  $82 \pm 79$  % for the FRNAPH-II genome (data not shown). These values were higher than those obtained with the same protocol applied to 200 mL of influent resulting in recovery rates of  $19 \pm 18$  % and  $47 \pm 50$  % for the SARS-CoV-2 genome (Fig. 1A) and the FRNAPH-II genome (data not shown), respectively. However, the difference was not significant for CUF applied to 100 mL or 200 mL of influent for both SARS-CoV-2 (Mann-Whitney test,  $p$ -value = 0.4818) and FRNAPH-II (Mann-Whitney test,  $p$ -value = 0.4637).

When processing 200 mL of influent, the SUF-P2 protocol showed a mean recovery rate of  $24 \pm 26$  % for the SARS-CoV-2 genome (Fig. 1A), and  $43 \pm 32$  % for the FRNAPH-II genome (data not shown). Increasing the volume to 500 mL and using the SUF-P3 protocol did not significantly affect the mean recovery rates since they reached  $31 \pm 29$  % for the SARS-CoV-2 genome (Mann-Whitney test,  $p$ -value = 0.4686) and  $38 \pm 36$  % for the FRNAPH-II genome (Mann-Whitney test,  $p$ -value = 0.8410), respectively.

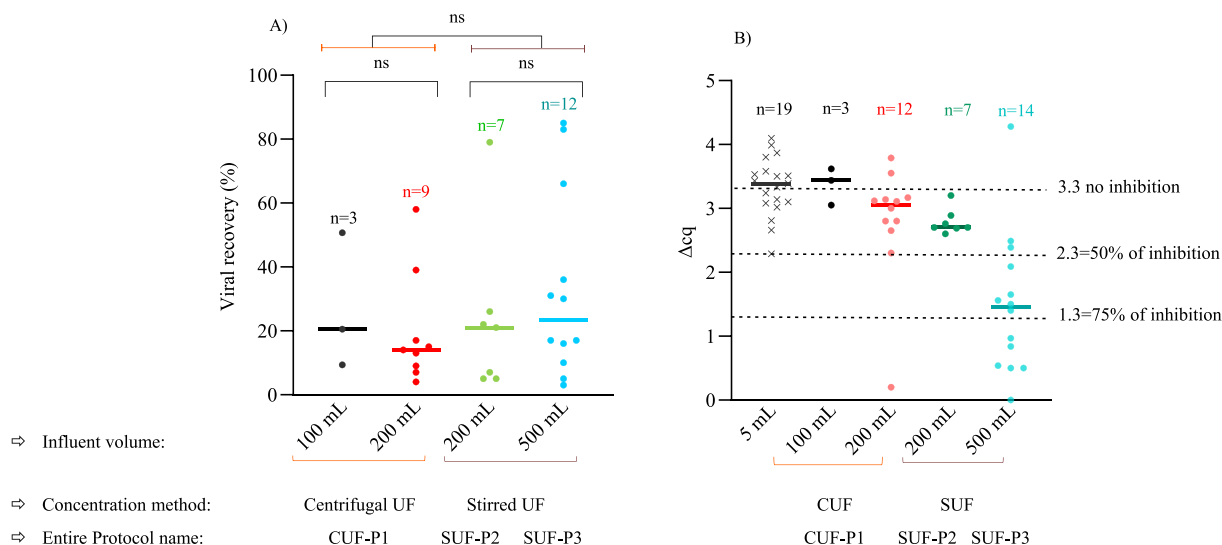
Likewise, no significant difference was observed between SUF and CUF concerning the recovery rate of both SARS-CoV-2 genome (Mann-Whitney test,  $p$ -value = 0.8563) and FRNAPH-II genome (Mann-Whitney test,  $p$ -value = 0.6665). The mean recovery rates of the non-enveloped virus (FRNAPH-II) were higher than those of the enveloped virus (SARS-CoV-2). However, this difference was not statistically significant ( $p = 0.0535$ , Mann-Whitney test).

We also assessed the presence of RT-qPCR inhibitors in both non-concentrated and concentrated WW samples (Fig. 1B). No inhibition was observed in the non-concentrated samples (5 mL). Among the concentrated samples (500 mL), inhibition was detected only in the SUF-P3 sample series, for which a 75 % median inhibition rate was observed.

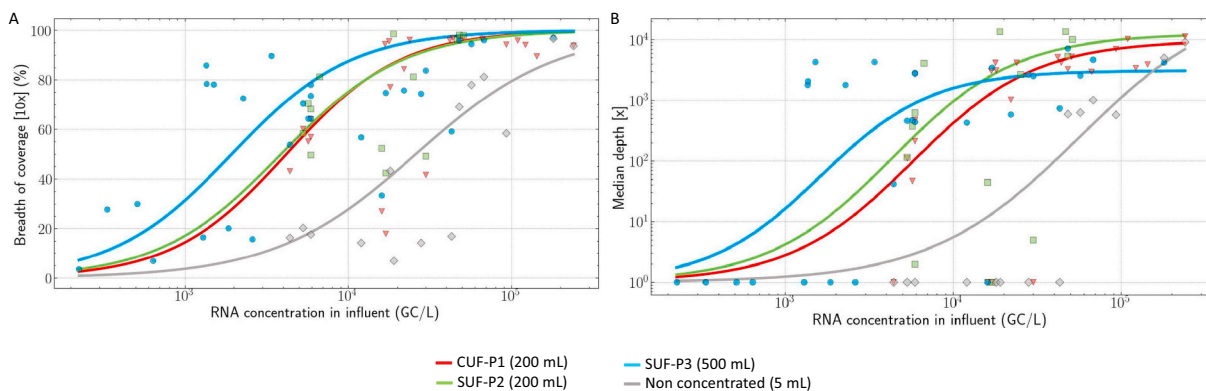
Based on recovery rates obtained for both SARS-CoV-2 and FRNAPH-II, we can conclude that the combination of a CUF or SUF method with the adapted NucliSENS extraction protocols proved effective for concentrating both enveloped and non-enveloped RNA viruses from influent samples of up to 500 mL. Analyzing 500 mL of influent could be considered the best option in terms of the quantity of concentrated RNA, but the analysis of this sample size is more frequently disturbed by RT-qPCR inhibitors that could also affect sequencing analysis.

### 3.2. Definition of the optimal concentration/extraction protocol for high-quality sequencing of the SARS-CoV-2 genome in influent

In this part of the study, a sequencing analysis was conducted for both non-concentrated and corresponding concentrated influent samples, collected between January 2023 and October 2023 from eight WWTPs (WWTP-1, -3, -10, -11, -12, -13, -14, and -15) (Table 1). Sequencing quality was evaluated by considering the breadth of the coverage and the median depth. A coverage of 50 % and a median depth of  $10\times$  are considered the minimal thresholds for satisfactory sequencing analysis of WW samples (Van Poelvoorde et al., 2021). Of the 79 samples that underwent sequencing, 71 % ( $n = 56/79$ ) exhibited both satisfactory coverage and depth values (Fig. 2A and B). Thus, the use of UF concentration methods enabled coverage rates of between 50 % and 99 % and median depth rates of between  $10\times$  and  $13,722\times$  (Fig. 2A and B). The majority of samples with insufficient coverage and depth belonged to the non-concentrated sample group ( $n = 8/14$ ). For the 6 non-concentrated samples with satisfactory values, the coverage ranged from 58 % to 97 %, and the median depth was between  $576\times$  and  $9177\times$  (Fig. 2A and B). Overall, coverage and depth values were significantly higher for concentrated samples compared to non-concentrated ones (Mann-Whitney test,  $p$ -value = 0.0159 and  $p$ -value = 0.00193, respectively).



**Fig. 1.** Comparison of stirred UF (Amicon® 10 kD) and centrifugal UF (Centricon® 100 kD) performed on 100 mL to 500 mL of influent. A) SARS-CoV-2 recovery rates (individual and median values). B) Analysis of RT-qPCR inhibitors by targeting FRNAPH-II genome.  $\Delta Cq$  values represent the difference between undiluted extract and the 1/10 diluted extract (individual values and median values). The asterisks represent significant differences. The Mann-Whitney test was used to compare the difference in viral recovery between the UF methods used (\*:  $p \leq 0.05$ ), the ns represents non-significant differences. UF: Ultrafiltration.



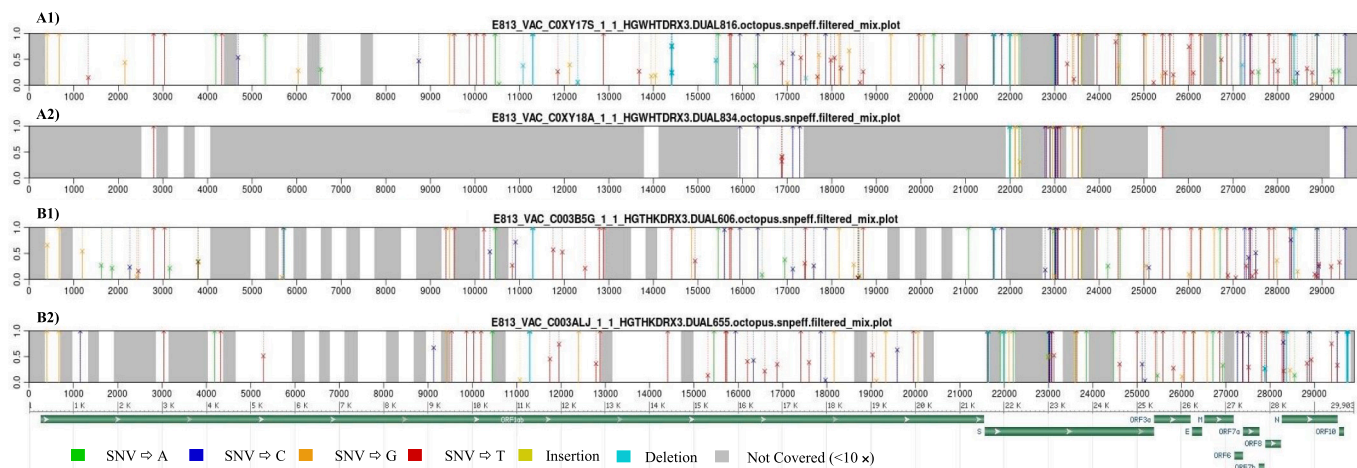
**Fig. 2.** Sequencing quality parameters of the SARS-CoV-2 RNA genome from influent samples ( $n = 79$  analyses from  $N = 43$ ) collected from January 2022 and October 2023 in 8 WWTPs. A) Sequencing coverage according to the viral RNA concentration in influent samples. B) Sequencing depth according to the viral RNA concentration in influent samples. Down-sampling was conducted to allow a comparison of the different methods due to sequencing performed on different flow cells.

The added-value of the concentration methods for sequencing is related to the initial concentration of SARS-CoV-2 genome in the influent samples. A significant difference was observed in both the breadth of coverage and the median depth of genome sequencing between non-concentrated and concentrated samples when the concentration of the SARS-CoV-2 genome was below  $5 \times 10^4$  GC/L (Mann-Whitney test,  $p$ -value = 0.0003 and  $p$ -value = 0.0007, respectively). However, when the concentration of the SARS-CoV-2 genome was higher than  $5 \times 10^4$  GC/L, the difference was no longer statistically significant for coverage (Mann-Whitney test,  $p$ -value = 0.2063) or median depth (Mann-Whitney test,  $p$ -value = 0.3095). Finally, even with 500 mL of influent, it was not possible to consistently achieve minimal sequencing analysis quality for SARS-CoV-2 when the concentration was below  $4 \times 10^3$  GC/L. Indeed, out of a total of 12 samples with concentrations below  $4 \times 10^3$  GC/L, only 42% (5/12) achieved coverage values ranging from 72 % to 90 %, with median depth values ranging from  $1000\times$  to  $10,000\times$ . Meanwhile, 58 % of these samples (7/12) showed coverage rates below 30 % and median depth values below  $10\times$  (Fig. 2A and B). As an example, for two separate samples showing the same RNA concentration of  $3 \times 10^3$  GC/L, the breadth of coverage was very wide: one reached 90 % (WWTP-14) while the other reached only 16 %

(WWTP-13) (Fig. 3A). Additionally, 500 mL of the same influent samples ( $N = 5$ ) collected between June and July 2023 in WWTP-1 were treated with the same protocol (SUF-P3) by laboratories L1 and L3. The resulting sequencing quality was similar with median breadth of coverage value ( $10\times$ ) reaching 80 % for Lab 1 and 63 % for Lab 3 (Fig. S1A, Wilcoxon test,  $p$ -value = 0.0625) and median depth value reaching  $5400\times$  for Lab 1 and  $3600\times$  for Lab 3 Fig. S1B, (Wilcoxon test,  $p$ -value = 0.2500). The variant calling profile of one of these five samples is also provided in Fig. 3B.

By analyzing the coverage profiles of the 79 samples represented in Fig. 2, we observed that some positions (4800, 20,000 and 25,500) in the genome show a fall in coverage across the samples, regardless of the method used (NC, CUF-200 mL, SUF-200 mL, or SUF-500 mL) (Fig. S2). This drop-in coverage is likely due to these positions being located in regions with high genetic variability or near deletions, which makes accurate detection using a PCR system challenging.

In order to determine the volume of influent to concentrate to ensure good quality sequencing, eight influent samples from WWTP-1 (with SARS-CoV-2 RNA concentrations not exceeding the threshold of  $5 \times 10^4$  GC/L) were analyzed using CUF on 200 mL (CUF-P1 protocol), SUF on 200 mL (SUF-P2 protocol), and SUF on 500 mL (SUF-P3 protocol). No



**Fig. 3.** Variant calling profiles. A) Comparison of variant calling profiles for two influent samples with similar SARS-CoV-2 RNA concentrations ( $3 \times 10^3$  GC/L), resulting in different breadths of coverage of 90 % (profile A1, sample from WWTP-14) and 16 % (profile A2, sample WWTP-13); B) Comparison of variant calling profiles of the same influent sample (WWTP-1) analyzed by laboratories 1 (profile B1) and 3 (profile B2) and resulting in similar breadths of coverage (81 % and 77 %, respectively).

significant differences in the breadth of coverage were observed ( $10\times$ ) (Friedman test,  $p$  value = 0.0789) (Fig. 4A). However, the median depth values were significantly higher when analyzing 500 mL than when analyzing 200 mL of influent (Friedman,  $p$  value = 0.0016) (Fig. 4B). Analyzing 500 mL of influent appears to be preferable, especially when seeking to enhance median sequencing depth. Nonetheless, opting for 200 mL could be beneficial in terms of cutting down processing time and minimizing the cost of sample handling and processing.

### 3.3. Sequencing for SARS-CoV-2 variant detection in influent from different cities

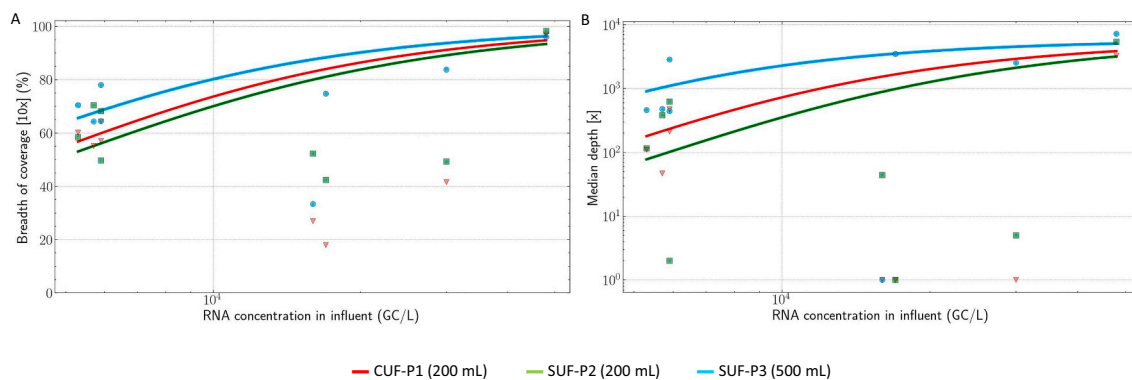
Among the samples discussed above, those with the highest values for breadth of coverage and mean depth were subsequently analyzed to detect SARS-CoV-2 variants. These samples had been collected between August and September 2023 in WWTP-1 ( $N = 5$ ), WWTP-3 ( $N = 4$ ), WWTP-11 ( $N = 4$ ), and WWTP-12 ( $N = 6$ ). The concentration of the SARS-CoV-2 genome was performed on 200 mL using either SUF (WWTP-1 and WWTP-11) or CUF (WWTP-3 and WWTP-12). The genome concentrations ranged from  $6.7 \times 10^3$  GC/L to  $5.0 \times 10^5$  GC/L. The breadth of coverage ranged from 96 % to 100 % (Fig. S3A) and the median depth was between  $6793\times$  and  $39,233\times$  (Fig. S3B). Variant EG.5, also known as Eris, consistently dominated, exhibiting a prevalence rate spanning from 66 % to 100 %.

### 4. Discussion

The most important factors to consider when addressing challenges related to the identification of emerging viruses in WW by sequencing are the volume of the sample being analyzed, the concentration methods of viral particles, and the RNA extraction protocol (Tamás et al., 2022). The significant influence of the concentration and RNA extraction protocols on the virome detected in raw sewage has been demonstrated by Hjelmsø et al. (2017).

Concentration methods are crucial for detecting low quantities of virus RNA in WW, especially during periods of low clinical incidence or low circulation of viruses in human populations (Gonçalves et al., 2023; Kaya et al., 2022). In our study, we employed two frontal UF methods on influent samples ranging from 100 mL to 500 mL, in conjunction with various adaptations of the RNA extraction protocol, which demonstrated effectiveness in extracting SARS-CoV-2 RNA. This RNA extraction protocol involves the treatment of samples with NPC and binding negatively charged nucleic acids to positively charged magnetic beads of silica in the presence of chaotropic salts (guanidium thiocyanate), followed by a final purification using an inhibitor removal kit based on resin columns.

Handling higher volumes of influent can negatively impact the recovery rate of viral RNA. Indeed, Peinado et al. (2022) detected higher quantities of SARS-CoV-2 RNA when concentrating 100 mL instead of 200 mL using CUF (Centricon® Plus-70, 10 kDa), meaning that the



**Fig. 4.** Comparison of the sequencing parameters between CUF applied to 200 mL of influent and SUF applied to 200 mL and 500 mL of influent ( $n = 8$ ). A) Comparison of sequencing coverage; B) Comparison of sequencing depth. Down-sampling was conducted as the sequencing was performed on different flow cells. CUF: Centrifugal Ultrafiltration, SUF: Stirred Ultrafiltration.



recovery rate was better with 100 mL than with 200 mL. In our study, increasing the influent sample volume to 500 mL when using the SUF method (10 kDa) did not impact the recovery rates for SARS-CoV-2 or FRNAPH-II. This may be related to the effective adjustment of the RNA extraction protocol for this volume by applying two NPC treatments and by a notable increase in the quantity of silica beads used per sample. [Tesauro et al. \(2022\)](#) observed an increase in the recovery of SARS-CoV-2 spiked in surface water by increasing the quantity of silica beads using the same RNA extraction protocol (NucliSENS). The modifications to the protocol for 500 mL samples of influent suggests that it would also be preferable to fractionate the concentrate for 200 mL samples, notably at low viral RNA concentrations, to further optimize viral RNA recovery. In the case of the CUF (100 kDa) method, increasing the volume of WW from 100 mL to 200 mL did not impact recovery rates for both viruses either. This may be partially due to the cutoff of the UF membrane (100 kDa), which could prevent filter clogging, thereby limiting the simultaneous concentration of WW pollutants ([Ahmed et al., 2020](#)).

Regarding the virus type, higher RNA recovery rates were observed for the non-enveloped virus (FRNAPH II) compared to the enveloped one (SARS-CoV-2), but the difference was not significant. The slight difference between these viruses may be related to the lower adsorption of FRNAPH-II to WW solids compared to enveloped viruses like SARS-CoV-2, as suggested by [Ye et al. \(2016\)](#). The same authors observed that 6 % of the non-enveloped bacteriophage MS2 (a representative member of FRNAPH-II) adsorbed to WW solids, whereas 26 % and 22 % of murine hepatitis viral particles and Phi 6 bacteriophage (enveloped viruses) adsorbed to WW solids, respectively. This suggests that the co-concentration of WW pollutants and solids could affect the concentration and RNA extraction steps, especially for enveloped viruses.

In sequencing studies, the sequencing depth was often not reported as frequently as coverage. However, when it was mentioned, the average depth ranged from  $7 \times$  to  $37,000 \times$  ([Child et al., 2023](#); [Crits-Christoph et al., 2021](#); [Fontenele et al., 2021](#); [Herold et al., 2021](#); [Levy et al., 2022](#); [Vo et al., 2022](#)). Moreover, the exact viral RNA concentrations required for the sequencing of WW samples were also frequently unspecified. According to studies that specified this, viral RNA concentrations were most often high and exceeded  $10^5$  GC/L ([Child et al., 2023](#); [Crits-Christoph et al., 2021](#); [Vo et al., 2022](#)). We observed that concentration methods for WW sequencing were not necessary when concentrations exceeded  $5 \times 10^4$  GC/L and that direct RNA extraction on 5 mL samples could be sufficient to obtain high-quality sequencing. This minimal concentration needed to sequence SARS-CoV-2 from 5 mL of unconcentrated influent agrees with the  $2 \times 10^5$  GC/L previously reported for a comparable extraction method from 5 mL samples ([El Soufi et al., 2024](#)). However, to our knowledge, the added value of these methods has not been evaluated in the literature data. In our study, the concentration methods (CUF or SUF) used in conjunction with the adapted RNA extraction protocols enabled us to achieve up to 99 % breadth of coverage and a median depth of up to  $32,953 \times$  for WW samples with SARS-CoV-2 genome concentrations of below 20 GC/ $\mu$ L of extracted RNA. Previous studies indicated that only samples with approximately  $\geq 20$  GC/ $\mu$ L of extracted RNA enabled a sequencing coverage of  $\geq 90$  % ([Crits-Christoph et al., 2021](#); [El Soufi et al., 2024](#); [Prado et al., 2021](#)), with a mean depth ranging from  $7 \times$  to  $107 \times$  ([Crits-Christoph et al., 2021](#)). It is worth mentioning that when the concentration of viral genetic material was below  $\sim 4 \times 10^3$  GC/L of influent sample, even with 500 mL of WW, only 42 % of the samples reached a coverage of  $>50$  % (70 % to 90 %) and a median depth of  $>10 \times$  ( $1000 \times$  to  $10,000 \times$ ). The high-quality sequencing achieved despite low concentrations of the SARS-CoV-2 genome could be related to various factors. These include the previously discussed adjustment of the extraction protocol, the diversity in the composition and physicochemical characteristics of WW samples—which led to variability in results across samples from different origins ([El Soufi et al., 2024](#); [Fontenele et al., 2021](#))—and, also, the sequencing process itself.

In terms of the volume of WW analyzed, 200 mL could be enough for

achieving high-quality sequencing, thereby reducing processing time and minimizing the cost associated with sample handling and processing. The use of 200 mL samples resulted in a breadth of coverage ranging from 96 % to 100 % and a median depth ranging from  $6793 \times$  to  $39,233 \times$  for samples with RNA concentration values ranging from  $6.7 \times 10^3$  GC/L to  $5 \times 10^5$  GC/L.

Sequencing for SARS-CoV-2 variant detection in diagnostic samples involves adhering to quality criteria established by the European Centre for Disease Prevention and Control (ECDC). While achieving (near)-complete genome coverage suffices for most genomic surveillance objectives, high sequencing depth is advised to effectively identify low-frequency variants, which is essential for studying direct transmission or reinfection. To precisely detect such variants with different levels of frequency, specific sequencing depths are necessary, namely mean values exceeding  $200 \times$  ([Child et al., 2023](#); [Van Poelvoorde et al., 2021](#)). In our study, we aimed to apply the same criteria to influent samples to ensure accurate sequencing analysis. Consequently, our samples that met this criterion were used to detect SARS-CoV-2 variants, both predominant variants and low-frequency variants. The analyses allowed the detection of variant EG.5, also known as Eris, and allowed us to determine that it was predominant.

## 5. Conclusion

The RNA extraction protocol based on silica beads and combined with neutral phenol-chloroform treatment (NS2) described in Chaqroun et al. (under review) for low volume (5 mL) of influent was adapted to concentrates obtained from large volumes of influent (up to 500 mL) by increasing the amount of silica beads and by splitting the concentrate lysate. The entire protocol allowed successful sequencing, but its added value varied based on the initial concentration of viral RNA in the influent. When the viral RNA concentration was above  $5 \times 10^4$  GC/L, concentration methods were not needed. For RNA concentrations between  $6 \times 10^3$  GC/L and  $5 \times 10^4$  GC/L, using a concentrate from 200 mL of influent was sufficient for successful sequencing. Below  $6 \times 10^3$  GC/L, analyzing 500 mL of WW showed satisfactory results in nearly half of the samples.

In the context of coronaviruses, the emergence of SARS-CoV-2 was followed by the accumulation of new mutations. This raises concerns about increased transmissibility, infectivity, and immunological resistance. Therefore, ongoing surveillance and research efforts are critical not only for monitoring SARS-CoV-2's genetic variability, but also for effectively addressing the challenges posed by the genetic evolution of coronaviruses and other emerging viruses ([Leifels et al., 2022](#); [Tosta et al., 2023](#)). Sequencing analysis of influent samples has to overcome the common technical challenge of fragmented genomes ([Otero et al., 2022](#)), especially during epidemic fluctuations and upon the emergence of new variants, when virus circulation is low. Consequently, the integrity of the viral genome needs further investigation to improve the quality of environmental sequencing analysis.

## CRedit authorship contribution statement

**Ahlam Chaqroun:** Writing – original draft, Visualization, Methodology, Investigation, Formal analysis. **Ghina El Soufi:** Writing – review & editing, Investigation, Formal analysis. **Zuzana Gerber:** Writing – review & editing, Visualization, Investigation, Formal analysis. **Julie Loutreul:** Investigation, Formal analysis. **Nicolas Cluzel:** Writing – review & editing, Visualization, Formal analysis. **Damien Delafoy:** Formal analysis. **Florian Sandron:** Formal analysis. **Léo Di Jorio:** Investigation. **Stéphanie Raffestin:** Resources. **Vincent Maréchal:** Writing – review & editing, Supervision, Funding acquisition. **Christophe Gantzer:** Writing – review & editing. **Robert Olasso:** Supervision. **Jean-François Deleuze:** Supervision, Funding acquisition, Conceptualization. **Olivier Rohr:** Writing – review & editing, Supervision, Resources, Funding acquisition, Conceptualization. **Nicolas Boudaud:**

Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization. **Clémentine Wallet:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Conceptualization. **Isabelle Bertrand:** Writing – review & editing, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

### Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Isabelle Bertrand, Olivier Rohr, Nicolas Boudaud reports financial support was provided by ANRS. Isabelle Bertrand, Christophe Gantzer, Vincent Marechal, Nicolas Boudaud, Olivier Rohr reports a relationship with OBEPINE that includes: board membership. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Data availability

Data will be made available on request.

### Acknowledgements

This work was performed within the scope of the EmerEaUde project (ANRS0160) funded by the ANRS-Maladies Infectieuses Emergentes and also supported by the French scientific group of interest GIS-OBEPINE and the French project OBEPINE+. This study was also supported by the UMT Actia VIROcontrol (Joint Technological Unit), a partnership tool shared by LCPME and Actalia, which has been established and supported by the French Ministry for Food. For LCPME laboratory, additional support was obtained from the Zone Atelier Moselle (ZAM). The UPR 9002 laboratory would like to thank the Agence de l'Eau Rhin Meuse, the Grand Est region, and the national research ANR agency (COVIDEU project) for their financial support, and the Syndicat des Eaux et de l'Assainissement Alsace-Moselle (SDEA) for the sampling efforts made in Alsace WWTPs.

### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.scitotenv.2024.175823>.

### References

- Ahmed, W., Bertsch, P.M., Bivins, A., Bibby, K., Farkas, K., Gathercole, A., Haramoto, E., Gyawali, P., Korajkic, A., McMinn, B.R., Mueller, J.F., Simpson, S.L., Smith, W.J.M., Symonds, E.M., Thomas, K.V., Verhagen, R., Kitajima, M., 2020. Comparison of virus concentration methods for the RT-qPCR-based recovery of murine hepatitis virus, a surrogate for SARS-CoV-2 from untreated wastewater. *Sci. Total Environ.* 739, 139960 <https://doi.org/10.1016/j.scitotenv.2020.139960>.
- Ahmed, W., Simpson, S.L., Bertsch, P.M., Bibby, K., Bivins, A., Blackall, L.L., Boffill-Mas, S., Bosch, A., Brandão, J., Choi, P.M., Ciesielski, M., Donner, E., D'Souza, N., Farnleitner, A.H., Gerrity, D., Gonzalez, R., Griffith, J.F., Gyawali, P., Haas, C.N., Shanks, O.C., 2022. Minimizing errors in RT-PCR detection and quantification of SARS-CoV-2 RNA for wastewater surveillance. *Sci. Total Environ.* 805, 149877 <https://doi.org/10.1016/j.scitotenv.2021.149877>.
- Bertrand, I., Challant, J., Jeulin, H., Hartard, C., Mathieu, L., Lopez, S., Schvoerer, E., Courtois, S., Gantzer, C., 2021. Epidemiological surveillance of SARS-CoV-2 by genome quantification in wastewater applied to a city in the northeast of France: comparison of ultrafiltration- and protein precipitation-based methods. *Int. J. Hyg. Environ. Health* 233, 113692. <https://doi.org/10.1016/j.ijheh.2021.113692>.
- Bofill-Mas, S., Rusiñol, M., 2020. Recent trends on methods for the concentration of viruses from water samples. *Curr. Opin. Environ. Sci. Health* 16, 7–13. <https://doi.org/10.1016/j.coesh.2020.01.006>.
- Borchardt, M.A., Boehm, A.B., Salit, M., Spencer, S.K., Wigginton, K.R., Noble, R.T., 2021. The environmental microbiology minimum information (EMMI) guidelines: qPCR and dPCR quality and reporting for environmental microbiology. *Environ. Sci. Technol.* 55 (15), 10210–10223.
- Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Wittwer, C.T., 2009. The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments.
- Cancela, F., Ramos, N., Smyth, D.S., Etchebehere, C., Berois, M., Rodríguez, J., Rufo, C., Alemán, A., Borzacconi, L., López, J., González, E., Botto, G., Thornhill, S.G., Mirazo, S., Trujillo, M., 2023. Wastewater surveillance of SARS-CoV-2 genomic populations on a country-wide scale through targeted sequencing. *PLoS One* 18 (4), e0284483. <https://doi.org/10.1371/journal.pone.0284483>.
- Carducci, A., Federigi, I., Liu, D., Thompson, J.R., Verani, M., 2020. Making waves: coronavirus detection, presence and persistence in the water environment: state of the art and knowledge needs for public health. *Water Res.* 179, 115907 <https://doi.org/10.1016/j.watres.2020.115907>.
- Cha, G., Graham, K.E., Zhu, K.J., Rao, G., Lindner, B.G., Kocaman, K., Woo, S., D'Amico, I., Bingham, L.R., Fischer, J.M., Flores, C.I., Spencer, J.W., Yathiraj, P., Chung, H., Biliya, S., Djeddar, N., Burton, L.J., Mascuch, S.J., Brown, J., Konstantinidis, K.T., 2023. Parallel deployment of passive and composite samplers for surveillance and variant profiling of SARS-CoV-2 in sewage. *Sci. Total Environ.* 866, 161101 <https://doi.org/10.1016/j.scitotenv.2022.161101>.
- Child, H.T., O'Neill, P.A., Moore, K., Rowe, W., Denise, H., Bass, D., Wade, M.J., Loose, M., Paterson, S., Van Aerle, R., Jeffries, A.R., 2023. Optimised protocol for monitoring SARS-CoV-2 in wastewater using reverse complement PCR-based whole-genome sequencing. *PLoS One* 18 (4), e0284211. <https://doi.org/10.1371/journal.pone.0284211>.
- Corman, V.M., Landt, O., Kaiser, R., Molenkamp, R., Meijer, A., Chu, D.K., Bleicker, T., Brünink, S., Schneider, J., Schmidt, M.L., 2020. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Eurosurveillance* 25 (3), 2000045.
- Crits-Christoph, A., Kantor, R.S., Olm, M.R., Whitney, O.N., Al-Shayeb, B., Lou, Y.C., Flamholz, A., Kennedy, L.C., Greenwald, H., Hinkle, A., Hetzel, J., Spitzer, S., Koble, J., Tan, A., Hyde, F., Schroth, G., Kuersten, S., Banfield, J.F., Nelson, K.L., 2021. Genome sequencing of sewage detects regionally prevalent SARS-CoV-2 variants. *mBio* 12 (1), e02703–e02720. <https://doi.org/10.1128/mBio.02703-20>.
- Daughton, C.G., 2020. Wastewater surveillance for population-wide Covid-19: the present and future. *Sci. Total Environ.* 736, 139631 <https://doi.org/10.1016/j.scitotenv.2020.139631>.
- El Soufi, G., Di Jorio, L., Gerber, Z., Cluzel, N., Van Assche, J., Delafoy, D., Olasso, R., Daviaud, C., Loustau, T., Schwartz, C., Trebouet, D., Hernalsteens, O., Marechal, V., Raffestin, S., Rousset, D., Van Lint, C., Deleuze, J., Boni, M., Rohr, O., Wallet, C., 2024. Highly efficient and sensitive membrane-based concentration process allows quantification, surveillance, and sequencing of viruses in large volumes of wastewater. *Water Res.* 249, 120959 <https://doi.org/10.1016/j.watres.2023.120959>.
- Feng, S., Owens, S.M., Shrestha, A., Poretzky, R., Hartmann, E.M., Wells, G., 2023. Intensity of sample processing methods impacts wastewater SARS-CoV-2 whole genome amplicon sequencing outcomes. *Sci. Total Environ.* 876, 162572 <https://doi.org/10.1016/j.scitotenv.2023.162572>.
- Fontenele, R.S., Kraberger, S., Hadfield, J., Driver, E.M., Bowes, D., Holland, L.A., Faleye, T.O.C., Adhikari, S., Kumar, R., Inchausti, R., Holmes, W.K., Deitrick, S., Brown, P., Duty, D., Smith, T., Bhatnagar, A., Yeager, R.A., Holm, R.H., Von Reitzenstein, N.H., Varsani, A., 2021. High-throughput sequencing of SARS-CoV-2 in wastewater provides insights into circulating variants. *Water Res.* 205, 117710 <https://doi.org/10.1016/j.watres.2021.117710>.
- Gerber, Z., Daviaud, C., Delafoy, D., Sandron, F., Alidjainou, E.K., Mercier, J., Gerber, S., Meyer, V., Boland, A., Bocket, L., Olasso, R., Deleuze, J.-F., 2022. A comparison of high-throughput SARS-CoV-2 sequencing methods from nasopharyngeal samples. *Sci. Rep.* 12 (1), 12561. <https://doi.org/10.1038/s41598-022-16549-w>.
- Gonçalves, J., Gomes Da Silva, P., Koritnik, T., Bosilj, M., Torres-Franco, A., Diaz, I., Rodríguez, E., Marcos, E., Mesquita, J.R., García-Encina, P., 2023. Quantification and whole genome characterization of SARS-CoV-2 RNA in wastewater and air samples. *J. Vis. Exp.* 196, 65053. <https://doi.org/10.3791/65053>.
- Hart, O.E., Halden, R.U., 2020. Computational analysis of SARS-CoV-2/COVID-19 surveillance by wastewater-based epidemiology locally and globally: feasibility, economy, opportunities and challenges. *Sci. Total Environ.* 730, 138875 <https://doi.org/10.1016/j.scitotenv.2020.138875>.
- Hartard, C., Rivet, R., Banas, S., Gantzer, C., 2015. Occurrence of and sequence variation among F-specific RNA bacteriophage subgroups in feces and wastewater of urban and animal origins. *Appl. Environ. Microbiol.* 81 (18), 6505–6515. <https://doi.org/10.1128/AEM.01905-15>.
- Hata, A., Kitajima, M., Katayama, H., 2013. Occurrence and reduction of human viruses, F-specific RNA coliphage genogroups and microbial indicators at a full-scale wastewater treatment plant in Japan. *J. Appl. Microbiol.* 114 (2), 545–554. <https://doi.org/10.1111/jam.12051>.
- Herold, M., d'Hérouël, A.F., May, P., Delogo, F., Wienecke-Baldacchino, A., Tapp, J., Walczak, C., Wilmes, P., Cauchie, H.-M., Fournier, G., Ogorzaly, L., 2021. Genome sequencing of SARS-CoV-2 allows monitoring of variants of concern through wastewater. *Water* 13 (21), 3018. <https://doi.org/10.3390/w13213018>.
- Hjelmso, M.H., Hellmér, M., Fernandez-Cassi, X., Timoneda, N., Lukjancenko, O., Seidel, M., Elsässer, D., Aarestrup, F.M., Löfström, C., Boffill-Mas, S., Abril, J.F., Girones, R., Schultz, A.C., 2017. Evaluation of methods for the concentration and extraction of viruses from sewage in the context of metagenomic sequencing. *PLoS One* 12 (1), e0170199. <https://doi.org/10.1371/journal.pone.0170199>.
- Huggett, J.F., 2020. The digital MIQE guidelines update: minimum information for publication of quantitative digital PCR experiments for 2020. *Clin. Chem.* 66 (8), 1012–1029.
- 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. Geneva (Switzerland).

- Kabdash, I., Tunay, O., 2021. Concentration techniques tailored for the detection of SARS-CoV-2 genetic material in domestic wastewater and treatment plant sludge: a review. *J. Environ. Chem. Eng.* 9 (5), 106296 <https://doi.org/10.1016/j.jece.2021.106296>.
- Karthikeyan, S., Levy, J.I., De Hoff, P., Humphrey, G., Birmingham, A., Jepsen, K., Farmer, S., Tubb, H.M., Valles, T., Tribelhorn, C.E., Tsai, R., Aigner, S., Sathe, S., Moshiri, N., Henson, B., Mark, A.M., Hakim, A., Baer, N.A., Barber, T., Knight, R., 2022. Wastewater sequencing reveals early cryptic SARS-CoV-2 variant transmission. *Nature* 609 (7925), 101–108. <https://doi.org/10.1038/s41586-022-05049-6>.
- Kaya, D., Niemeier, D., Ahmed, W., Kjellerup, B.V., 2022. Evaluation of multiple analytical methods for SARS-CoV-2 surveillance in wastewater samples. *Sci. Total Environ.* 808, 152033 <https://doi.org/10.1016/j.scitotenv.2021.152033>.
- Klymus, K.E., Merkes, C.M., Allison, M.J., Goldberg, C.S., Helbing, C.C., Hunter, M.E., Richter, C.A., 2020. Reporting the limits of detection and quantification for environmental DNA assays. *Environ. DNA* 2 (3), 271–282.
- Kuroiwa, M., Gahara, Y., Kato, H., Morikawa, Y., Matsui, Y., Adachi, T., Kurosawa, S., Kuroita, T., Ando, Y., Rokushima, M., 2023. Targeted amplicon sequencing of wastewater samples for detecting SARS-CoV-2 variants with high sensitivity and resolution. *Sci. Total Environ.* 893, 164766 <https://doi.org/10.1016/j.scitotenv.2023.164766>.
- Leifels, M., Khalilur Rahman, O., Sam, I.-C., Cheng, D., Chua, F.J.D., Nainani, D., Kim, S. Y., Ng, W.J., Kwok, W.C., Sirikanchana, K., Wuertz, S., Thompson, J., Chan, Y.F., 2022. The one health perspective to improve environmental surveillance of zoonotic viruses: lessons from COVID-19 and outlook beyond. *ISME Commun.* 2 (1), 107. <https://doi.org/10.1038/s43705-022-00191-8>.
- Levy, A., Gazeley, J., Lee, T., Jardine, A., Gordon, C., Cooper, N., Theobald, R., Huppatz, C., Sjollem, S., Hodge, M., Speers, D., 2022. Whole genome sequencing of SARS-CoV-2 from wastewater links to individual cases in catchments. *Sci. Total Environ.* 851, 158266 <https://doi.org/10.1016/j.scitotenv.2022.158266>.
- Li, L., Uppal, T., Hartley, P.D., Gorzalski, A., Pandori, M., Picker, M.A., Verma, S.C., Pagilla, K., 2022. Detecting SARS-CoV-2 variants in wastewater and their correlation with circulating variants in the communities. *Sci. Rep.* 12 (1), 16141. <https://doi.org/10.1038/s41598-022-20219-2>.
- Lin, K., Schulte, C.R., Marr, L.C., 2020. Survival of MS2 and  $\Phi$ 6 viruses in droplets as a function of relative humidity, pH, and salt, protein, and surfactant concentrations. *PLoS One* 15 (12), e0243505. <https://doi.org/10.1371/journal.pone.0243505>.
- Michael-Kordatou, I., Karaolia, P., Fatta-Kassinos, D., 2020. Sewage analysis as a tool for the COVID-19 pandemic response and management: the urgent need for optimised protocols for SARS-CoV-2 detection and quantification. *J. Environ. Chem. Eng.* 8 (5), 104306 <https://doi.org/10.1016/j.jece.2020.104306>.
- Otero, M.C.B., Murao, L.A.E., Limen, M.A.G., Caalim, D.R.A., Gaité, P.L.A., Bacus, M.G., Acaso, J.T., Miguel, R.M., Corazo, K., Knot, I.E., Sajonia, H., De Los Reyes, F.L., Jaraula, C.M.B., Baja, E.S., Del Mundo, D.M.N., 2022. Multifaceted assessment of wastewater-based epidemiology for SARS-CoV-2 in selected urban communities in Davao City, Philippines: a pilot study. *Int. J. Environ. Res. Public Health* 19 (14), 8789. <https://doi.org/10.3390/ijerph19148789>.
- O'Toole, Á., Scher, E., Underwood, A., Jackson, B., Hill, V., McCrone, J.T., Colquhoun, R., Ruis, C., Abu-Dahab, K., Taylor, B., Yeats, C., Du Plessis, L., Maloney, D., Medd, N., Attwood, S.W., Aanensen, D.M., Holmes, E.C., Pybus, O.G., Rambaut, A., 2021. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evol.* 7 (2), veab064 <https://doi.org/10.1093/ve/veab064>.
- Peinado, B., Martínez-García, L., Martínez, F., Nozal, L., Sánchez, M.B., 2022. Improved methods for the detection and quantification of SARS-CoV-2 RNA in wastewater. *Sci. Rep.* 12 (1), 7201. <https://doi.org/10.1038/s41598-022-11187-8>.
- Polo, D., Quintela-Baluja, M., Corbishley, A., Jones, D.L., Singer, A.C., Graham, D.W., Romalde, J.L., 2020. Making waves: wastewater-based epidemiology for COVID-19 – approaches and challenges for surveillance and prediction. *Water Res.* 186, 116404 <https://doi.org/10.1016/j.watres.2020.116404>.
- Prado, T., Fumian, T.M., Mannarino, C.F., Resende, P.C., Motta, F.C., Eppinghaus, A.L.F., Chagas Do Vale, V.H., Braz, R.M.S., De Andrade, J.D.S.R., Maranhão, A.G., Miagostovich, M.P., 2021. Wastewater-based epidemiology as a useful tool to track SARS-CoV-2 and support public health policies at municipal level in Brazil. *Water Res.* 191, 116810 <https://doi.org/10.1016/j.watres.2021.116810>.
- Rouchka, E.C., Chariker, J.H., Saurabh, K., Waigel, S., Zacharias, W., Zhang, M., Talley, D., Santisteban, I., Puccio, M., Moyer, S., Holm, R.H., Yeager, R.A., Sokoloski, K.J., Fuqua, J., Bhatnagar, A., Smith, T., 2021. The rapid assessment of aggregated wastewater samples for genomic surveillance of SARS-CoV-2 on a city-wide scale. *Pathogens* 10 (10), 1271. <https://doi.org/10.3390/pathogens10101271>.
- Sims, N., Kasprzyk-Hordern, B., 2020. Future perspectives of wastewater-based epidemiology: monitoring infectious disease spread and resistance to the community level. *Environ. Int.* 139, 105689.
- Smyth, D.S., Trujillo, M., Gregory, D.A., Cheung, K., Gao, A., Graham, M., Guan, Y., Guldenpfennig, C., Hoxie, I., Kannoly, S., Kubota, N., Lyddon, T.D., Markman, M., Rushford, C., San, K.M., Sompanya, G., Spagnolo, F., Suarez, R., Teixeira, E., Dennehy, J.J., 2022. Tracking cryptic SARS-CoV-2 lineages detected in NYC wastewater. *Nat. Commun.* 13 (1), 635. <https://doi.org/10.1038/s41467-022-28246-3>.
- Tamáš, M., Potocarova, A., Konecna, B., Klucar, L., Mackulak, T., 2022. Wastewater sequencing—an innovative method for variant monitoring of SARS-CoV-2 in populations. *Int. J. Environ. Res. Public Health* 19 (15), 9749. <https://doi.org/10.3390/ijerph19159749>.
- Tesauro, M., Terraneo, M., Consonni, M., Fappani, C., Colzani, D., Stevanin, C., Amendola, A., Masseroni, D., Tanzi, E., 2022. A methodological approach to water concentration to investigate the presence of SARS-CoV-2 RNA in surface freshwaters. *Pathogens* 11 (8), 845. <https://doi.org/10.3390/pathogens11080845>.
- Tosta, S., Moreno, K., Schuab, G., Fonseca, V., Segovia, F.M.C., Kashima, S., Elias, M.C., Sampaio, S.C., Ciccozzi, M., Alcantara, L.C.J., Slavov, S.N., Lourenço, J., Cella, E., Giovanetti, M., 2023. Global SARS-CoV-2 genomic surveillance: what we have learned (so far). *Infect. Genet. Evol.* 108, 105405 <https://doi.org/10.1016/j.meegid.2023.105405>.
- Van Poelvoorde, L.A.E., Delcourt, T., Coucke, W., Herman, P., De Keersmaecker, S.C.J., Saelens, X., Roosens, N.H.C., Vanneste, K., 2021. Strategy and performance evaluation of low-frequency variant calling for SARS-CoV-2 using targeted deep Illumina sequencing. *Front. Microbiol.* 12, 747458 <https://doi.org/10.3389/fmicb.2021.747458>.
- Vo, V., Tillett, R.L., Papp, K., Shen, S., Gu, R., Gorzalski, A., Siao, D., Markland, R., Chang, C.-L., Baker, H., Chen, J., Schiller, M., Betancourt, W.Q., Buttery, E., Pandori, M., Picker, M.A., Gerrity, D., Oh, E.C., 2022. Use of wastewater surveillance for early detection of Alpha and Epsilon SARS-CoV-2 variants of concern and estimation of overall COVID-19 infection burden. *Sci. Total Environ.* 835, 155410 <https://doi.org/10.1016/j.scitotenv.2022.155410>.
- Wolf, S., Hewitt, J., Greening, G.E., 2010. Viral multiplex quantitative PCR assays for tracking sources of fecal contamination. *Appl. Environ. Microbiol.* 76 (5), 1388–1394. <https://doi.org/10.1128/AEM.02249-09>.
- Wurtzer, S., Waldman, P., Ferrier-Rembert, A., Frenois-Veyrat, G., Mouchel, J.M., Boni, M., Moulin, L., 2021. Several forms of SARS-CoV-2 RNA can be detected in wastewaters: implication for wastewater-based epidemiology and risk assessment. *Water Res.* 198, 117183.
- Ye, Y., Ellenberg, R.M., Graham, K.E., Wigginton, K.R., 2016. Survivability, partitioning, and recovery of enveloped viruses in untreated municipal wastewater. *Environ. Sci. Technol.* 50 (10), 5077–5085. <https://doi.org/10.1021/acs.est.6b00876>.