

Article

Key SARS-CoV-2 Mutations of Alpha, Gamma, and Eta Variants Detected in Urban Wastewaters in Italy by Long-Read Amplicon Sequencing Based on Nanopore Technology

Giuseppina La Rosa ^{1,*} , David Brandtner ², Pamela Mancini ¹, Carolina Veneri ¹, Giusy Bonanno Ferraro ¹, Lucia Bonadonna ¹ , Luca Lucentini ¹ and Elisabetta Suffredini ³ 

¹ Department of Environment and Health, Istituto Superiore di Sanità, 00161 Rome, Italy; pamela.mancini@iss.it (P.M.); carolina.veneri@guest.iss.it (C.V.); giusy.bonannoferraro@iss.it (G.B.F.); lucia.bonadonna@iss.it (L.B.); luca.lucentini@iss.it (L.L.)

² Independent Researcher, 00156 Rome, Italy; restless.v2@gmail.com

³ Department of Food Safety, Nutrition and Veterinary Public Health, Istituto Superiore di Sanità, 00161 Rome, Italy; elisabetta.suffredini@iss.it

* Correspondence: giuseppina.larosa@iss.it; Tel.: +39-06-49902768



Citation: La Rosa, G.; Brandtner, D.; Mancini, P.; Veneri, C.; Bonanno Ferraro, G.; Bonadonna, L.; Lucentini, L.; Suffredini, E. Key SARS-CoV-2 Mutations of Alpha, Gamma, and Eta Variants Detected in Urban Wastewaters in Italy by Long-Read Amplicon Sequencing Based on Nanopore Technology. *Water* **2021**, *13*, 2503. <https://doi.org/10.3390/w13182503>

Academic Editors: Silvia Monteiro and Alba Pérez-Cataluña

Received: 14 August 2021

Accepted: 10 September 2021

Published: 13 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: The emergence of SARS-CoV-2 variants of concern (VOCs) and variants of interest (VOIs) poses an increased risk to global public health and underlines the need to prioritise monitoring and research to better respond to the COVID-19 pandemic. Wastewater monitoring can be used to monitor SARS-CoV-2 spread and to track SARS-CoV-2 variants. A long read amplicon sequencing approach based on the Oxford Nanopore technology, targeting the spike protein, was applied to detect SARS-CoV-2 variants in sewage samples collected in central Italy on April 2021. Next-generation sequencing was performed on three pooled samples. For variant identification, two approaches—clustering (unsupervised) and classification (supervised)—were implemented, resulting in the detection of two VOCs and one VOI. Key mutations of the Alpha variant (B.1.1.7) were detected in all of the pools, accounting for the vast majority of NGS reads. In two different pools, mutations of the Gamma (P.1) and Eta (B.1.525) variants were also detected, accounting for 22.4%, and 1.3% of total NGS reads of the sample, respectively. Results were in agreement with data on variant circulation in Italy at the time of wastewater sample collection. For each variant, in addition to the signature key spike mutations, other less common mutations were detected, including the amino acid substitutions S98F and E484K in the Alpha cluster (alone and combined), and S151I in the Eta cluster. Results of the present study show that the long-read sequencing nanopore technology can be successfully used to explore SARS-CoV-2 diversity in sewage samples, where multiple variants can be present, and that the approach is sensitive enough to detect variants present at low abundance in wastewater samples. In conclusion, wastewater monitoring can help one discover the spread of variants in a community and early detect the emerging of clinically relevant mutations or variants.

Keywords: SARS-CoV-2; nanopore sequencing; variant of concern; variant of interest; sewage; wastewater; long reads; mutations

1. Introduction

Coronavirus disease 2019 (COVID-19), the viral disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is the cause of the ongoing worldwide public health emergency. SARS-CoV-2 is defined by a variety of lineages harbouring distinctive genetic changes in the genome and, more significantly, in the spike (S) protein. Some fast-spreading SARS-CoV-2 variants have become the dominant circulating strains. The emergence of variants that posed an increased risk to global public health prompted the characterisation of specific variants of concern (VOCs) and variants of interest (VOIs) in order to prioritise global monitoring and research, and ultimately to

inform the ongoing response to the COVID-19 pandemic (tracking SARS-CoV-2 variants (who.int)). VOCs show clear evidence of either increased transmissibility, increased virulence, change in clinical disease presentation, decreased efficacy of public health measures, diagnostics, or vaccines. VOIs have mutations with established or suspected phenotypic implications that could imply an effect on transmissibility, severity and/or immunity, realistically having an impact on the epidemiological situation (SARS-CoV-2 variants of concern as of 18 June 2021 (<https://www.ecdc.europa.eu/en/covid-19/variants-concern>, accessed on 11 August 2021). Finally, variants under monitoring are those for which there is some indication that they could have properties similar to those of a VOC, but the evidence is still weak or an assessment has not yet been completed. On 31 May, 2021, WHO assigned simple labels for key variants of SARS-CoV-2 using letters of the Greek alphabet (WHO announces simple, easy-to-say labels for SARS-CoV-2 variants of interest and concern). As of August 2021, there are four VOCs identified by WHO (<https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/>, accessed on 11 August 2021): Alpha, Beta, Gamma, Delta (Pango lineages B.1.1.7, B.1.351, P.1, and B.1.617.2, respectively), and four VOIs: Eta, Iota, Kappa, and Lambda (Pango lineages B.1.525, B.1.526, B.1.617.1, and C.37). As of 6 July 2021, WHO had removed P.2 (Zeta), P.3 (Theta) and B.1.427/4.29 (Epsilon) from the VOI list. Taking into account the likelihood of having an impact on the epidemiological situation in the EU/EEA, slightly different assignments are reported by the European Centre for Disease Control (ECDC) as of 5 August 2021, with Theta and Epsilon variants still on the VOI list, and the Iota variant instead included among variants under monitoring (<https://www.ecdc.europa.eu/en/covid-19/variants-concern>, accessed on 11 August 2021).

WHO recommends that all countries increase the sequencing of SARS-CoV-2 where possible, and share data to help monitoring and responding to the evolving pandemic (SARS-CoV-2 Evolution (who.int)). As on 13 August, 2021, there are approximately 2.7 million SARS-CoV-2 sequences deposited in GISAID, the most popular data-sharing platform for SARS-CoV-2. In Europe, to prevent, mitigate and respond to the potential impact of SARS-CoV-2 variants, a European bio-defence preparedness plan called “HERA Incubator” was launched on 17 February, 2021. Furthermore, the European Commission has adopted recommendation 2021/472 to support member states in establishing wastewater surveillance systems across the EU as a complementary data collection and management tool of the COVID-19 pandemic, focusing on the emergence and spread of SARS-CoV-2 variants (https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=uriserv:OJ.L_.2021.098.01.0003.01.ENG, accessed on 11 August 2021). Indeed, numerous studies have shown that wastewater monitoring can be used to investigate SARS-CoV-2 diversity and to identify SARS-CoV-2 variants [1–13].

In a previous study, we described a long nested RT-PCR assay targeting the spike protein (~1600 bps, covering amino acids 58–573), combined with Sanger sequencing [6]. This approach allowed the detection of panels of multiple nucleotide mutations distinctive of Alpha and Gamma SARS-CoV-2 variants as well as of the Pango lineage B.1.177 (20E.EU1). However, since classical Sanger sequencing on PCR amplicons may underestimate the existence of possibly less common sequences, in this study we aimed to combine the protocol with NGS for a more in-depth analysis, as previously successfully applied to enteric viruses [14–16]. The NGS technologies more commonly used, requiring sequencing of short DNA fragments and assembling in unique contigs, display some limitations in complex matrices like urban wastewaters, where a mixture of different genomes is frequently present. Using this technology, indeed, due to DNA fragmentation, it is not possible to pinpoint co-occurrence of mutations on the same DNA strand, while it is possible to detect signature mutations of SARS-CoV-2 variants by considering the frequency of mutations within a sample. The introduction of single-molecule sequencing platforms, such as single-molecule sensing technologies on the Oxford Nanopore Technologies (ONT) MinION platform, has opened the way to obtaining ultra-long reads. Herein we applied the ONT technology to

identify a panel of key spike mutations, in order to discriminate the different SARS-CoV-2 variants in wastewater samples.

2. Materials and Methods

2.1. Sampling, Nucleic Acids Extraction, Real-Time RT-qPCR, RT-Nested-PCR, and Sanger Sequencing

Twelve samples were collected in three wastewater Treatments Plants located in the Province of Latina, Latium, Central Italy between 28 April and 28 May 2021. These WTPs were selected since a high-risk COVID red zone was declared in that area, due to the high proportion of local Indian farm labourers who have tested positive for the virus and the concern about the possible circulation of the Delta variant (COVID: red zone declared near Latina as dozens of Indians test positive-English-ANSA.it).

Before viral concentration, samples underwent a 30 min treatment at 56 °C to increase the safety of the analytical protocol for the laboratory personnel and environment, and were then concentrate using polyethylene glycol (PEG)-based concentration, according to the protocol by Wu et al. (2020) [17], with slight modifications. Briefly, sewage samples (45 mL) were centrifuged at $4500\times g$ for 30 min; after centrifugation 40 mL of sample were mixed with polyethylene glycol 8000 8% (wt/vol) and NaCl (0.3 M) (both supplied by Sigma-Aldrich, St. Louis, MO, USA), spiked with a known quantity of mengovirus, used as a process control; sample was then centrifuged at $12,000\times g$ for 2 h and the viral pellet was resuspended in 2 mL NucliSENS Lysis Buffer reagent (bioMerieux, Marcy-l'Étoile, France) for RNA extraction. Nucleic acids were extracted using the MiniMAG extraction system (bioMerieux) and eluted RNA (100 µL) was further purified from PCR inhibitors using the OneStep PCR Inhibitor Removal Kit (Zymo Research, Irvine, CA, USA) and stored at $-80\text{ }^{\circ}\text{C}$ until molecular analysis. Viral recovery was assessed by mengovirus analysis as described in Costafreda et al. (2006) [18]. Screening for SARS-CoV-2 detection was performed by real-time RT-qPCR as reported in La Rosa et al. (2021) [19]. Conventional RT-PCR was performed using the SuperScript IV OneStep RT-PCR System with Platinum SuperFi RT-PCR Master mix (Invitrogen, Carlsbad, CA, USA) and previously described primers amplifying a long fragment (1592 bps) of the spike region [6]. Amplifications were performed in a T100 PCR thermal cycler (Bio-Rad Laboratories, Hercules, CA, USA). Mutations detectable by the long PCR assays for VOCs and VOIs variants are shown in Table 1.

Amplification conditions were as follows: reverse transcription at 45 °C for 10 min, denaturation at 98 °C for 2 min, followed by 35 cycles of 98 °C for 10 s, 58 °C for 10 s, 72 °C for 1 min (PCR ID 979), and a final extension at 72 °C for 5 min. PCR reactions was performed using 2 µL of each primer (10 µM) and 8 µL of RNA in a final volume of 50 µL. After the first round PCR, nested PCR (PCR ID 980) was performed in a 50 µL volume using the Phusion Hot Start II DNA Polymerase with GC buffer (Thermo Fisher Scientific, Waltham, MA, USA), 4 µL of the first PCR product, 2 µL of each primer (10 µM) and the following conditions: initial denaturation at 98 °C for 30 s, followed by 45 cycles of 98 °C for 10 s, 62 °C for 30 s, 72 °C for 1 min, and a final extension at 72 °C for 10 min. Standard precautions were taken to avoid laboratory contamination. RNA from SARS-CoV-2 Wuhan prototype strain (provided by the Robert Koch Institute under the RefBio project) and molecular grade water were included in each PCR run as positive and negative controls, respectively. PCR products were observed by gel electrophoresis (2% agarose gel, stained with GelRed, Biotium; Fremont, CA, USA). Positive PCR products were purified using Montage PCRm96 Micro-Well Filter Plate (Millipore, Burlington, MA, USA) and were sequenced on both strands (Bio-Fab Research, Rome, Italy, and Eurofins Genomics, Ebersberg, Germany). Positive PCR products confirmed by Sanger sequencing were mixed in three different pools (barcode BC01, BC02, and BC03) based on sampling points (see Table 2) for subsequent nanopore sequencing.

Table 1. Mutations detectable by sequencing of the PCR assay amplicon.

Alpha (B.1.1.7)		Beta (B.1.351)		Gamma (P.1)		Delta (B.1.617.2)		Eta (B.1.525)		Kappa (B.1.617.1)		Lambda (C.37)		Lota (B.1.526)	
Mutation	%	Mutation	%	Mutation	%	Mutation	%	Mutation	%	Mutation	%	Mutation	%	Mutation	%
H69del	94%	D80A	87%	D138Y	93%	K77T	0.8%	A67V	95%	T95I	45%	G75V	85%	T95I	74%
V70del	95%	D215G	85%	R190	92%	T95I	41%	H69-	92%	G142D	54%	T76I	86%	D253G	74%
Y144del	94%	L241-	77%	K417T	95%	G142D	54%	V70-	92%	E154K	67%	R246-	76%	S477N	29%
E484K	0.2%	L242-	78%	E484K	96%	E156G	88%	Y144-	92%	L452R	93%	S247-	76%	E484K	42%
S494P	0.3%	L243-	77%	N501Y	96%	F157-	89%	E484K	97%	E484Q	93%	Y248-	77%		
N501Y	97%	P384L	1.2%			R158-	88%					L249-	77%		
A570D	98%	K417N	88%			W258L	0.2%					T250-	77%		
		E484K	87%			K417N	0.4%					P251-	77%		
		N501Y	87%			L452R	97%					G252-	77%		
		E516Q	0.3%			T478K	97%					D253N	77%		
												L452Q	90%		
												F490S	90%		

Percentage represent the frequency of mutation occurrence in each variant according to GISAID data (data extracted on 6 August, 2021). Key mutations used for the BAMQL queries are reported in bold.

Table 2. Samples under study and results obtained by Sanger and nanopore sequencing.

Sample ID	Sampling Location	Date	Sanger Sequencing		Nanopore Sequencing		
			Mutations Detected	Associated Variant	Pool Barcode	Clustering	Classification (BAMQL)
4158	Sabaudia (Belsito)	28 April 2021	D138Y, R190S, K417T, E484K, N501Y	Gamma	BC01	Alpha + Gamma	Alpha + Gamma
4159	Sabaudia (Belsito)	29 April 2021	H69del, V70del, Y144del, N501Y, A570D	Alpha			
4162	Sabaudia (Caprolace)	28 April 2021	S98F, H69del, V70del, Y144del, N501Y, A570D	Alpha	BC02	Alpha	Alpha + Eta
4163	Sabaudia (Caprolace)	29 April 2021	H69del, V70del, Y144del, N501Y, A570D	Alpha			
4172	Sabaudia (Caprolace)	4 May 2021	H69del, V70del, Y144del, N501Y, A570D	Alpha	BC03	Alpha	Alpha
4173	Sabaudia (Caprolace)	6 May 2021	H69del, V70del, Y144del, N501Y, A570D	Alpha			

2.2. Library Preparation of SARS-CoV-2 PCR Products for Nanopore Sequencing, Data Acquisition, Basecalling, and Bioinformatics Analyses

The workflow for library preparation, nanopore sequencing, and data analysis is shown in Figure 1. Nanopore sequencing was performed at Bio-Fab Research (Rome, Italy) using a MinION sequencer. Libraries were prepared using the cDNA-PCR Sequencing kit (SQK-PCS109) from Oxford Nanopore Technologies (Oxford, UK) following the manufacturer's instructions, combined with the Native barcoding (EXP-NBD104) kit. MinION flow cells (FLO-MIN106 R9.4.1) were loaded with 50 fmol of pooled libraries and sequencing was carried on over 17 h using the MinKNOW software 3.6.5. The raw data (FAST5 files) were base called using the high-accuracy models (dna_r9.4.1_450bps_hac.cfg) and then demultiplexed by Guppy basecalling suite v3.6.1 + 249406c on Ubuntu 18.04 machine, to obtain the final FASTQ files. Filtering parameters were applied to select only for reads preserving the 5' and 3' adaptors. Subsequently, reads were selected for a length range consistent with the PCR amplicon product (1400–1700 bases), and for an average quality equal or above 8. In order to eliminate non-specific signals, individual sequence reads were mapped to the SARS-CoV-2 reference sequence (NC_045512) and only reads aligned with this reference were used for further analysis. The resulting BAM file was the input of the subsequent analysis for amplicon assignment. Both unsupervised and supervised approaches were used for NGS reads assignment.

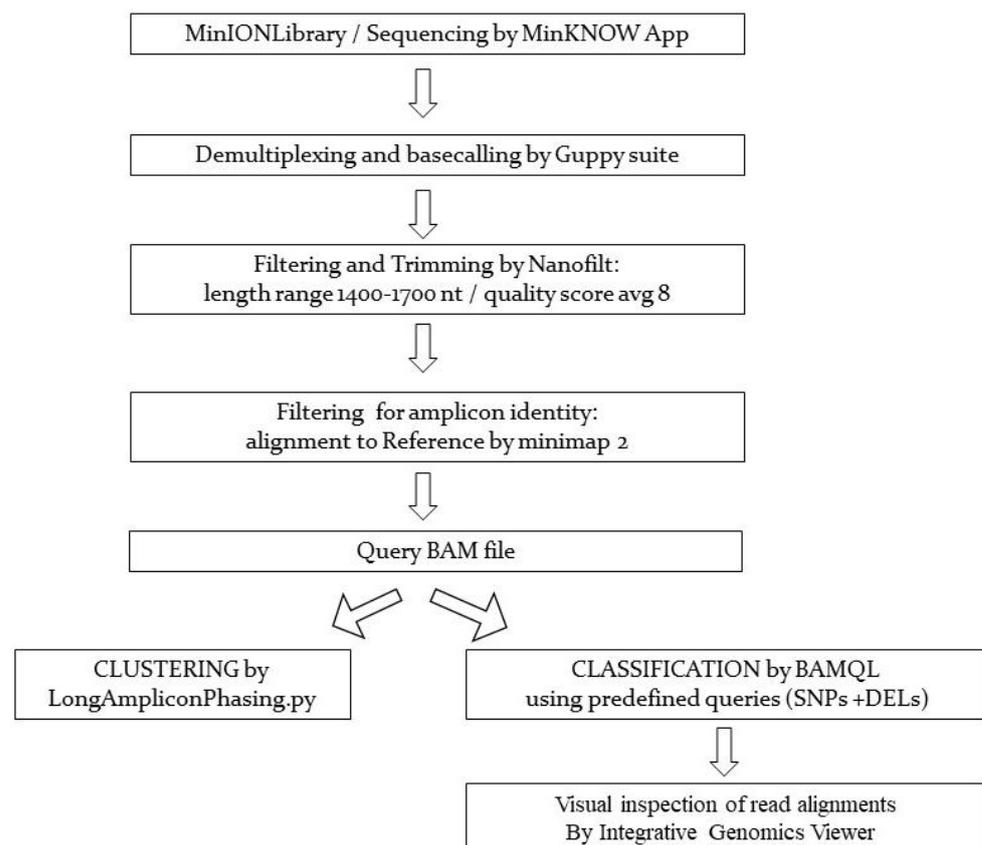


Figure 1. Workflow for library preparation, nanopore sequencing and data analysis.

The unsupervised clustering was performed by using the phasing by the alignment model of the LongAmpliconPhasing.py tool (<https://github.com/PacificBiosciences/pbampliconclustering>, accessed on 11 August 2021). Most similar reads were clustered, based on SNPs present in BAM files, calling a unique haplotype variant. Results were shown in an alleleClusterSummary file, which illustrates the different clusters and associated mutations.

The supervised classification approach was performed extracting reads subsets from BAM files, querying them with BAMQL tool v1.6 [20] using ad hoc command lines designed to search a panel of deletions and/or amino acid substitutions in linkage, each characteristic for specific VOCs or VOIs. Mutations included in the command lines for each VOC/VOI are illustrated in the Supplementary Material (Table S2). BAMQL reads subset outputs were loaded in IGV v2.9.4 for visual confirmation of the alignment and of selected mutations.

When the alignment showed the presence of additional mutations besides the key mutations associated with the selected variant (e.g., the sub-variant of the Alpha cluster, including the amino acid substitution E484K, or other), a second query was performed to extract the reads subsets containing the additional amino acid changes. Further to the presence of queried mutations, the visualization step in IGV was used to check for alignment anomalies, such as Ins/Del heavy patterns (see Supplementary Material, Figures S4 and S5) and positive/negative strand heavy unbalance. In presence of such unbalancing, the alignment was considered unreliable and therefore discarded. A sequencing coverage level of at least 100 reads after the bioinformatic analysis was required for variant assignment. Finally, a comparison between classification outputs (variants and sub-variants) was performed extracting read IDs from the BAMQL output and using them as input for the VIB web tool to draw Venn diagrams (Draw Venn Diagram (ugent.be)). Bioinformatics analysis was performed under Linux Ubuntu 18.04 machine.

3. Results

Viral recovery, assessed by mengovirus analysis was on average 2.9% (range 1.3–6.2%). SARS-CoV-2 RNA was detected in 11 of the 12 samples by real-time RT-qPCR, with concentrations ranging from 1.9×10^3 to 9.1×10^4 genome copies (g.c.)/L of sewage (see Supplementary Materials, Table S1). Amplification by the long RT-nested-PCR assay (ID 980) was achieved in 6 of the 12 samples, and resulting amplicons were sequenced by conventional Sanger sequencing. Mutations characteristic of both Alpha (B.1.1.7) and Gamma (P.1) VOCs were detected, in these sewage samples, as shown in Table 2. These positive samples were then combined into three pools for nanopore sequencing (pool BC01, BC02, and BC03), each containing two samples collected at the same WTP in close dates. A total of 1.367.887, 1.925.893, and 764.517 raw reads were obtained for barcoded pool BC01, BC02, and BC03, respectively. After selecting only reads preserving the 5' and 3' adaptors, trimming, filtering for fragment length, and aligning to the reference Wuhan prototype strain, a total of 312.697 reads were obtained for BC01, 328.701 for BC02, and 330.950 for BC03 (see Supplementary Material, Table S1). The number of reads assigned to each variant/sub-variant by the two approaches (clustering and BAMQL classification) is shown in Table 3.

Table 3. Number of reads assigned to each variant/sub-variant by the two bioinformatics analysis approaches.

Panel of Deletion and/or Amino Acid Substitutions in Linkage of Each Variant	Clustering			Classification (BAMQL)		
	BC01	BC02	BC03	BC01	BC02	BC03
Alpha (H69del, V70del, Y144del, N501Y, A570D)	11,055	70,714	85,971	54,489	85,191	105,315
Alpha + S98F (C21855T)	/	8604	/	/	38,460	/
Alpha + E484K (G23012A)	/	/	/	/	761	/
Alpha + S98F + E484K (C21855T + G23012A)	/	/	/	/	931	/
Gamma (D138Y, R190S, K417T, E484K, N501Y)	8443	/	/	69,980	/	/
Eta (A67V, H69del, V70del, Y144del, E484K)	/	/	/	/	4325	/
Eta + S151I (G22014T)	/	/	/	/	2036	/

3.1. Clustering Results

The results of clustering analysis using the LongAmpliconPhasing.py tool showed the presence of two clusters, corresponding to the Alpha (11,055 reads) and Gamma (8443) variants in pool BC01 (see Supplementary Materials, Figure S1); two clusters, Alpha

(70,714 reads) and Alpha with substitution S98F (C21855T) (8604 reads) in BC02; and only one cluster, the Alpha variant (85,971 reads) in BC03.

3.2. Classification Results

With the analysis using BAMQL queries (see Supplementary Materials, Table S2), a total of 54,489 reads and 69,980 reads were assigned to the Alpha and Gamma clusters, respectively, in pool BC01. In pool BC02, 85,191 reads were assigned to the Alpha cluster. Of these, 38,460 included also the mutation S98F (C21855T), 761 harboured the mutation E484K (G23012A), and 931 displayed both S98F and E484K (see Table 3 and Supplementary Material, Figure S1). The Eta variant was also detected in pool BC02, with 4325 total reads, including 2036 reads displaying also the S151I mutation (G22014T) (see Table 3 and Supplementary Material, Figure S2). Venn diagrams, which use overlapping circles or other shapes to illustrate the relationships between sets of items, confirmed the results obtained in BC02 for Alpha and Eta variants/sub-variants (see Supplementary Material, Figure S3). In pool BC03, 105,315 reads were assigned to the Alpha variant. No sub-variants or other variants were detected in this pool.

4. Discussion

The appearance and spread of new SARS-CoV-2 variants represent a major challenge for the control of the COVID-19 pandemic. Several studies have shown that detecting SARS-CoV-2 in untreated wastewaters can be a useful tool to establish outbreak trends, identify new outbreaks, and assess the prevalence of infections [21–26]. A number of studies have also reported the importance of sequencing of SARS-CoV-2 in the environment as a tool to determine strains circulating in the community and to study SARS-CoV-2 diversity [1–13]. A few of these studies also reported the detection of VOCs in sewage [2,4,6,13].

In a previous investigation, we detected SARS-CoV-2 VOCs (Alpha and Gamma) in Italy by a long nested PCR (~1600 bps) targeting the spike region, followed by Sanger sequencing [6]. All SARS-CoV-2 variants involve genetic mutations of the spike protein, which is therefore a significant region to be sequenced in order to discriminate VOCs and VOIs. In the present study, we implemented this method by combining it with a long read sequencing approach based on the Oxford Nanopore MinION platform.

Key mutations associated with two VOCs (Alpha and Gamma) and a VOI (Eta) were detected in urban sewage. Specifically, a panel of mutations characteristic of the Alpha variant (H69del, V70del, Y144del, N501Y, A570D) was detected in all the tested pools, confirming that the B.1.1.7 lineage was widespread in Italy at the time of sample collection (April/May 2021), as reported by the surveillance data on VOCs in Italy (prevalenza delle tre varianti SARS-CoV-2: lineage B.1.1.7, P1, in Italia). In all the pools, the Alpha cluster accounted for the vast majority of reads and three alpha sub-clusters were identified based on the presence of additional mutations (S98F, E484K, and S98F + E484K in combination). These two mutations are indeed occasionally found in SARS-CoV-2 sequences of the Alpha variant. As of 13 August 2021, of the 1,049,488 Alpha sequences in GISAID, 18,490 (1.8%) also contain the amino acid substitution S98F. This mutation was also detected by Sanger sequencing in one sample (ID 4162) later included in the pool (Table 2). With regard to mutation E484K, it is usually present in variants as Beta, Gamma, Eta and Zeta, but only in some isolates of the Iota and Alpha variants. In detail, mutation E484K is reported in 2959 (0.3%) of the Alpha sequences from GISAID, including 26 sequences from Italy. According to ECDC, the variant B.1.1.7 + E484K shows evidence for impact on transmissibility, immunity, and severity of the infection (<https://www.ecdc.europa.eu/en/covid-19/variants-concern>, accessed on 11 August 2021). The combination Alpha + mutations S98F and E484K, found in BC02, is overall extremely rare, having being found thus far only in 12 Alpha sequences, none of which are from Italy. However, the visual control of the alignment confirmed the reads being of good quality and not artefacts of sequencing.

In pool sample BC01, a significant number of reads (approximately 70,000) were assigned to the Gamma cluster. All the reads showed the key panel of mutations characteristic of this variant (D138Y, R190S, K417T, E484K, N501Y) and no additional mutations were found in the sample.

VOCs other than Alpha and Gamma were not identified in sewage collected in April 2021. This is in agreement with clinical data at the time of sewage collection. Since March 2021, the Italian National Institute of Health (ISS) has been reporting on SARS-CoV-2 variants in nasopharyngeal swabs by whole genome sequencing of samples collected in Italy from cases notified in specific dates (so called 'flash survey'). Monthly reports are available in the ISS website (Comunicati stampa-ISS). According to the report related to 20 April 2021, the Alpha variant prevalence in Italy was estimated at 91.6%, the Gamma at 4.5%, and all the other variants were below 0.5%. Interestingly, the Eta variant was also detected in this survey in 11 of the 2041 sequences samples, with an estimated prevalence of 0.4%. The panel of mutations of the Eta variant (A67V, H69del, V70del, Y144del, E484K) was detected in more than 4300 reads in pool BC02. An Eta sub-cluster with the additional mutation S151I (G22014T) was also detected, accounting for 2036 reads, and the visual control of the alignment confirmed these reads to be of good quality. According to GISAID database, this mutation has been detected in 741 sequences from 45 countries worldwide thus far, but has rarely been associated with the Eta variant (only three sequences). To further explore the extent of the circulation of this variant in Italy at the time of collection, the GISAID database was interrogated, and, interestingly, five B.1.525 sequences were reported for nasopharyngeal swabs collected since February 2021 and from patients of the hospital S.M. Goretti of Latina, the city where the sewage samples with the Eta variant were collected, suggesting the occurrence of a small cluster of the Eta variant in this area. Significantly, the last of these swab samples had been collected on 21 April, which is consistent with our findings on the presence of eta variants in sewage on 28–29 April in the WTP catchment area.

In this study, the Delta variant was not detected in sewage samples, despite samples being collected specifically in an area where there was concern over the possible spread of this variant in the community, due to the high number of local Indian farm labourers who had tested positive for SARS-CoV-2 (COVID: Red zone declared near Latina as dozens of Indians test positive-English-ANSA.it). It is worthy of note, however, that the absence of the Delta variant in wastewater samples in April 2021 is not surprising since, at the time of sewage sample collection, only one sequence of the Delta variant had been submitted to GISAID from Italy (hCoV-19/Italy/FVG-ASUGI_4_S4_Run10/2021) and three weeks later, on May 18, the estimated prevalence of this variant in Italy was still below 1% (Prevalenza delle tre varianti SARS-CoV-2: lineage B.1.1.7, P1, in Italia). Moreover, the epidemiological investigation on the epidemic in the province of Latina, carried out by 'Lazzaro Spallanzani' National Institute for Infectious Diseases in Rome, confirmed by extended sequencing of positive nasopharyngeal swabs that no infection attributable to the Delta variant was present in the community of Latina (https://newsrnd.com/life/2021-04-29-early-tests-latin-sikh-community--no-indian-variant-cases.HkWhSZL_vu.html, accessed on 11 August 2021).

With regard to the deep sequencing approach used in this study, both clustering and classification by specific queries were implemented for amplicon reads assignment. Clustering is a type of unsupervised analysis, used to group data points having similar characteristics as clusters. The supervised BAMQL classification analysis, instead, is able to group NGS reads in clusters on the base of selected target characteristics (query based on known key mutations typical of each variant). Significantly, while BAMQL offers direct SNPs call, it doesn't allow to directly call for deletions. In this study, we implemented a strategy to filter deletions by querying the absence of a specific nucleotide in the deleted region; along with the presence of wild type nucleotides at its 5' and 3', to minimise selection of random deletions generated by nanopore sequencing. This "deletion calling strategy", concatenated with the panel of SNPs has a strong discriminatory power for

VOCs/VOIs. Our results showed that the unsupervised clustering approach missed the Eta variant, as well as Alpha cluster, with the additional mutation E484K or the additional mutations S98F plus E484K, suggesting the inability of this approach to identify small clusters in the presence of significantly larger clusters [27,28]. The BAMQL classification, on the other hand, was able to identify also the small clusters, including the Eta variant, which represented only the 1.3% of the total reads in pool sample BC02. This method is characterised by high stringency: of the 972,348 reads obtained in total for the three pools after the filters (reads preserving both the 5' and 3' adaptor, consistent with the PCR amplicon product size, with an average quality equal or above 8, and aligning to the SARS-CoV-2 reference sequence) only 319,300 (32.8%) could be assigned. The Oxford Nanopore MinION, third generation sequencing technology, has several advantages such as portability, long read length, user friendliness, speed of sequencing, capability for multiplexing samples, and it has been demonstrated that Nanopore sequencing can identify low frequency variants in a mix of sequences [29]. However, it is important to underline that reverse transcription and PCR amplification in complex matrices display variable degrees of efficiency on different molecular targets and may significantly affect the final proportions of the amplified sequences. Therefore, the number of amplicon reads in sequenced samples do not necessarily represent variant proportions in wastewater samples and sequencing results should not be considered in a quantitative manner.

MinION's main disadvantage is a relatively low read accuracy compared with the highly accurate sequencing of Sanger and Illumina methods, the estimated error rate being between 5 and 15% [30–32] despite its constant improvement. Therefore, the analysis of ONT sequence reads is a complex task, particularly when variant detection is required. Here we demonstrated that, despite the low read accuracy of the system, SARS-CoV-2 variant assignment is feasible, first using stringent filtering parameters (i.e., selection only of reads with double barcodes and of the expected length), then applying a supervised tool for the assignation of reads based on stringent queries, and lastly by inspecting the sequence alignment to check for anomalies. Manual review of aligned reads for confirmation and interpretation of variant calls is an important step that can significantly increase the confidence in calls and reduce the risk of false positives. For example, if an alignment check reveals that variant bases are only on strands of the same polarity the subsequent reads assignation is likely to be a false positive, since the expected strands distribution is 50–50 with the protocol used for this sample sequencing [33]. In this study, the protocol was designed to avoid false positive mutation calls, in the awareness that the sequencing system has a non-negligible rate of errors. However, it should be considered that a high stringency of the analysis may also lead to false negative results, missing the detection of SARS-CoV-2 variants that could be present in the sample. We can assume that the vast majority of the unassigned reads were composed by PCR chimera products, or were of low quality, including too many nucleotide substitutions or insertion/deletions to be assigned. In particular, a characterization via Ins/Del ratio of unassigned reads versus assigned reads highlights that unassigned reads display a considerable higher number of small deletions, therefore hindering their correct alignment and assignation. Furthermore, sequences with an exceedingly high number of deletions and misalignments may easily be missed by the BAMQL stringent queries of the classification approach, due to absence/mismatch of one of the queried targets.

In the current study, sequencing of PCR products was not performed on separate samples, but on pooled PCR amplicons obtained from samples collected at the same WTP in close dates, thus reducing analytical costs and time. Indeed, despite the reduction of NGS costs in recent years, sequencing of large numbers of samples is still challenging. Therefore, barcoding and pooling of samples is a cost- and time-effective approach. One of the most challenging problems of sample pooling is the correct identification of rare variants. In this study, however, we were able to identify mutations related to the Eta variant accounting for only 1.3% of total NGS reads of the sample, confirming that rare

variants can be successfully detected after pooling, as already demonstrated by other studies [34].

In this study, 6 of the 12 samples could not be amplified using the long RT-nested-PCR assay. Of these, 5 samples displayed presence of SARS-CoV-2 RNA by real-time RT-qPCR in low concentrations (equal or below 3.5 g.c./ μ L of tested RNA). This means that SARS-CoV-2 infection was present in the community, but variant analysis could not be completed by partial spike gene sequencing since the amplification of long fragments poses significant challenges in complex matrices as wastewaters. Indeed, while false-positive errors can be reduced through strict adherence to best laboratory practices with appropriate controls, false-negative results may be associated to intrinsic sensitivity limits of the analytical methods. For a detailed review on factors contributing to SARS-CoV-2 RT-PCR false-negative errors in wastewater surveillance, see Ahmed et al., 2021 [35].

5. Conclusions

In conclusion, the present study, the first to our knowledge using long-read amplicon sequencing based on nanopore technology to detect SARS-CoV-2 variants in sewage, provides the results of a preliminary analysis of VOCs and VOIs based on partial spike gene amplicon sequencing, using the ONT MinION. The developed bioinformatics analysis approach for variants identification allowed the detection of mutations characteristic of the Alpha, Gamma, and Eta variants in sewage samples from central Italy. Regular collection of wastewater samples and systematic sequencing of the spike gene of SARS-CoV-2 in sewage will help identify VOCs and VOIs and track their circulation in the population.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/w13182503/s1>. Figure S1. AlleleClusterSummary for pool BC01 (Clustering unsupervised results); Figure S2. Mutations detected in the alpha cluster; Figure S3. Mutations detected in the eta cluster; Figure S4. Venn diagrams for Alpha and Eta reads in pool BC02; Figure S5. Distribution of small deletions in assigned reads (A) and in unassigned reads (B): small deletions are significantly more frequent in unassigned reads; Figure S6. Insertion/Deletion analysis in assigned reads (A) and unassigned reads (B): in unassigned reads Ins/Del ratio is exceedingly low in the range of 10 bases indicating an abundance of small deletions in comparison to assigned reads. Table S1. Results of real-time RT-qPCR detection of SARS-CoV-2 in the tested samples; Table S2. Number of raw and final reads selected after filtering steps; Table S3. Concatenated mutations queried using the BAMQL tool v1.6 for the detection of each VOC/VOI.

Author Contributions: Conceptualization, G.L.R., E.S., L.L., L.B.; methodology, G.B.F., P.M., C.V., D.B.; writing—original draft preparation, G.L.R. and E.S.; writing—review and editing, G.L.R., E.S., L.L., L.B., G.B.F., P.M., C.V., D.B. All authors have read and agreed to the published version of the manuscript.

Funding: The study has received funding from the Italian Ministry of Health in the framework of the national project “Wastewater based epidemiology: implementation of a surveillance system for the early detection of pathogens with a focus on SARS-CoV-2” (CCM 2020 program).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: We thank Project RefBio-Germany for strengthening the bio-analytical reference laboratories in the UNSGM (2017–2021), EQAE on SARS-CoV-2 (2020), Centre for Biological Threats and Special Pathogens (ZBS 1–3) at RKI (Project Coordinator), Berlin, Germany, for providing the Wuhan strain. We thank the water manager at Acquatina S.p.A. for wastewater sampling.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Agrawal, S.; Orschler, L.; Lackner, S. Metatranscriptomic Analysis Reveals SARS-CoV-2 Mutations in Wastewater of the Frankfurt Metropolitan Area in Southern Germany. *Microbiol. Resour. Announc.* **2021**, *10*, e00280-21. [[CrossRef](#)]
2. Bar-Or, I.; Weil, M.; Indenbaum, V.; Bucris, E.; Bar-Ilan, D.; Elul, M.; Levi, N.; Aguvaev, I.; Cohen, Z.; Shirazi, R.; et al. Detection of SARS-CoV-2 variants by genomic analysis of wastewater samples in Israel. *Sci. Total Environ.* **2021**, *789*, 148002. [[CrossRef](#)]
3. 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.; et al. Genome Sequencing of Sewage Detects Regionally Prevalent SARS-CoV-2 Variants. *mBio* **2021**, *12*, e02703-20. [[CrossRef](#)]
4. Heijnen, L.; Elsinga, G.; de Graaf, M.; Molenkamp, R.; Koopmans, M.P.; Medema, G. Droplet digital RT-PCR to detect SARS-CoV-2 signature mutations of variants of concern in wastewater. *Sci. Total Environ.* **2021**, *799*, 149456. [[CrossRef](#)]
5. Izquierdo-Lara, R.; Elsinga, G.; Heijnen, L.; Munnink, B.B.O.; Schapendonk, C.M.; Nieuwenhuijse, D.; Kon, M.; Lu, L.; Aarestrup, F.M.; Lycett, S.; et al. Monitoring SARS-CoV-2 Circulation and Diversity through Community Wastewater Sequencing, the Netherlands and Belgium. *Emerg. Infect. Dis.* **2021**, *27*, 1405–1415. [[CrossRef](#)]
6. La Rosa, G.; Mancini, P.; Ferraro, G.B.; Veneri, C.; Iaconelli, M.; Lucentini, L.; Bonadonna, L.; Brusaferrero, S.; Brandtner, D.; Fasanella, A.; et al. Rapid screening for SARS-CoV-2 variants of concern in clinical and environmental samples using nested RT-PCR assays targeting key mutations of the spike protein. *Water Res.* **2021**, *197*, 117104. [[CrossRef](#)] [[PubMed](#)]
7. Martin, J.; Klapsa, D.; Wilton, T.; Zambon, M.; Bentley, E.; Bujaki, E.; Fritzsche, M.; Mate, R.; Majumdar, M. Tracking SARS-CoV-2 in Sewage: Evidence of Changes in Virus Variant Predominance during COVID-19 Pandemic. *Viruses* **2020**, *12*, 1144. [[CrossRef](#)] [[PubMed](#)]
8. Mondal, S.; Feirer, N.; Brockman, M.; Preston, M.A.; Teter, S.J.; Ma, D.; Goueli, S.A.; Moorji, S.; Saul, B.; Cali, J.J. A direct capture method for purification and detection of viral nucleic acid enables epidemiological surveillance of SARS-CoV-2. *Sci. Total Environ.* **2021**, *795*, 148834. [[CrossRef](#)] [[PubMed](#)]
9. Nemudryi, A.; Nemudraia, A.; Wiegand, T.; Surya, K.; Buyukyoruk, M.; Cicha, C.; Vanderwood, K.K.; Wilkinson, R.; Wiedenheft, B. Temporal Detection and Phylogenetic Assessment of SARS-CoV-2 in Municipal Wastewater. *Cell Rep. Med.* **2020**, *1*, 100098. [[CrossRef](#)]
10. Prado, T.; Fumian, T.M.; Mannarino, C.F.; Resende, P.C.; Motta, F.C.; Eppinghaus, A.L.F.; Vale, V.H.C.D.; Braz, R.M.S.; Andrade, J.D.S.R.D.; Maranhão, A.G.; et al. Wastewater-based epidemiology as a useful tool to track SARS-CoV-2 and support public health policies at municipal level in Brazil. *Water Res.* **2021**, *191*, 116810. [[CrossRef](#)] [[PubMed](#)]
11. Rimoldi, S.G.; Stefani, F.; Gigantiello, A.; Polesello, S.; Comandatore, F.; Mileto, D.; Maresca, M.; Longobardi, C.; Mancon, A.; Romeri, F.; et al. Presence and infectivity of SARS-CoV-2 virus in wastewaters and rivers. *Sci. Total Environ.* **2020**, *744*, 140911. [[CrossRef](#)]
12. Rubio-Acero, R.; Beyerl, J.; Muenchhoff, M.; Roth, M.S.; Castelletti, N.; Paunovic, I.; Radon, K.; Springer, B.; Nagel, C.; Boehm, B.; et al. Spatially resolved qualified sewage spot sampling to track SARS-CoV-2 dynamics in Munich—One year of experience. *Sci. Total Environ.* **2021**, *797*, 149031. [[CrossRef](#)] [[PubMed](#)]
13. Wilton, T.; Bujaki, E.; Klapsa, D.; Majumdar, M.; Zambon, M.; Fritzsche, M.; Mate, R.; Martin, J. Rapid Increase of SARS-CoV-2 Variant B.1.1.7 Detected in Sewage Samples from England between October 2020 and January 2021. *mSystems* **2021**, *6*, e0035321. [[CrossRef](#)]
14. Mancini, P.; Ferraro, G.B.; Iaconelli, M.; Suffredini, E.; Valdazo-González, B.; Della Libera, S.; Divizia, M.; La Rosa, G. Molecular characterization of human Sapovirus in untreated sewage in Italy by amplicon-based Sanger and next-generation sequencing. *J. Appl. Microbiol.* **2019**, *126*, 324–331. [[CrossRef](#)] [[PubMed](#)]
15. Suffredini, E.; Iaconelli, M.; Equestre, M.; Valdazo-González, B.; Ciccaglione, A.R.; Marcantonio, C.; della Libera, S.; Bignami, F.; La Rosa, G. Genetic Diversity Among Genogroup II Noroviruses and Progressive Emergence of GII.17 in Wastewaters in Italy (2011–2016) Revealed by Next-Generation and Sanger Sequencing. *Food Environ. Virol.* **2018**, *10*, 141–150. [[CrossRef](#)]
16. Iaconelli, M.; Valdazo-González, B.; Equestre, M.; Ciccaglione, A.R.; Marcantonio, C.; della Libera, S.; La Rosa, G. Molecular characterization of human adenoviruses in urban wastewaters using next generation and Sanger sequencing. *Water Res.* **2017**, *121*, 240–247. [[CrossRef](#)] [[PubMed](#)]
17. Wu, F.; Zhang, J.; Xiao, A.; Gu, X.; Lee, W.L.; Armas, F.; Kauffman, K.; Hanage, W.; Matus, M.; Ghaeli, N.; et al. SARS-CoV-2 Titers in Wastewater Are Higher than Expected from Clinically Confirmed Cases. *mSystems* **2020**, *5*, e00614-20. [[CrossRef](#)]
18. Costafreda, M.I.; Bosch, A.; Pintó, R.M. Development, Evaluation, and Standardization of a Real-Time TaqMan Reverse Transcription-PCR Assay for Quantification of Hepatitis A Virus in Clinical and Shellfish Samples. *Appl. Environ. Microbiol.* **2006**, *72*, 3846–3855. [[CrossRef](#)]
19. La Rosa, G.; Mancini, P.; Ferraro, G.B.; Veneri, C.; Iaconelli, M.; Bonadonna, L.; Lucentini, L.; Suffredini, E. SARS-CoV-2 has been circulating in northern Italy since December 2019: Evidence from environmental monitoring. *Sci. Total Environ.* **2021**, *750*, 141711. [[CrossRef](#)]
20. Masella, A.P.; Lalansingh, C.M.; Sivasundaram, P.; Fraser, M.; Bristow, R.G.; Boutros, P.C. BAMQL: A query language for extracting reads from BAM files. *BMC Bioinform.* **2016**, *17*, 305. [[CrossRef](#)]
21. Panchal, D.; Prakash, O.; Bobde, P.; Pal, S. SARS-CoV-2: Sewage surveillance as an early warning system and challenges in developing countries. *Environ. Sci. Pollut. Res. Int.* **2021**, *28*, 22221–22240. [[CrossRef](#)]

22. Macku'ak, T.; Gál, M.; Špalková, V.; Fehér, M.; Briestenská, K.; Mikušová, M.; Tomčíková, K.; Tamáš, M.; Butor Škulcová, A. Wastewater-Based Epidemiology as an Early Warning System for the Spreading of SARS-CoV-2 and Its Mutations in the Population. *Int. J. Environ. Res. Public Health* **2021**, *18*, 5629. [[CrossRef](#)]
23. Lundy, L.; Fatta-Kassinos, D.; Slobodnik, J.; Karaolia, P.; Cirka, L.; Kreuzinger, N.; Castiglioni, S.; Bijlsma, L.; Dulio, V.; Deviller, G.; et al. Making Waves: Collaboration in the time of SARS-CoV-2—Rapid development of an international co-operation and wastewater surveillance database to support public health decision-making. *Water Res.* **2021**, *199*, 117167. [[CrossRef](#)] [[PubMed](#)]
24. Hamouda, M.; Mustafa, F.; Maraqa, M.; Rizvi, T.; Aly Hassan, A. Wastewater surveillance for SARS-CoV-2: Lessons learnt from recent studies to define future applications. *Sci. Total Environ.* **2021**, *759*, 143493. [[CrossRef](#)] [[PubMed](#)]
25. Rooney, C.M.; Moura, I.B.; Wilcox, M.H. Tracking COVID-19 via sewage. *Curr. Opin. Gastroenterol.* **2021**, *37*, 4–8. [[CrossRef](#)] [[PubMed](#)]
26. Medema, G.; Been, F.; Heijnen, L.; Pettersen, S. Implementation of environmental surveillance for SARS-CoV-2 virus to support public health decisions: Opportunities and challenges. *Curr. Opin. Environ. Sci. Health* **2020**, *17*, 49–71. [[CrossRef](#)]
27. Ertöz, L.; Steinbach, M.; Kumar, V. Finding Clusters of Different Sizes, Shapes, and Densities in Noisy, High Dimensional Data. In Proceedings of the 2003 SIAM International Conference on Data Mining; Society for Industrial and Applied Mathematics, San Francisco, CA, USA, 1–3 May 2003; pp. 47–58.
28. Liang, M.; Raley, C.; Zheng, X.; Kutty, G.; Gogineni, E.; Sherman, B.T.; Sun, Q.; Chen, X.; Skelly, T.; Jones, K.; et al. Distinguishing highly similar gene isoforms with a clustering-based bioinformatics analysis of PacBio single-molecule long reads. *BioData Min.* **2016**, *9*, 13. [[CrossRef](#)]
29. Riaz, N.; Leung, P.; Barton, K.; Smith, M.A.; Carswell, S.; Bull, R.; Lloyd, A.R.; Rodrigo, C. Adaptation of Oxford Nanopore technology for hepatitis C whole genome sequencing and identification of within-host viral variants. *BMC Genom.* **2021**, *22*, 148. [[CrossRef](#)]
30. Kono, N.; Arakawa, K. Nanopore sequencing: Review of potential applications in functional genomics. *Dev. Growth Differ.* **2019**, *61*, 316–326. [[CrossRef](#)]
31. Lin, B.; Hui, J.; Mao, H. Nanopore Technology and Its Applications in Gene Sequencing. *Biosensors* **2021**, *11*, 214. [[CrossRef](#)]
32. Ogden, R.; Vasiljevic, N.; Prost, S. Nanopore sequencing in non-human forensic genetics. *Emerg. Top. Life Sci.* **2021**. [[CrossRef](#)] [[PubMed](#)]
33. Robinson, J.T.; Thorvaldsdóttir, H.; Wenger, A.M.; Zehir, A.; Mesirov, J.P. Variant Review with the Integrative Genomics Viewer. *Cancer Res.* **2017**, *77*, e31–e34. [[CrossRef](#)] [[PubMed](#)]
34. Anand, S.; Mangano, E.; Barizzone, N.; Bordoni, R.; Sorosina, M.; Clarelli, F.; Corrado, L.; Boneschi, F.M.; D'Alfonso, S.; De Bellis, G. Next Generation Sequencing of Pooled Samples: Guideline for Variants' Filtering. *Sci. Rep.* **2016**, *6*, 33735. [[CrossRef](#)] [[PubMed](#)]
35. Ahmed, W.; Simpson, S.L.; Bertsch, P.M.; Bibby, K.; Bivins, A.; Blackall, L.L.; Bofill-Mas, S.; Bosch, A.; Brandão, J.; Choi, P.M.; et al. Minimizing errors in RT-PCR detection and quantification of SARS-CoV-2 RNA for wastewater surveillance. *Sci. Total Environ.* **2021**, 149877. [[CrossRef](#)]