Science of the Total Environment 800 (2021) 149562

Contents lists available at ScienceDirect

Science of the Total Environment

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

Exploring the diversity of coronavirus in sewage during COVID-19 pandemic: Don't miss the forest for the trees

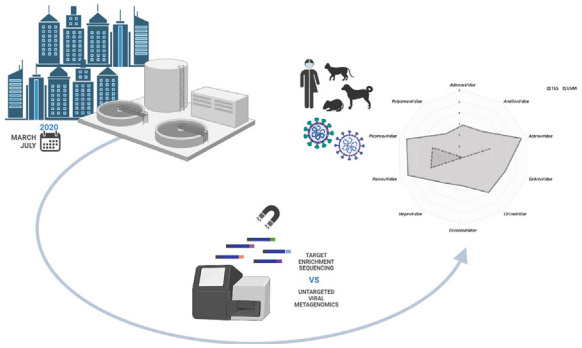
Sandra Martínez-Puchol [a,b,*], Marta Itarte [a,b], Marta Rusiñol [c], Eva Forés [a,b], Cristina Mejías-Molina [a], Cristina Andrés [d], Andrés Antón [d], Josep Quer [e], Josep F. Abril [f], Rosina Girones [a,b], Sílvia Bofill-Mas [a,b]

[a] Laboratory of Viruses Contaminants of Water and Food, Genetics, Microbiology & Statistics Dept., Universitat de Barcelona. Barcelona, Catalonia, Spain
[b] The Water Research Institute (IdRA), Universitat de Barcelona, Barcelona, Catalonia, Spain
[c] Institute of Environmental Assessment & Water Research (IDAEA), CSIC, Barcelona, Catalonia, Spain
[d] Respiratory Viruses Unit, Virology Section, Microbiology Department, Hospital Universitari Vall d'Hebron, Vall d'Hebron Research Institute, Universitat Autònoma de Barcelona, Barcelona, Catalonia, Spain
[e] Liver Unit, Liver Diseases - Viral Hepatitis, Vall d'Hebron Institut de Recerca (VHIR), Vall d'Hebron Hospital Universitari, Vall d'Hebron Barcelona Hospital Campus, Passeig Vall d'Hebron 119-129, 08035 Barcelona, Spain
[f] Computational Genomics Lab, Genetics, Microbiology & Statistics Dept., Universitat de Barcelona; Institut de Biomedicina (IBUB), Barcelona, Catalonia, Spain

HIGHLIGHTS

- NGS methods have been applied to study the sewage virome during COVID-19 pandemic.
- *Coronaviridae* sequences were not detected in sewage using untargeted metagenomics.
- Target Enrichment provided with SARS-CoV-2 sequences as part of the sewage virome.
- Human and animal CoV co-circulation in sewage only detected with Target Enrichment.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 9 June 2021
Received in revised form 5 August 2021
Accepted 6 August 2021
Available online 9 August 2021

Keywords:
Next generation sequencing
Viral metagenomics
Coronavirus
Sewage virome
Target enrichment sequencing

ABSTRACT

In the wake of the COVID-19 pandemic, the use of next generation sequencing (NGS) has proved to be an important tool for the genetic characterization of SARS-CoV-2 from clinical samples. The use of different available NGS tools applied to wastewater samples could be the key for an in-depth study of the excreted virome, not only focusing on SARS-CoV-2 circulation and typing, but also to detect other potentially pandemic viruses within the same family. With this aim, 24-hours composite wastewater samples from March and July 2020 were sequenced by applying specific viral NGS as well as target enrichment NGS. The full virome of the analyzed samples was obtained, with human *Coronaviridae* members (CoV) present in one of those samples after applying the enrichment. One contig was identified as HCoV-OC43 and 8 contigs as SARS-CoV-2. CoVs from other animal hosts were also detected when applying this technique. These contigs were compared with those obtained from contemporary clinical specimens by applying the same target enrichment approach. The results showed that there is a co-circulation in urban areas of human and animal coronaviruses infecting domestic animals and rodents. NGS enrichment-based protocols might be crucial to describe the occurrence and genetic characteristics of SARS-CoV-2 and other *Coronaviridae* family members within the excreted virome present in wastewater.

* Corresponding author at: Laboratory of Viruses Contaminants of Water and Food, Genetics, Microbiology & Statistics Dept., Universitat de Barcelona, Barcelona, Catalonia, Spain.
E-mail address: smartinezpuchol@ub.edu (S. Martínez-Puchol).

https://doi.org/10.1016/j.scitotenv.2021.149562

1. Introduction

SARS-CoV-2 was identified in China at the end of 2019 and has become the first pandemic coronavirus (CoV) (Wu et al., 2020a, 2020b). Upon confirmation that COVID-19 patients shed SARS-CoV-2 in feces, different studies provided significant correlation between the concentration of SARS-CoV-2 in wastewater and the prevalence of COVID-19 in the served population, increasing the evidence that wastewater is a good indicator of the prevalence of the excreted virus in a population (wastewater-based epidemiology, WBE). SARS-CoV-2 RNA has been reported in feces of 55% of COVID-19 patients in the study cohort, prolonged fecal shedding was also reported with fecal samples remaining positive for nearly 5 weeks after the patients' respiratory samples tested negative for SARS-CoV-2 RNA (Wu et al., 2020a, 2020b). The concentration of SARS-CoV-2 excreted from infected patients has been estimated to be in the range of 1E+02 to 1E+07 copies per gram of feces (Wölfel et al., 2020).

It is then presumable that, in the context of a pandemic, SARS-CoV-2 could become a member of wastewater virome where viruses commonly present in human excreta have been described to occur (Cantalupo et al., 2011; Ng et al., 2012). The presence of this virus in wastewater samples has been extensively reported by RT-PCR based methods. Also, an extensive effort for SARS-CoV-2 sequencing has been done based on metagenomic approaches focused on variant analysis in both clinical and sewage samples. However, the use of metagenomic approaches to describe its occurrence together with other excreted viruses in the context of sewage virome has not been described. Knowing that SARS-CoV-2 is found in sewage in moderate concentrations and sewage is a complex matrix comprising a wide variety of viruses, studying SARS-CoV-2 in this context may pose a challenge. Previous studies report procedures to describe wastewater virome and compare high throughput viral metagenomics to target enrichment and amplicon deep sequencing approaches (Martínez-Puchol et al., 2020). In this study, we applied a probe-capture target enrichment NGS for describing wastewater virome from 2 wastewater 24-h composite samples collected in March and July 2020 from a wastewater treatment plant located in the city of Barcelona, when COVID-19 incidence was of approximately 258 cases and 175 cases/100,000 inhabitants respectively. The results obtained were compared with those obtained from clinical samples collected at the same time and sequenced applying the same enrichment approach.

2. Materials and methods

2.1. Sample collection and viral concentration

Urban 24-h composite sewage samples were collected on March 19th and July 14th 2020 from a wastewater treatment plant (WWTP) located in the city of Barcelona that serves up to 2.8 million population equivalents and receives domestic and industrial waste from the sewer system. Samples were collected and kept at 4 °C in a sterile container until viral particles from 70 ml of sewage were concentrated by centrifugal ultrafiltration. After a debris removal (15 min, 4500 ×g), the samples were ultrafiltered with a Centricon® Plus-70 device (30 kDa) following the manufacturer instructions, obtaining a viral concentrates of 200 μl. Additionally, two SARS-CoV-2-positive naso/oropharyngeal swabs were obtained on March 15th from a male (clinical sample A) and on March 24th from a female (clinical sample B) patients attended at Hospital Universitari Vall d'Hebron (HUVH).

2.2. SARS-CoV-2 quantification by (RT)-qPCR

Nucleic acid extraction (NA) was performed as described previously (Fernandez-Cassi et al., 2018) with QIAamp Viral RNA Mini Kit. The concentration of SARS-CoV-2 RNA in wastewater samples was measured by (RT)-qPCR of two viral targets in nucleocapsid phosphoprotein (N1 and N2 region) and clinical samples in HUVH by commercial real-time multiplex RT-PCR (Allplex™ 2019-nCoV Assay, Seegene, South Korea). EURM-019 single stranded RNA (ssRNA) fragments of SARS-CoV-2 (Joint Research Centre, EC, https://crm.jrc.ec.europa.eu/p/EURM-019) were used to construct the standard for quantitation. JC polyomavirus (JCPyV) was quantified in the samples as an indicator of human fecal viral contamination as previously described (Bofill-Mas et al., 2006).

2.3. Library construction and probe-based capture of viral sequences

Before library preparation, NA were retrotranscribed to cDNA, tagged and complemented to obtain dsDNA. This viral randomly tagged dsDNA was then amplified (25 cycles) to obtain the sufficient amount of DNA for the preparation of libraries, as previously described (Fernandez-Cassi et al., 2018). Libraries were prepared in duplicate using KAPA HyperPrep Kit following the instructions provided by the manufacturer (Roche-Kapa Biosystems). One replicate of the libraries was hybridized with probes designed to capture sequences from vertebrate viral pathogens (VirCapSeq Enrichment Kit, Roche). Two negative controls were also processed, one of them with the enrichment kit.

After the capture, quality and concentration were re-checked and sequencing of the libraries from the captured, non-captured, and negative controls was performed (Illumina Miseq 2x300bp).

Clinical samples were sequenced in an independent Illumina Miseq 2×300bp run (in this case, consensus sequences were assembled with reads produced from another Illumina TruSeq high coverage sequencing run on same samples, and are already available at GISAID as EPI_ISL_418860 and EPI_ISL_418861, respectively).

2.4. Bioinformatic analysis

The sequencing raw data obtained was analyzed with Genome Detective Virus Tool (Vilsker et al., 2018) and the contigs with a nucleotide identity ≥70% (when comparing against the known viral genomes database) were further processed with Geneious (v11.1.5; https://www.geneious.com). Simultaneously, raw reads from wastewater and clinical samples were cleaned of technical sequences and trimmed by quality using Trimmomatic (v0.38; (Bolger et al., 2014)) in order to remove low quality segments and Illumina adapters (min Phred score = Q20 on 4 bp window, min read length = 30 bp, leading/trailing clip = 15 bp, max mismatch count = 2, palindrome clip threshold = 30, and simple clip threshold = 10). When one of the reads, either R1 or R2, was discarded, the remaining one was collected into a single-ended (SG) reads file to use along with the resulting filtered paired-end reads (PE) reads later on. Samtools (v1.9; [11]) and bamtools (v2.5.1; (Barnett et al., 2011)) sets of commands were used to process, sort, and index those alignments made by bowtie2 (v2.3.4.3, 64bit; (Langmead and Salzberg, 2012); with parameters k = 5, L = 12, and "sensitive-local" switch), mapping the PE/SG reads against the SARS-CoV-2 reference genome (GenBank entry: NC_045512.2) and ensuring that the stored alignments were position sorted on the final bam files. Trinity (r20190503git; (Grabherr et al., 2013; Haas et al., 2013); min contig length = 100, k-mer size 31) produced contigs that later on were mapped over the reference genome obtaining the final scaffolds, then manually curated to finish the sequences. Due to the low number of reads recovered from wastewater sample, the corresponding assembly was filled with 93% of Ns to place the contigs into the assembled scaffold (this sequence is provided as Supplementary File 1). Mafft (v7.407; (Katoh and Standley, 2013); with localpair switch on) was chosen to calculate the multiple sequence alignment shown on Fig. 4, comparing the wastewater scaffold against a randomly sampled set of 600 Spanish and 600 international sequences from GISAID database (available on August 4th 2020; (Elbe and Buckland-Merrett, 2017; Shu and McCauley, 2017); see also Acknowledgments).

3. Results and discussion

3.1. SARS-CoV-2 quantification in clinical and wastewater samples

Viral SARS-CoV-2 copies were quantified in composite raw sewage samples collected in Barcelona different times of COVID-19 pandemic, March 2020, one week after the declaration of a 'state of alarm' in the country, and July 2020, 2020 just before the beginning of the second COVID-19 peak. The (RT)-qPCR were performed in quadruplicate with the direct NA extractions and in quadruplicate with 1/10 dilution of NA extractions, in order to avoid potential inhibitory effects. Concentration values obtained were 3.92E6 (± 0.154) GC/L for N1 and 2.71E6 (±0.117) GC/L for N2 in March 2020 and 3.80E4 (±0.06) GC/L for N1 and 1.25E4 (±0.017) GC/L for N2 in July 2020. The concentrations for the human fecal indicator JCPyV were 1.99E5 (±0.057) GC/L and 8.73E5 (±0.026) GC/L in March and July respectively (Bofill-Mas et al., 2006). Enzymatic inhibition was not observed. These values are in accordance with the ones observed in the same WTP after a full year of weekly monitoring surveillance and 2 peak registered periods (Rusiñol et al., 2021).

Regarding the two SARS-CoV-2 laboratory-confirmed clinical samples from HUVH, the E gene cycle-threshold values were 15.5 and 16.4 from EPI_ISL_418860 and EPI_ISL_418861, respectively.

3.2. Urban wastewater virome during COVID-19 pandemic

Sewage samples from March and July were mass-sequenced in parallel using high throughput sequencing and probe-based target enrichment sequencing. The virome obtained in March, using target enrichment, resulted in almost 1 million reads belonging to 27 viral families. The distribution of the number of reads obtained for each vertebrate viral family is shown in Fig. 1. As expected, a wide variety of viral families infecting vertebrates was observed when applying the probed-based capture methodology. This approach was successfully applied in the past to improve the deep sequencing in human-focused virome studies (Briese et al., 2015; Hjelmsø et al., 2019; Martínez-Puchol et al., 2020), which is a key point in the study of viral species that are present in a low concentration in environmental and clinical samples. Traditional virome studies have shown that the most abundant viral sequences belong to bacteriophages and plant viruses (Cantalupo et al., 2011; Fernandez-Cassi et al., 2018), but depending on the sequencing platform used or the number of samples multiplexed, the total number of reads obtained could be reduced and relevant human and animal viral reads could go undetected. Target enrichment methods could overcome these limitations enabling the possibility of viral discovery.

In this study we used the VirCapSeq Enrichment Kit (Roche) intended to capture vertebrate viruses from complex samples prior to metagenomic sequencing, trying to improve the limitations of high throughput sequencing. It employs approximately 2 million biotinylated oligonucleotide probes designed to bind to the coding sequences of all viral taxa known to infect vertebrates at intervals of 50–100 nt. Libraries prepared from random primed cDNA are hybridized with the biotinylated probes and trapped with streptavidin magnetic beads. After magnetic capture and washing, NA are released from the beads and subjected to post-hybridization PCR prior to sequencing. In silico studies of the VirCapSeq panel performed in this study showed that although the platform was designed before the emergence of SARS-CoV-2, the kit nonetheless contained 21,414 fragments from 1838 sequences described for 346 *Coronaviridae* species. NCBI-BLASTn of the full sequences and the probes was run, with default parameters apart from the e-value set at 10e-25, against SARS-CoV-2 reference genome (RefSeq ID: NC_045512; Wu et al., 2020a, 2020b). Among the 1838 full targets already contained in the VirCapSeq kit, 277 were found on 456 alignment hits, and 29 of those had hits above 90% identity; on the other hand, 104 probe fragments from 28 different sequences returned 104 hits.

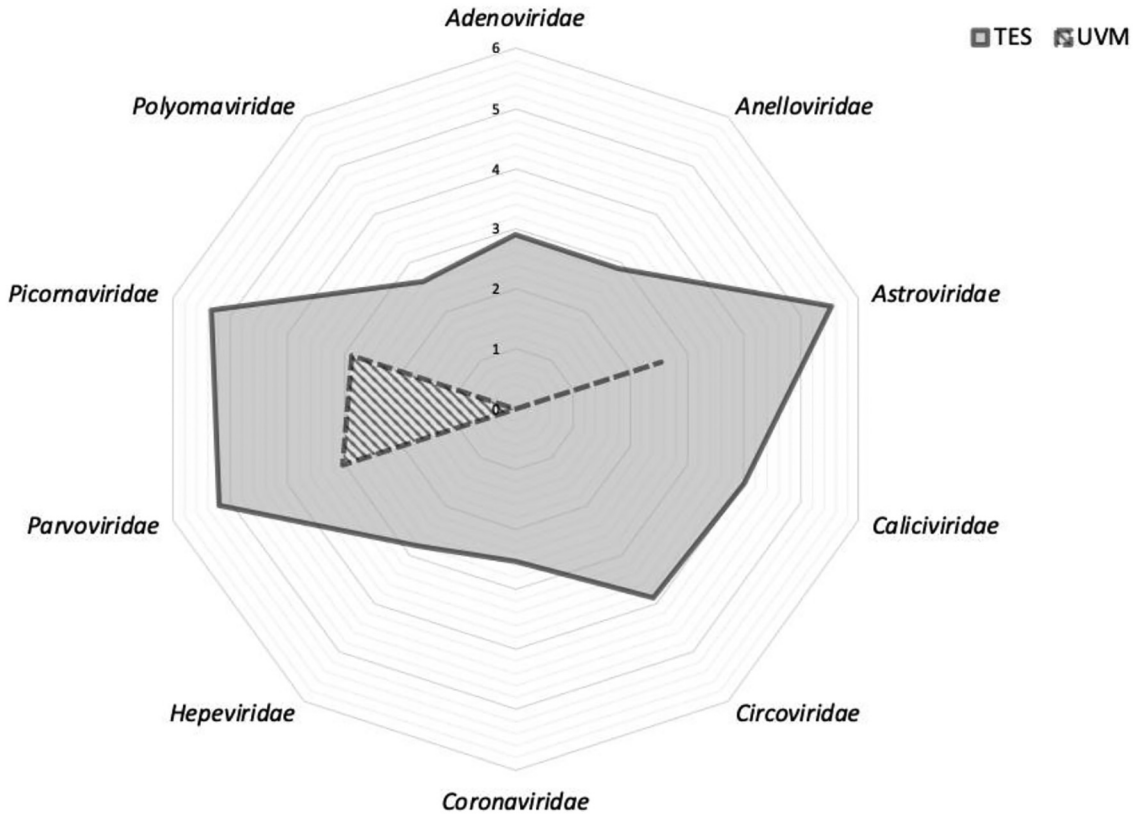


Fig. 1. Diversity and relative abundance of vertebrate viral families reads obtained from March sample. The results for untargeted viral metagenomics (UVM) are represented in dashed lines, regular lines represent the results after Target Enrichment NGS (TES). Relative abundance of reads is represented in log10.

Table 1
Coronaviridae family contigs description after performing Probe-capture targeted NGS in one urban wastewater sample (March 2020).

Coronaviridae species	Host	Reads	Contigs	Length (bp)	Region	Nt ID (%)	AA ID (%)
SARS-CoV-2	Human	45	8	203	ORF1ab polyprotein	100	100
				217	ORF1ab polyprotein	100	100
				126	ORF1ab polyprotein	100	100
				126	ORF1ab polyprotein	100	100
				178	ORF1ab polyprotein	100	100
				266	ORF3 protein	100	100
				516	Nucleocapsid protein	99.6	99.1
				399	Nucleocapsid protein	99.8	99.3
Betacoronavirus 1 (HCoV OC43)	Human	11	1	235	2'-*O*-methyltransferase	99.5	98.7
Feline coronavirus	Other vertebrates	33	1	524	ORF1ab polyprotein	91.8	86.5
Lucheng Rn rat coronavirus	Other vertebrates	233	5	354	ORF1ab polyprotein	97.4	97.1
				416	ORF1ab polyprotein	97.0	99.3
				542	ORF1ab polyprotein	97.0	99.4
				848	ORF1ab polyprotein	96.6	96.6
				1129	ORF1ab polyprotein	97.3	99.4
Canine coronavirus	Other vertebrates	14	1	231	ORF1ab polyprotein	96.1	94.8

The most abundant families obtained in March sample using target enrichment were *Astroviridae*, *Picornaviridae* and *Parvoviridae*. A complete list of all the identified viral species and families is presented in the Supplementary File 2. Homology searches against known viral genomes database retrieved *Coronaviridae* sequences listed in Table 1. One contig of a Betacoronavirus 1, typed as HCoV-OC43, was mapped with a high identity (99.5% at nucleotide level, 98.7% at aminoacid level) to the CoV 2'-*O*-methyltransferase, an enzyme that enables the mRNA cap formation, essential for viral RNA stability (Krafcikova et al., 2020). HCoV-OC43,one of the four seasonal HCoVs that are known to cause the common cold in humans, is especially prevalent during the winter months (Van der Hoek, 2015), and is known to have rodents as natural hosts and bovines as intermediate ones (Ye et al., 2020).

Eight SARS-CoV-2 contigs were obtained, with a total of 2.03 Kb representing 6.8% of the genome. Five contigs corresponded to the ORF1ab polyprotein region and one contig to the ORF3 protein, all of them with 100% of nucleotide and amino acid identity with the genome Reference Sequence NC_045512. Additionally, two contigs, with sizes 516 bp and 316 bp, mapped against the nucleocapsid protein, separated with a gap between them of 231 bp, and presented a nucleotide identity of 99.6% and 99.8% to the reference sequence.

Regarding other members of the *Coronaviridae* family, one feline CoV contig (524 bp) matching the replicase ORF1ab polyprotein region was obtained in the analysis of sequences from the March wastewater sample. This virus causes asymptomatic persistent enteric infections in a high percentage of household and catteries' cats (Vogel et al., 2010). While feline CoV is highly prevalent in their hosts, its survival in sewage has been reported to be limited (Gundy et al., 2009). One contig from the same region (231 bp) was typed as canine CoV, known for being responsible of mild or moderate enteritis in dogs of all breeds and ages and with a high establishment in the environment (Pratelli, 2006). Finally, five contigs of Lucheng Rn rat CoV, comprising 3.28 Kb of the ORF1ab polyprotein, were obtained.

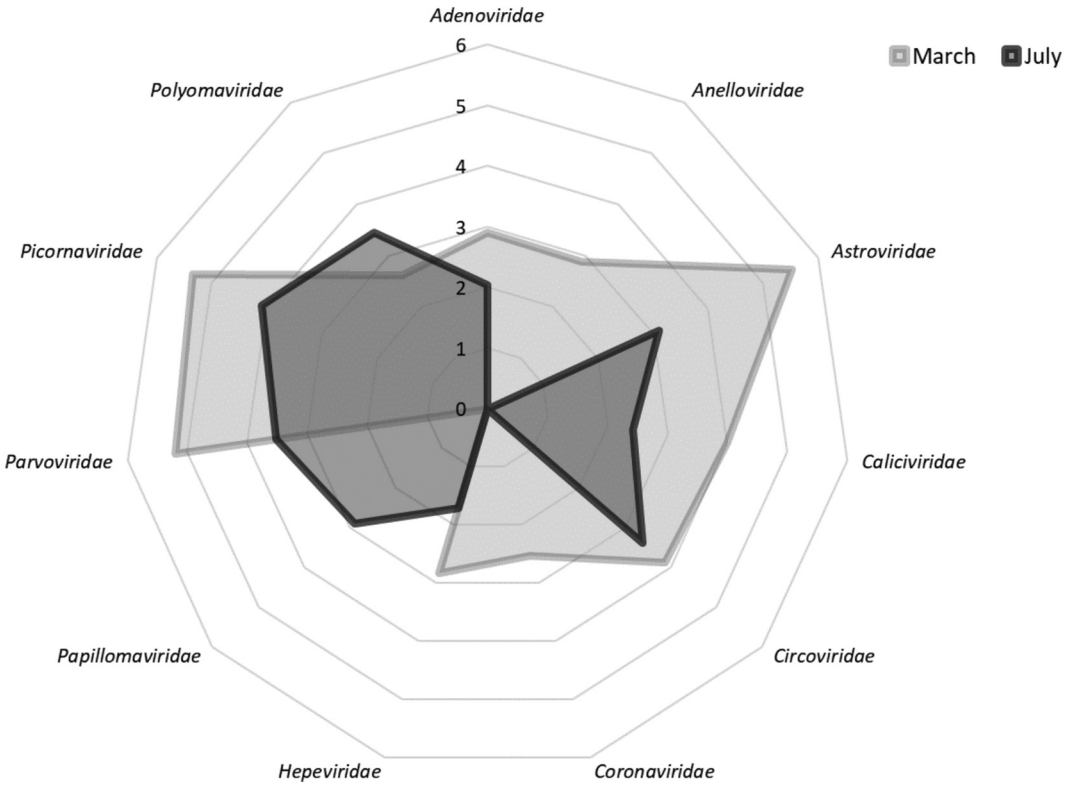


Fig. 2. Diversity and relative abundance of vertebrate viral families reads for March and July samples using Target Enrichment NGS. Relative abundance of reads is represented in log10.

This Alphacoronavirus was firstly described in 2014 in rat samples from China (Wang et al., 2015). At the time, rodents were not considered to be a relevant CoV reservoir, however it is now clear that this group of animals are in fact important reservoirs for alphacoronaviruses and also betacoronaviruses (Wartecki and Rzymski, 2020).

New CoV members in bats has been a topic of interest within the scientific community due to the fact that these viruses could be the origin of important outbreaks. In this field, the use of metagenomics is important for a better understanding of evolution, epidemiology, and host-relationships of zoonotic and human viruses (Kivistö et al., 2020). These approaches made possible the discovery of new alphacoronaviruses (De Sabato et al., 2019; Hall et al., 2014), and recently the description of a betacoronavirus closely related to SARS-CoV-2 in Rhinolophus bats from China (Zhou et al., 2020). The low sensitivity of high throughput techniques or the lack of sequencing depth in samples with a high viral load has revealed enrichment NGS as a successful strategy for CoV surveillance and discovery in Asia (Li et al., 2020; Lim et al., 2019). Thus, NGS for CoV discovery should ideally be wide enough to detect unknown viruses but targeting viruses belonging to the family of interest among many other present in the analyzed samples.

Regarding the wastewater sample from July, its virome composition showed a lower proportion for almost all vertebrate viral families, except for *Polyomaviridae* and *Papillomaviridae*, compared with the sample from March (Fig. 2), members from the *Coronaviridae* family where no detected. The absence of SARS-CoV-2 could be due to the fact that COVID-19 incidence recorded when this sample was collected was low (175.2 cases/100,000 inhabitants), with wastewater quantifications of SARS-CoV-2 two orders of magnitude lower than in March. The absence of HCoV-OC43 and animal CoV reads in the July sample could be explained by winter seasonality or by the fact that the transmission of these viruses could have been reduced due to the massive use of masks and the increase in hand washing from March to July as has been observed for other viruses.

3.3. Comparison of wastewater strains vs. clinical strains

SARS-CoV-2 contigs retrieved after applying targeted NGS to the March wastewater sample were compared with sequences obtained by applying the VirCapSeq enrichment approach to two clinical samples isolated in the same period. A short summary of the sequencing results for the three samples is provided in Supplementary File 2 (see Fig. 3 for an overview of the coverage distribution of the reads mapped over reference genome).

From a total of 5,006,516 paired-end reads, only 8 aligned uniquely at a single location and 20 aligned more than one time over the reference genome for the wastewater sample. On the contrary, VirCapSeq clinical samples starting with less clean paired-end reads, 375,769 and 316,398 for samples A and B, ended up with 1042 and 1856 uniquely mapped reads, respectively. The final scaffolds assembled from those sets of reads recovered up to 6.79% for wastewater sample, 88.38% for clinical sample A, and 89.83% for clinical sample B, of the 29,903 bp of reference SARS-CoV-2 genome sequence.

Other studies have reported the use of VirCapSeq and Twist Bioscience panels for sequencing of clinical samples, determining the

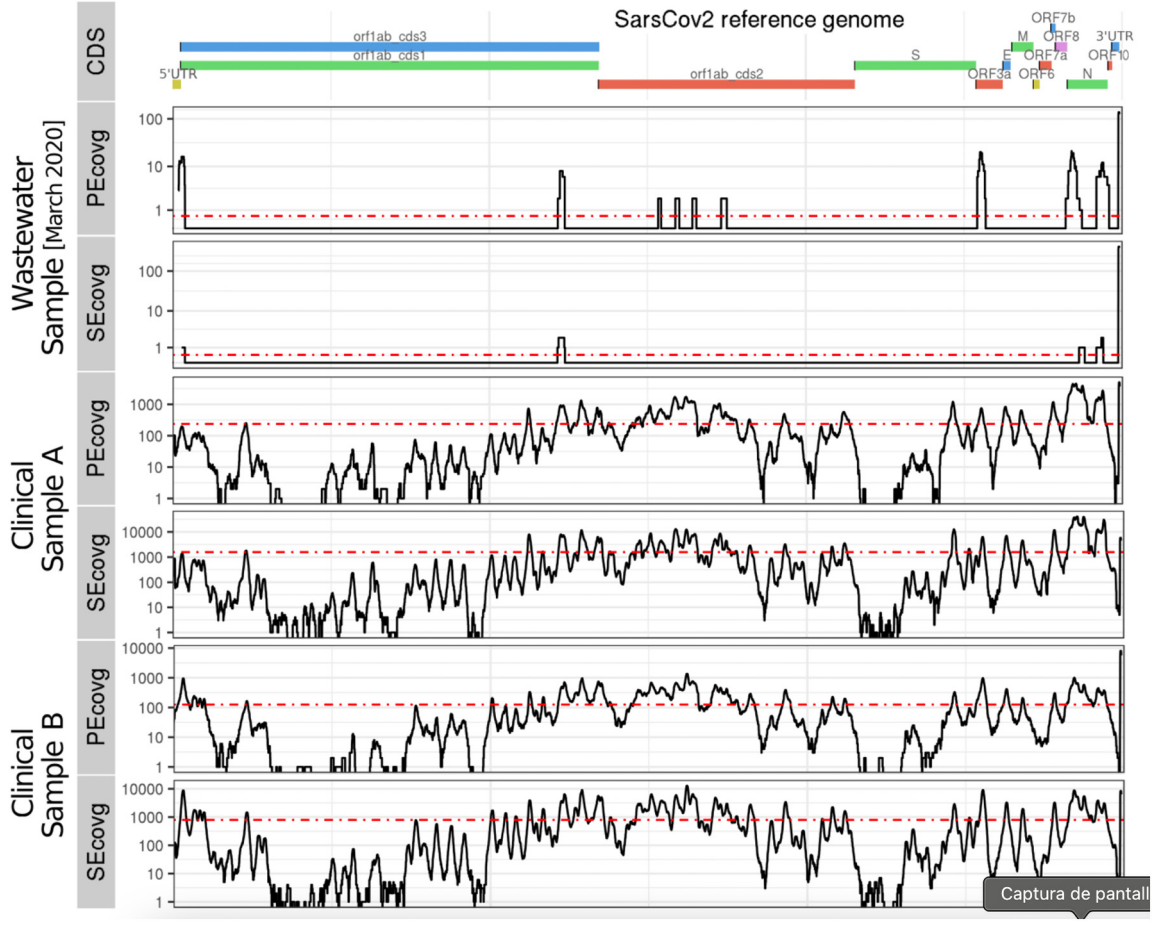


Fig. 3. Summary of reads coverage along the SARS-CoV-2 reference genome sequence. The number of reads per position for the clean Pair-End reads (PEcovg) and for the Single-End reads (SEcovg) is show here for the three samples. SE-reads were produced when cleaning the raw PE-reads, corresponding to the single member of a pair or reads that passed the cleaning and trimming criteria while the other read was discarded. Top track represents the segments defining the open reading frames encoding for the viral proteins (except the trailing 5'UTR and 3'UTR segments that are not protein-coding and were only included for illustrative purposes). The dashed red line represents the average coverage.

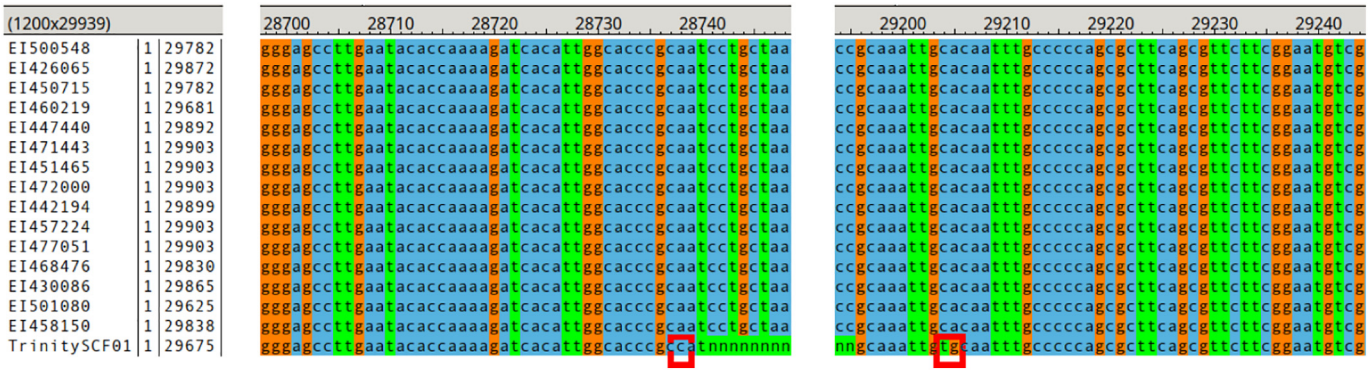


Fig. 4. Nucleotide mismatches found in the nucleocapsid protein region. The red boxes highlight the 3 nucleotides that differ on the sequence reconstructed from the recovered reads with respect to the alignment with 1200 randomly chosen GISAID SARS-CoV-2 sequences (including the reference). Those nucleotide substitutions correspond to 1 synonymous and 2 non-synonymous changes: 28721.c**a**a[Q] <> 28,514.c**C**a[P], 29,187.tg**c**[C] == 28,959.tg**T**[C], 29,188.**a**ca[T] <> 28,960.**G**ca[A]. Those three positions do not change on the alignment for all the other provided sequences yet are supported by only three reads which in fact does not allow defining them as variants. Scale on top corresponds to the positions relative to the full alignment (29,939 columns) that is available as Supplementary File 4.

importance of having low qPCR Ct values in a given sample for obtaining a high genome coverage after conducting NGS (Carbo et al., 2020; Klempt et al., 2020).

SARS-CoV-2 sequences obtained from sewage were compared with other 1200 sequences obtained from clinical isolates available on the GISAID database (Elbe and Buckland-Merrett, 2017) (https://www.gisaid.org/epiflu-applications/hcov-19-genomic-epidemiology/) to elucidate the presence of viral variants. While a high degree of similarity was found for the assembled contigs, 3 nucleotide mismatches were identified within the nucleocapsid protein region (Fig. 4). These divergences could represent single nucleotide variants (SNV), but they were only supported by 3 overlapping reads. More read coverage would be needed in this region to further assess these differences, although it is worth noting that those changes were not described in clinical samples at the moment. Amplicon sequencing panels covering the totality of the genome are available now for studying SARS-CoV-2 genetic variants. These approaches, as well as specific SARS-CoV-2 and Human CoV capture panels, would have enabled the acquisition of sufficient genome coverage and consequently a better SNV typing (Nasir et al., 2020; Xiao et al., 2020) than other targeted approaches directed towards wider groups of viruses, like VirCapSeq capture used in this study. Even so, NGS methods still could not compete with other molecular approaches, as RT-qPCR, as both the time to obtain results and the processing price are higher. In contrast, these sequencing-based methodologies may be more sensitive than RT-qPCR (Charlebois et al., 2020) and give broader information about genetic variants and characteristics of other members from the same family that could not be detected by a single primer amplification-based method. This could be the case of Amplicon Deep Sequencing approaches based on massive sequencing of a PCR produced amplicon. High throughput NGS approaches have as a main advantage that allow detection of unknown viral sequences, in contrast, PCR-based detection relies of previous knowledge and primer design of viral targets. RT-qPCR and NGS-based methods do not necessarily compete but complement each other to provide useful information in terms of wastewater-based epidemiology. Moreover, NGS methods are in constant development, thus it is expected that in a short period of time could compete in terms of time and cost with other molecular approaches.

4. Conclusions

- The application of a targeted NGS has provided nearly the whole genome of SARS-CoV-2 from two clinical samples and eight SARS-CoV-2 contigs from one wastewater sample, both type of samples obtained in March 2020 from the same geographical area.
- Because of low genome coverage obtained in the wastewater sample, the 3 nucleotide differences observed among environmental and clinical samples could not be further assessed but results obtained show a high degree of similarity between environmental and clinical samples.
- The results of the excreted virome in wastewater showed that there is co-circulation, in urban areas, of human and animal coronaviruses infecting domestic animals and rodents.
- The use of a Target Enrichment panel designed to cover vertebrate viruses allowed the acquisition of SARS-CoV-2 sequences as part of the wastewater virome within the context of a COVID-19 pandemic.
- Specific SARS-CoV-2/human CoV panels should be used for studying SARS-CoV-2 and other Human CoV genetic diversity while panels targeting a wide variety of viral families would be a better choice for evaluating the co-circulation of known and unknown human and animal CoV, which may be of relevance regarding potentially zoonosis and viral discovery.
- SARS-CoV-2 can be now considered a new member of the sewage virome. Its presence in this type of samples after global vaccination campaigns should be further evaluated as well as the presence of other human CoV that could change their circulation patterns due to immune cross reactivity phenomena.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2021.149562.

Availability of data and material

The NGS datasets generated during the current study are available in zenodo under the DOI number https://doi.org/10.5281/zenodo.4252765; raw sequencing reads for all the analyzed samples can be also accessed from NCBI-SRA Bioproject PRJNA689221.

Funding

This study was partially supported by the Ministry of Science, Innovation and Universities (AGL2017-86797-C2-1-R) through the University of Barcelona and the Direcció General de Recerca i Innovació en Salut (DGRIS) Catalan Health Ministry Generalitat de Catalunya through Vall d'Hebron Research Institute (VHIR). Sílvia Bofill-Mas is a Serra-Hunter fellow at the University of Barcelona.

CRediT authorship contribution statement

Conceptualization, S.B-M. and S.M-P.; methodology, S.M—P, M.I, M. R, E.F, C.M; software, J.A., S.M—P; validation, S.B-M, R.G and J.A.; formal

analysis, S.M—P, S.B-M and J.A; investigation, S.M-P and S.B-M.; resources, S.B-M, R.G, J.Q, A.A.; data curation, S.M-P and C.A; writing—original draft preparation, S.M-P and S.B-M.; writing—review and editing, S.M—P, S.B-M., M.I, M.R, E.F, C.M, C.A, A.A. J.Q, R.G a J.A; visualization, S.M-P and J.A.; supervision, S.B-M.; project administration, S.B-M, R.G, A.A, J.Q.; funding acquisition, S.B-M, R.G, A.A, J.Q. All authors have read and agreed to the published version of the manuscript.

Declaration of competing interest

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

Acknowledgments

We thank the members of ICRA and the sewage treatment plants personnel for the coordination of the wastewater sampling. We also gratefully acknowledge the authors originating and submitting laboratories of the SARS-Cov-2 sequences from GISAID's EpiCov Database (https://www.epicov.org/epi3/) used to build the Fig. 4 alignment; access to the individual isolates is facilitated through GISAID web site (https://www.gisaid.org/).

References

Barnett, D.W., Garrison, E.K., Quinlan, A.R., Str¨mberg, M.P., Marth, G.T., 2011. Bamtools: a C API and toolkit for analyzing and managing BAM files. Bioinformatics 27, 1691–1692. https://doi.org/10.1093/bioinformatics/btr174.

Bofill-Mas, S., Albinana-Gimenez, N., Clemente-Casares, P., Hundesa, A., Rodriguez-Manzano, J., Allard, A., Calvo, M., Girones, R., 2006. Quantification and stability of human adenoviruses and polyomavirus JCPyV in wastewater matrices. Appl. Environ. Microbiol. https://doi.org/10.1128/AEM.00965-06.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Briese, T., Kapoor, A., Mishra, N., Jain, K., Kumar, A., Jabado, O.J., Ian Lipkina, W., 2015. Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. MBio 6, 1–12. https://doi.org/10.1128/mBio.01491-15.

Cantalupo, P.G., Calgua, B., Zhao, G., Hundesa, A., Wier, A.D., Katz, J.P., Grabe, M., Hendrix, R.W., Girones, R., Wang, D., Pipas, J.M., 2011. Raw sewage harbors diverse viral populations. MBio 2. https://doi.org/10.1128/mBio.00180-11 e00180-11.

Carbo, E.C., Sidorov, I.A., Zevenhoven-Dobbe, J.C., Snijder, E.J., Claas, E.C., Laros, J.F.J., Kroes, A.C.M., de Vries, J.J.C., 2020. Coronavirus discovery by metagenomic sequencing: a tool for pandemic preparedness. J. Clin. Virol. 131, 104594. https://doi.org/10.1016/j.jcv.2020.104594.

Charlebois, R.L., Sathiamoorthy, S., Logvinoff, C., Gisonni-Lex, L., Mallet, L., Ng, S.H.S., 2020. Sensitivity and breadth of detection of high-throughput sequencing for adventitious virus detection. npj Vaccines 5, 1–8. https://doi.org/10.1038/s41541-020-0207-4.

De Sabato, L., Lelli, D., Faccin, F., Canziani, S., Di Bartolo, I., Vaccari, G., Moreno, A., 2019. Full genome characterization of two novel alpha-coronavirus species from italian bats. Virus Res. 260, 60–66. https://doi.org/10.1016/j.virusres.2018.11.007.

Elbe, S., Buckland-Merrett, G., 2017. Data, disease and diplomacy: GISAID's innovative contribution to global health. Glob. Chall. 1, 33–46. https://doi.org/10.1002/gch2.1018.

Fernandez-Cassi, X., Timoneda, N., Martínez-Puchol, S., Rusiñol, M., Rodriguez-Manzano, J., Figuerola, N., Bofill-Mas, S., Abril, J.F., Girones, R., 2018. Metagenomics for the study of viruses in urban sewage as a tool for public health surveillance. Sci. Total Environ. 618, 870–880. https://doi.org/10.1016/j.scitotenv.2017.08.249.

Grabherr, M.G., Haas, Brian J., Joshua, Moran Yassour, Levin, Z., Thompson, Dawn A., Amit, Ido, Adiconis, Xian, Fan, Lin, Raychowdhury, Raktima, Zeng, Qiandong, Chen, Zehua, Mauceli, Evan, Hacohen, Nir, Gnirke, Andreas, Rhind, Nicholas, di Palma, Federica, Bruce, W., N. and A.R., Friedman, 2013. Trinity: reconstructing a full-length transcriptome without a genome from RNA-seq data. Nat. Biotechnol. 29, 644–652. https://doi.org/10.1038/nbt.1883.Trinity.

Gundy, P.M., Gerba, C.P., Pepper, I.L., 2009. Survival of coronaviruses in water and wastewater. Food Environ. Virol. 1, 10–14. https://doi.org/10.1007/s12560-008-9001-6.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Philip, D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Macmanes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., Leduc, R.D., Friedman, N., Regev, A., 2013. De novo transcript sequence recostruction from RNA-seq: reference generation and analysis with trinity. Nat. Protoc. https://doi.org/10.1038/nprot.2013.084.De.

Hall, R.J., Wang, J., Peacey, M., Moore, N.E., McInnes, K., Tompkins, D.M., 2014. New alphacoronavirus in Mystacina tuberculata bats, New Zealand. Emerg. Infect. Dis. 20, 697–700. https://doi.org/10.3201/eid2004.131441.

Hjelmsø, M.H., Mollerup, S., Jensen, R.H., Pietroni, C., Lukjancenko, O., Schultz, A.C., Aarestrup, F.M., Hansen, A.J., 2019. Metagenomic analysis of viruses in toilet waste from long distance flights—A new procedure for global infectious disease surveillance. PLoS One 14, e0210368. https://doi.org/10.1371/journal.pone.0210368.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. https://doi.org/10.1093/molbev/mst010.

Kivistö, I., Tidenberg, E.M., Lilley, T., Suominen, K., Forbes, K.M., Vapalahti, O., Huovilainen, A., Sironen, T., 2020. First report of coronaviruses in northern european bats. Vector-Borne Zoonotic Dis. 20, 155–158. https://doi.org/10.1089/vbz.2018.2367.

Klempt, P., Brož, P., Kašný, M., Novotný, A., Kvapilová, K., Kvapil, P., 2020. Performance of targeted library preparation solutions for SARS-CoV-2 whole genome analysis. Diagnostics 10, 769. https://doi.org/10.3390/diagnostics10100769.

Krafcikova, P., Silhan, J., Nencka, R., Boura, E., 2020. Structural analysis of the SARS-CoV-2 methyltransferase complex involved in RNA cap creation bound to sinefungin. Nat. Commun. 11, 1–7. https://doi.org/10.1038/s41467-020-17495-9.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with bowtie 2. Nat. Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923.

Li, B., Si, H.-R., Zhu, Y., Yang, X.-L., Anderson, D.E., Shi, Z.-L., Wang, L.-F., Zhou, P., 2020. Discovery of bat coronaviruses through surveillance and probe capture-based next-generation sequencing. mSphere 5, 1–10. https://doi.org/10.1128/msphere.00807-19.

Lim, X.F., Lee, C.B., Pascoe, S.M., How, C.B., Chan, S., Tan, J.H., Yang, X., Zhou, P., Shi, Z., Sessions, O.M., Wang, L.F., Ng, L.C., Anderson, D.E., Yap, G., 2019. Detection and characterization of a novel bat-borne coronavirus in Singapore using multiple molecular approaches. J. Gen. Virol. 100, 1363–1374. https://doi.org/10.1099/jgv.0.001307.

Martínez-Puchol, S., Rusiñol, M., Fernández-Cassi, X., Timoneda, N., Itarte, M., Andrés, C., Antón, A., Abril, J.F., Girones, R., Bofill-Mas, S., 2020. Characterisation of the sewage virome: comparison of NGS tools and occurrence of significant pathogens. Sci. Total Environ. 713. https://doi.org/10.1016/j.scitotenv.2020.136604.

Nasir, J.A., Kozak, R.A., Aftanas, P., Raphenya, A.R., Smith, K.M., Maguire, F., Maan, H., Alruwaili, M., Banerjee, A., Mbareche, H., Alcock, B.P., Knox, N.C., Mossman, K., Wang, B., Hiscox, J.A., McArthur, A.G., Mubareka, S., 2020. A comparison of whole genome sequencing of SARS-CoV-2 using amplicon-based sequencing, random hexamers, and bait capture. Viruses 12. https://doi.org/10.3390/v12080895.

Ng, T.F.F., Marine, R., Wang, C., Simmonds, P., Kapusinszky, B., Bodhidatta, L., Oderinde, B.S., Wommack, K.E., Delwart, E., 2012. High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage. J. Virol. 86, 12161–12175. https://doi.org/10.1128/JVI.00869-12.

Pratelli, A., 2006. Genetic evolution of canine coronavirus and recent advances in prophylaxis. Vet. Res. 37, 191–200. https://doi.org/10.1051/vetres:2005053.

Rusiñol, M., Zammit, I., Itarte, M., Forés, E., Martínez-Puchol, S., Girones, R., Borrego, C., Corominas, L., Bofill-Mas, S., 2021. Monitoring waves of the COVID-19 pandemic: inferences from WWTPs of different sizes. Sci. Total Environ. 787, 147463. https://doi.org/10.1016/j.scitotenv.2021.147463.

Shu, Y., McCauley, J., 2017. GISAID: global initiative on sharing all influenza data – from vision to reality. Eurosurveillance 22, 2–4. https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494.

Van der Hoek, L., 2015. Human coronaviruses: what do they cause? Antivir. Ther. 12.

Vilsker, M., Moosa, Y., Nooij, S., Fonseca, V., Ghysens, Y., Dumon, K., Pauwels, R., Alcantara, L.C., Vanden Eynden, E., Vandamme, A.-M., Deforche, K., de Oliveira, T., 2018. Genome detective: an automated system for virus identification from high-throughput sequencing data. Bioinformatics 1–3. https://doi.org/10.1093/bioinformatics/bty695.

Vogel, L., Van Der Lubben, M., Te Lintelo, E.G., Bekker, C.P.J., Geerts, T., Schuijff, L.S., Grinwis, G.C.M., Egberink, H.F., Rottier, P.J.M., 2010. Pathogenic characteristics of persistent feline enteric coronavirus infection in cats. Vet. Res. 41. https://doi.org/10.1051/vetres/2010043.

Wang, W., Lin, X., Guo, W., Zhou, R., Wang, M., Wang, C.-Q., Ge, S., Mei, S.-H., Li, M.-H., Shi, M., Holmes, E.C., Zhang, Y.-Z., 2015. Discovery, diversity and evolution of novel coronaviruses sampled from rodents in China. Virology 474, 19–27. https://doi.org/10.1016/j.virol.2014.10.017.

Wartecki, A., Rzymski, P., 2020. On the coronaviruses and their associations with the aquatic environment and wastewater. Water (Switzerland) 12, 1–27. https://doi.org/10.3390/W12061598.

Wölfel, R., Corman, V.M., Guggemos, W., Seilmaier, M., Zange, S., Müller, M.A., Niemeyer, D., Jones, T.C., Vollmar, P., Rothe, C., Hoelscher, M., Bleicker, T., Brünink, S., Schneider, J., Ehmann, R., Zwirglmaier, K., Drosten, C., Wendtner, C., 2020. Virological assessment of hospitalized patients with COVID-2019. Nature 581, 465–469. https://doi.org/10.1038/s41586-020-2196-x.

Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z.W., Tian, J.H., Pei, Y.Y., Yuan, M.L., Zhang, Y.L., Dai, F.H., Liu, Y., Wang, Q.M., Zheng, J.J., Xu, L., Holmes, E.C., Zhang, Y.Z., 2020. A new coronavirus associated with human respiratory disease in China. Nature 579, 265–269. https://doi.org/10.1038/s41586-020-2008-3.

Wu, Y., Guo, C., Tang, L., Hong, Z., Zhou, J., Dong, X., Yin, H., Xiao, Q., Tang, Y., Qu, X., Kuang, L., Fang, X., Mishra, N., Lu, J., Shan, H., Jiang, G., Huang, X., 2020. Prolonged presence of SARS-CoV-2 viral RNA in faecal samples. Lancet Gastroenterol. Hepatol. 5, 434–435. https://doi.org/10.1016/S2468-1253(20)30083-2.

Xiao, M., Liu, X., Ji, J., Li, M., Li, Jiandong, Yang, L., Sun, W., Ren, P., Yang, G., Zhao, J., Liang, T., Ren, H., Chen, T., Zhong, H., Song, W., Wang, Y., Deng, Z., Zhao, Y., Ou, Z., Wang, D., Cai, J., Cheng, X., Feng, T., Wu, H., Gong, Y., Yang, H., Wang, J., Xu, X., Zhu, S., Chen, F., Zhang, Y., Chen, W., Li, Y., Li, Junhua, 2020. Multiple approaches for massively parallel sequencing of SARS-CoV-2 genomes directly from clinical samples. Genome Med. 12, 1–15. https://doi.org/10.1186/s13073-020-00751-4.

Ye, Z.W., Yuan, S., Yuen, K.S., Fung, S.Y., Chan, C.P., Jin, D.Y., 2020. Zoonotic origins of human coronaviruses. Int. J. Biol. Sci. 16, 1686–1697. https://doi.org/10.7150/ijbs.45472.

Zhou, H., Chen, X., Hu, T., Li, J., Song, H., Liu, Y., Wang, P., Liu, D., Yang, J., Holmes, E.C., Hughes, A.C., Bi, Y., Shi, W., 2020. A novel bat coronavirus closely related to SARS-CoV-2 contains natural insertions at the S1/S2 cleavage site of the spike protein. Curr. Biol. 30, 2196–2203.e3. https://doi.org/10.1016/j.cub.2020.05.023.