Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company’s public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Spatial and temporal distribution of SARS-CoV-2 diversity circulating in wastewater

Alba Pérez-Cataluña a,b, Álvaro Chiner-Oms b, Enric Cuevas-Ferrando a, Azahara Díaz-Reolid a, Irene Falcó a, Walter Randazzo b, Inés Girón-Guzmán a, Ana Allende a, María A. Bracho a,c,d,e,f, Inaki Comas b,c, Gloria Sánchez a

a VISAFELab, Department of Preservation and Food Safety Technologies, Institute of Agrochemistry and Food Technology, IATA-CSIC, Av. Agustín Escardino 7, Paterna, Valencia 46980, Spain
b Instituto de Biomecánica de Valencia (IBV-CSIC), C/ Jaume Roig, 11, Valencia 46010, Spain
c Department of Food Science and Technology, CEBAS-CSIC, Research Group on Quality and Safety of Fruits and Vegetables, Campus Universitario de Espinardo, 25, Murcia 30100, Spain
d FISABIO - Public Health, Department of Genomics and Health, Av. Catalunya, 21, Valencia 46020, Spain
e Joint Research Unit "Infection and Public Health" FISABIO-Universitat de Valencia IDSyB, Av. Catalunya, 21, Valencia 46020, Spain
f CIBER in Epidemiology and Public Health (CIBERESP), Valencia, Spain

ARTICLE INFO

Keywords:
SARS-CoV-2
Wastewater
Genome sequencing
Spike mutations
Variants of concern
Variants of interest

ABSTRACT

Wastewater-based epidemiology (WBE) has proven to be an effective tool for epidemiological surveillance of SARS-CoV-2 during the current COVID-19 pandemic. Furthermore, combining WBE together with high-throughput sequencing techniques can be useful for the analysis of SARS-CoV-2 viral diversity present in a given sample. The present study focuses on the genomic analysis of SARS-CoV-2 in 76 sewage samples collected during the three epidemiological waves that occurred in Spain from 14 wastewater treatment plants distributed throughout the country. The results obtained demonstrate that the metagenomic analysis of SARS-CoV-2 in wastewater allows the detection of mutations that define the B.1.1.7 lineage and the ability of the technique to anticipate the detection of certain mutations before they are detected in clinical samples. The study proves the usefulness of sewage sequencing to track Variants of Concern that can complement clinical testing to help in decision-making and in the analysis of the evolution of the pandemic.

1. Introduction

The family Coronaviridae is a family of enveloped RNA viruses generally associated with mild respiratory and gastrointestinal infections (Shang et al., 2020). Nevertheless, in recent decades new and highly pathogenic zoonotic coronavirus (CoVs) have emerged such as the Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) (Drosten et al., 2003; Ksiazek et al., 2003), the Middle East Respiratory Syndrome Coronavirus (MERS-CoV) (Zaki et al., 2012) and, most recently, SARS-CoV-2 (Zhu et al., 2020) which has resulted in the Coronavirus Disease 2019 (COVID-19) pandemic. Transmission of SARS-CoV-2 occurs mainly through aerosols or respiratory secretions (Chan et al., 2020) but it has also been found that, due to its replication capacity in the gastrointestinal tract (Xiao et al., 2020), it is excreted in feces and urine, as was previously reported for its counterparts SARS-CoV and MERS-CoV. For this reason, it has been possible to detect the genetic material of the virus in the feces of not only symptomatic, but also asymptomatic people (Polo et al., 2020). These findings have led to the use of wastewater monitoring for SARS-CoV-2. As for other pathogens, the use of Wastewater-Based Epidemiology (WBE) has proven to be a very useful tool as an early detection warning system, allowing trend-estimations as well as establishing correlations between different epidemiological indicators (Bivins et al., 2020; Medema et al., 2020; Randazzo et al., 2020b, 2020a). One of the reasons for the success of WBE is that wastewater samples are a non-invasive and inexpensive source of information to investigate the spread of SARS-CoV-2 within a community. Moreover, it provides real-time information on the circulating lineages of SARS-CoV-2, which is essential for the development of vaccines and drugs. This is particularly relevant in view of the current situation where the world’s population is being vaccinated against...
SARS-CoV-2 and where, due to the appearance of emerging lineages, vaccine effectiveness might be compromised (Zhou et al., 2021).

Massive parallel sequencing techniques applied to sewage samples allow us to analyze a large number of SARS-CoV-2 genomes, including those present in symptomatic and asymptomatic persons. Through the analysis of sequences, it is possible to detect low-frequency variants (LFV) and to infer which lineages are circulating at a certain time and place (Bar-Or et al., 2021; Crits-Christoph et al., 2021; Dharmadhikari et al., 2021; Herold et al., 2021; Izquierdo-Lara et al., 2021; La Rosa et al., 2021; Nemudryi et al., 2020; Rios et al., 2021). Additionally, genomic analyses may allow to detect the entry of described lineages or Variants of Concern (VOCs) into populations, as well as the appearance of emerging lineages, to characterize new outbreaks, and to aid in viral strains tracking (Bar-Or et al., 2021; Crits-Christoph et al., 2021; Izquierdo-Lara et al., 2021; La Rosa et al., 2021; Nemudryi et al., 2020; Rios et al., 2021). These studies also evidenced that improvement on sequencing techniques must be performed in order to reduce error rates, as the case of Nanopore sequencing (Nemudryi et al., 2020). Despite these limitations, the published works showed that genomic analysis of SARS-CoV-2 in wastewater should be used as a complementary tool in epidemiological surveillance. This aspect has grown in significance because during the spread of SARS-CoV-2, different mutations (i.e. D614G, Δ69/70, N501Y, E484K, K417N) present in VOCs (i.e. B.1.1.7, B.1.351, B.1.617.2, and B.1.1.28.1) have emerged (https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/), which have a considerable impact on transmissibility, infection severity (Singh et al., 2021), or immunity. These characteristics, if they occur, can aggravate the epidemiological situation in certain areas, so the detection of new lineages and the appearance of VOCs in any specific population is crucial to overcome the current pandemic situation and control the spread of the virus. Variants of interest (VOI) (for example, B.1.427) have also been defined which are currently under investigation and must be monitored to ensure a prompt response should they pose a greater threat to the population. The usefulness of these techniques is evident from the fact that the European Commission published, on March 17, 2021, recommendations for the establishment of SARS-CoV-2 surveillance in wastewater, highlighting the importance of SARS-CoV-2 sequencing in wastewater as a tool for the detection of VOC and VOI (https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=uriserv:OJ.L_.2021.098.01.000.01.ENG). The aim of this study was to analyze SARS-CoV-2 genomes in wastewater through high-throughput sequencing in order to monitor the emergence of mutations, lineages or the detection of signature mutations of VOCs and VOIs.

2. Materials and methods

2.1. Sample processing

In the framework of SARS-CoV-2 wastewater monitoring in Spain, grab samples were collected from 14 treatment plants located in different parts of the Spanish territory, with equivalent inhabitant values ranging from 60,600 to 1900,800. The samples taken between April 2020 and January 2021 encompass the three waves that have affected the country. The first wave occurred between March and April of 2020, the second wave in November 2020, and the third wave between January and February 2021. For each sample, 200 mL of wastewater samples were artificially inoculated with porcine epidemic diarrhea virus (PEDV) as process control with a final concentration of 4.5 log (PCRU/L), and concentrated following an aluminum-based adsorption precipitation method (AAVV, 2018; Pérez-Cataluña et al., 2021; Randazzo et al., 2020b). Then, 200 mL of wastewater was adjusted to pH 6.0. Precipitation by Al(OH)$_3$ was carried out by mixing 1 part of 0.9 N AlCl$_3$ per 100 parts of sample. The resulting precipitate was resuspended in 10 mL of 3% beef extract (pH 7.4) and centrifuged at 150 rpm for 10 min at room temperature (RT). The supernatant was centrifuged at 1900 × g for 30 min at 6°C and the pellet resuspended in 1 mL of phosphate buffered saline solution (PBS, pH 7.4). After this, concentrated samples were stored at −80°C until analysis.

2.2. Nucleic acid extraction and SARS-CoV-2 RT-qPCR quantification

Nucleic acid extraction from wastewater concentrates was performed using an automated method with the Maxwell RSC Pure Food GMO and authentication kit (Promega) with slight modifications (Pérez-Cataluña et al., 2021). Firstly, 300 μL of concentrated samples were mixed with 400 μL of cetyltrimethyl ammonium bromide (CTAB) and 40 μL of proteinase K solution. The mixed sample was incubated at 60°C for 10 min and centrifuged for 10 min at 16,000 × g. Next, the resulting supernatant was transferred to the loading cartridge and 300 μL of lysis buffer added. The cartridge was then loaded in the Maxwell® RSC Instrument (Promega) using the ‘Maxwell RSC Viral total Nucleic Acid’ running program for the nucleic acid extraction. The obtained RNA was eluted in 100 μL nuclease-free water. Negative controls were included by using nuclease-free water instead of concentrated sample.

SARS-CoV-2 nucleic acid detection was performed by RT-qPCR using One Step PrimeScriptTM RT-PCR Kit (Perfect Real Time) (Takara Bio, USA) targeting a genomic region of the nucelocapsid gene (N1 region) using primers, probes and conditions previously described (CDC, 2020). The complete genomic RNA of SARS-CoV-2 (ATCC VR-1986D) and nuclease free water were used as positive and negative controls, respectively.

2.3. SARS-CoV-2 genome sequencing and analysis

Samples with RT-qPCR cycle threshold (Ct) values below 36 were selected for sequencing analysis. Genomic sequencing of SARS-CoV-2 present in selected wastewater samples was carried out following ARTIC protocol version 3 for retrotranscription and amplification by multiplex PCR (Quick, 2020; https://www.protocols.
A. Pérez-Cataluna et al.

| Wave | ORF1a polyprotein | ORF1b | ORF3a | Intergenic region | ORF5a | Intergenic region | ORF6 | Intergenic region | ORF7a | Intergenic region | Nucleocapsid protein | Membrane protein | ORF10 protein |
|------|------------------|-------|-------|------------------|-------|------------------|------|------------------|-------|------------------|-------------------|----------------|-------------|
| 1st and 2nd | NA | 32 | 15 | 6 | 1 | 6 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 3rd | NA | 52 | 43 | 14 | 4 | 2 | 1 | NA | 0 | 0 | 0 | 0 | 0 |
| Total | NA | 138 | 100 | 58 | 21 | 14 | 16 | 23 | 0 | 0 | 0 | 0 | 0 |

### Table 1: Overview of the nucleotide substitutions and deletions detected in SARS-CoV-2 genomes from wastewater samples (n = 76) as compared to the SARS-CoV-2 isolate Wuhan-Hu-1 reference genome (MN908947.3). NA, not applicable.

| Region | ORF1a polyprotein | ORF1b | ORF3a | Intergenic region | ORF5a | Intergenic region | ORF6 | Intergenic region | ORF7a | Intergenic region | Nucleocapsid protein | Membrane protein | ORF10 protein |
|--------|------------------|-------|-------|------------------|-------|------------------|------|------------------|-------|------------------|-------------------|----------------|-------------|
| ORF1a | Δ21 | Δ18 | Δ17 | Δ16 | Δ15 | Δ14 | Δ13 | Δ12 | Δ11 | Δ10 | Δ9 | Δ8 | Δ7 |
| ORF1b | Δ12 | Δ11 | Δ10 | Δ9 | Δ8 | Δ7 | Δ6 | Δ5 | Δ4 | Δ3 | Δ2 | Δ1 | Δ0 |

3. Results and discussion

#### 3.1. SARS-CoV-2 quantification and genome coverage

A total of 76 sewage samples positive for SARS-CoV-2 by RT-qPCR (Ct < 36) collected throughout the three epidemiological waves were sequenced during this study (Supplementary Fig. S1). Samples were grouped in three regions: north (2 WWTPs, n = 8), center (7 WWTPs, n = 39), and south (5 WWTPs, n = 29). Results showed Ct values of SARS-CoV-2 target N1 ranged from 26.59 to 34.75 (Table S1). From the 76 sequenced samples, 11 (14.5%) showed percentages of 20X coverage values higher than 90% (Figs. 1 and S2), and a mean genomic percentage of coverage of 50.1 ± 30.6%. In order to study the potential correlation between viral loads and genome coverage in wastewater samples, correlation analyses between RT-qPCR outputs (genome copies (gc) per liter) versus genome coverage were carried out for each individual sample. No correlations were found for the analyzed target, as occurred in the study of Izquierdo-Lara et al. (2021) for Illumina reads. Fig. 1 shows the number of samples that covered a certain nucleotide position with depths higher than 20X among the samples with genomic coverage greater than 20% (n = 59). However, some areas were only covered by less than 20 samples (< 33.90%), as is the case of two regions at the end of ORF 1b (nucleotides 21,456–21,467 and several regions between nucleotides 21,162 and 21,600), the regions of the S gene from nucleotides 22,303 to 22,342 and from 22,364 to 22,523, most of the ORF7a (nucleotides 27,529–27,790), and the central region of the N gene (nucleotides 28,773–28,853 and 28,901–28,993) (Fig. 1).

#### 3.2. Overview of detected nucleotide substitutions and deletions

Sequence analysis showed a total of 627 nucleotide substitutions and 20 deletions (Table 1) in comparison with the reference genome of SARS-CoV-2 isolate Wuhan-Hu-1 (MN908947.3). Among detected nucleotide substitutions: 248 were found in ORF1a polyprotein; 171 in ORF1b; 71 in the spike glycoprotein; 32 in ORF3a; 29 in the membrane glycoprotein; 20 in ORF8; 39 in the nucleocapsid gene; 31 in intergenic regions; 3 in ORF10; and one in the envelope protein, ORF6, and ORF7a each (Table 1, Fig. 2). Regarding deletions, a total of 8 deletions were found in samples of the first and second waves: 5 of them in the ORF1a region (Δ21–23, Δ82–84, Δ84–86, Δ141–143, and Δ682); one in the spike glycoprotein (Δ385); and two in the ORF3a (Δ80 and Δ111–20). Two of these deletions were found in two samples: ORF1a:Δ141–143 in samples N1–18–2020 and C3–40–2020; and ORF3a:Δ11–20 in samples

io/view/ncov-2019-sequencing-protocol-v3-locost-bh42j8ye).
3.3. Nucleotide substitutions and deletions in the spike glycoprotein gene

Table 2 shows the non-synonymous nucleotide substitutions (n = 49) and deletions (n = 3) found in the spike glycoprotein gene. Among these polymorphisms, 18 of them were not previously described in genomes obtained from Spanish sequences, according to the database available at https://outbreak.info/ (Mullen et al., 2020). However, two of these nucleotide substitutions (amino acid substitutions G404V and G648V) have been found at low frequencies among the reads obtained in the sequencing of Spanish genomes from clinical samples. These results evidence the ability of this technique to detect mutations that are in low percentage in the viral population and from different lineages.

Interestingly, some of these amino acid substitutions in the spike protein were found in sewage at the same time or even weeks or months before their appearance in genomes from clinical samples. For example, among nucleotide substitutions that have been detected in Spain in clinical samples, two spike mutations (G639S and V642G) were found in waters around the same time that they appeared in clinical genomes, while spike mutations A648V was found in waters 6 weeks before, and mutations S884F, G404V, and A372T were found in waters between 4 and 5 months before their detection in clinical genomes. It should be noted that, in the case of G404V, its first detection occurred at very low percentages of sequencing reads in some clinically obtained genomes (n = 2), and its appearance at higher frequencies in one clinical genome was 5 months later. Additionally, for these genomic mutations, the number of clinical cases was very low, ranging from 1 to 6 cases. Moreover, mutations A893T, L1152S, and N1173K had not been detected in Spanish clinical genomes but their detection in other countries occurred after detection in Spanish wastewater, more specifically 3, 4, and 8 months before, respectively. These results, along with those obtained by other authors who found genomes and single nucleotide polymorphisms (SNPs) widely described in the clinical samples (Critic-Christoph et al., 2021; Izquierdo-Lara et al., 2021), show that high-throughput sequencing of SARS-CoV-2 in wastewater is a very useful complementary tool for studies and decision-making related to the epidemiology of the virus.

3.4. Identification of B.1.1.7 (VOC 202,012/01) mutation signatures

The highly transmissible B.1.1.7 lineage of SARS-CoV-2 contains 16 characteristic non-synonym nucleotide substitutions and deletions (Rambaut et al., 2020) and was first detected in Spain in week 52 of
Table 2
Non-synonymous nucleotide substitutions and deletions detected in the spike glycoprotein region as compared to the SARS-CoV-2 isolate Wuhan-Hu-1 reference genome (MN908947.3). Reference (ref) and alternative (alt) depth relate to the percentage of the total depth that corresponded to the nucleotide present in the reference genome MN908947.3 and the alternative nucleotide, respectively. Mixed samples related to the number of samples showing nucleotides according to reference and alternative sequence. Light gray, region of the receptor binding domain; dark gray, region of the S1/S2 cleavage region. NA, not applicable.

| Positiona | No. Samples | No. mixed samples | Wave | % ref deep | % alt deep | Aminosacrd Change | Previously described in Spain (No. genomes) | Lineagesb |
|-----------|-------------|-------------------|------|------------|------------|------------------|------------------------------------------|-----------|
| 21701     | 1           | 1                 | 3    | 85%        | 15%        | V47L            | No                                       | < 0.5%    |
| 21764     | 2           | 2                 | 3    | See Figure 3 | Δ69/70     | Yes (13,739)    | B.1.258, B.1.1.7, B.1.620, B.1.525, A.28, B.1.1.298, B.1.375, B.1.388 |
| 21852     | 1           | 0                 | 3    | 0%         | 100%       | K97M            | No                                       | < 0.5%    |
| 21855     | 6           | 2                 | 3    | Variable   | S98F       | Yes (453)       | B.1.221                                  |
| 21990     | 2           | 2                 | 3    | See Figure 3 | Δ144       | Yes (12,691)    | Xa, B.1.1.482, B.1.1.1.7, B.1.620, B.1.616, B.1.525, B.1.526.1, B.1.1.318 |
| 22227     | 4           | 1                 | 2.3  | Variable   | A222V      | Yes (8,829)     | B.1.177                                  |
| 22276     | 1           | 1                 | 2    | 32%        | 68%        | A372T           | Yes (1)                                  | < 0.5%    |
| 22773     | 1           | 1                 | 2    | 83%        | 17%        | G404V           | Yes (1)                                  | < 0.5%    |
| 22785     | 1           | 1                 | 3    | 36%        | 64%        | R408L           | Yes (4)                                  | B.1.416, B.1.201, B.1.1.354 |
| 22841     | 2           | 2                 | 3    | 84±35%     | 16±35%     | D427Y           | Yes (3)                                  | A.2.5.2   |
| 22933     | 1           | 1                 | 3    | 76%        | 24%        | R457S           | No                                       | < 0.5%    |
| 22950     | 1           | 1                 | 3    | 53%        | 47%        | P463L           | No                                       | B.1.211   |
| 22992     | 2           | 1                 | 3    | Variable   | S477N      | Yes (328)       | B.1.1.160, C.1.1, B.1.1.27, B.1.526, B.1.438, B.1.1.413, B.1.1.445, B.1.620, B.1.404, B.1.227, B.1.1.375, B.1.1.25.2 |
| 23063     | 2           | 1                 | 3    | See Figure 3 | N501Y      | Yes (14,981)    | P.1, Xa, P.3, A.27, B.1.1.7, B.1.1.70, B.1.623, B.1.521, B.1.351, B.1.694 |
| 23069     | 1           | 0                 | 3    | 0%         | 100%       | V503F           | Yes (3)                                  | B.1.194, B.1.587 |
| 23161     | 1           | 0                 | 3    | 0%         | 100%       | L533F           | No                                       | < 0.5%    |
| 23271     | 1           | 1                 | 3    | 80%        | 20%        | A570D           | Yes (14,125)                             | Xa, B.1.1.7 |
| 23277     | 2           | 2                 | 3    | 80±25%     | 20±25%     | T572I           | Yes (228)                                | B.1.177.88, B.1.1.486 |
| 23403     | 23          | 0                 | 2.3  | 0%         | 100%       | D614G           | Yes (36,787)                             | Fixed |
| 23426     | 1           | 1                 | 3    | 95%        | 5%         | V622F           | Yes (2)                                  | B.1.375, B.1.1.335, B.1.1.225, B.1.177.81, B.1.1.282, B.4 |
| 23477     | 3           | 2                 | 2    | Variable   | G699S      | Yes (9)         | B.1.177.33, B.1.224 |
| 23487     | 1           | 1                 | 2    | 41%        | 59%        | V642G           | Yes (3)                                  | B.1.146, B.1.1.517, N.9, B.1.177.42, B.1.1.274 |
| 23505     | 1           | 1                 | 2    | 55%        | 45%        | G648V           | No                                       | < 0.5%    |
| 23528     | 1           | 1                 | 3    | 67%        | 33%        | V656F           | No                                       | < 0.5%    |
| 23580     | 2           | 2                 | 3    | 51±16%     | 49±16%     | S673T           | Yes (45)                                 | B.1.1.433, B.1.1.39, B.1.1.277, B.1.160.32, B.1.36.31, B.1.177.23, Ab.1, B.1.1.434, B.1.160 |
| 23599     | 1           | 0                 | 3    | 0%         | 100%       | N679K           | Yes (27)                                 | B.1.529, B.1.1.435, K.2, B.1.111, B.1.9.5, B.1.1.385, B.1.560 |

(continued on next page)
The characteristic mutations described in the genome of the B.1.1.7 lineage were searched for in our sequencing data. These mutations corresponded to 2 nucleotide substitutions and one deletion in ORF1a, 6 nucleotide substitutions and 2 deletions in spike gene, 3 nucleotide substitutions in ORF8, and one nucleotide substitution in N gene (Fig. 3).

Amino acid substitutions S235F of nucleocapsid protein was not shown because it was absent or not covered. Samples with Ct values below 36 for N1 were analyzed, starting from the week 52 of 2020 up to week 7 in the case of samples from region C5. Among the analyzed samples, only samples from regions S1, S3, C3, C4, and C5 showed characteristic mutations of the B.1.1.7 lineage. None of the samples showed all the 18 markers that were searched for, at the same time. The highest presence

Table 2 (continued)

| 23604 | 1 | 1 | 3 | 33% | 67% | P681H | Yes (4/17) | P2, Av.1, B.1.1.538, Xa.1.474, B.1.1.519, B.1.1.318, B.1.1.37, B.1.1.575, B.1.1.351, 1.1.620, B.1.621, B.1.575, B.1.1.207, B.1.623, B.1.1.522, B.1.1.469 |
| 23612 | 1 | 1 | 3 | 88% | 12% | A684S | Yes (4) | B.1.160.32 |
| 23628 | 1 | 1 | 3 | 81% | 19% | S689N | No | < 0.5% |
| 23709 | 1 | 1 | 3 | 37% | 63% | T716I | Yes (14/982) | B.1.214.3, Xa.1.214.4, B.1.1.1.7, B.1.575.1, B.1.214.2, B.1.575 |
| 23851 | 1 | 1 | 3 | 51% | 49% | L763F | No | B.1.160.27 |
| 24213 | 1 | 1 | 2 | 83% | 17% | S884F | Yes (7) | B.1.1.122, 1.1.351.1, B.1.177.74, B.1.538, B.1.1.513, B.1.1.322, B.1.517 |
| 24227 | 1 | 0 | 3 | 0% | 100% | G889C | No | B.1.340 |
| 24239 | 1 | 1 | 2 | 84% | 16% | A893T | No | B.1.604 |
| 24375 | 1 | 1 | 3 | 79% | 21% | L938P | No | < 0.5% |
| 24479 | 1 | 1 | 3 | 67% | 33% | I973V | No | < 0.5% |
| 24506 | 2 | 2 | 3 | See Figure 3 | S982A | Yes (14/991) | Xa.1.1.1.7, B.1.411, B.1.1 |
| 24654 | 1 | 1 | 3 | 81% | 19% | E1031V | No | < 0.5% |
| 24655 | 1 | 1 | 3 | 81% | 19% | E1031D | No | < 0.5% |
| 24764 | 1 | 0 | 3 | 10% | 90% | V1068F | Yes (30) | B.1.36.12, B.1.1.447, W.1, D.2, B.1.159, B.1.139, C.26, B.1.1.362, B.1.9.4, B.1.36.28, B.1.190.16 |
| 24812 | 1 | 1 | 3 | 78% | 22% | D1084Y | Yes (123) | B.1.1.269, A.15, B.1.1.319, C.9, C.29, B.1.411, B.4, R.2, B.1.326, B.1.595, B.1.560, B.1.1.273, B.1.416, B.1.9, B.1.189, B.1.354, B.1.623, B.6 |
| 24872 | 1 | 1 | 3 | 69% | 31% | V1104L | Yes (45) | B.1.1.341, B.1.160.32, B.1.1.299, B.1.1.228, B.1.177.75, B.1.287, B.1.1.324, B.1.177.89, B.1.178, B.1.219, B.1.177.88, B.1.1.433, A.19 |
| 24914 | 2 | 1 | 3 | See Figure 3 | D1118H | Yes (14/988) | Xa.1.1.7, B.1.620, B.1.1.34, B.4.11, B.1.1.462, B.1.2.213, B.1.1.378, B.1.1.194, B.1.533, B.1.1.135, P.1,1, B.1.1.375, B.1.1.319, B.1.570 |
| 25017 | 1 | 1 | 2 | 85% | 15% | L1152S | No | < 0.5% |
| 25047 | 1 | 1 | 2 | 83% | 17% | P1162R | Yes (40) | P.1, B.1.177.52, B.1.9 |
| 25049 | 1 | 1 | 3 | 82% | 18% | D1163Y | Yes (226) | B.1.1.77, B.1.406, B.1.1.332, B.53, B.1.36.20, B.1.240.2, B.1.1.359, B.1.291, B.6.3 |
| 25081 | 1 | 0 | 2 | 0.00% | 100.00% | N1173K | No | < 0.5% |
| 25088 | 2 | 2 | 3 | 58±12% | 41±12% | V1176F | Yes (672) | P.1, P.2, P.3, B.1.1.28, B.1.1.332, B.1.324, B.1.1.378, B.1.1.775, B.1.466, B.1.1.382, B.1.369, B.1.566, B.1.1.184, B.1.1.775, B.1.1.218, B.1.545, B.1.1.333, N.1 |
| 25244 | 1 | 1 | 3 | 82% | 18% | V1228L | Yes (67) | B.1.258.15, B.1.177.66, L.3, B.1.181, B.1.1.512, B.1.36.21, B.1.1.112, B.1.390, B.1.596, B.1.426, B.1.238.7, C.7, B.1.236, B.1.570, B.1.393, B.1.626, B.1.533, B.4.1, N.3 |

*Nucleotide position in SARS-CoV-2 isolate Wuhan-Hu-1 genome (MN908947.3).

*Lineage, according to PANGO lineages, that carries the mutation (with frequencies higher than 1%) as stated in outbreak.info.
4. Conclusions

SARS-CoV-2 has created a pandemic scenario unprecedented in modern times. The rapid spread of this virus together with the appearance of emerging lineages has also mobilized the scientific community like never before. Its detection in wastewater has been very helpful for the epidemiological study in large populations and is currently being implemented worldwide. However, few studies using mass sequencing have been published.

- The present study describes the mutations found in SARS-CoV-2 genomes isolated from wastewater in 14 different regions of Spain. This is the first study carried out in Spain that analyzes the diversity of SARS-CoV-2 present in wastewater in the three epidemiological waves which occurred between 2020 and 2021.
- These results confirm the potential of sewage sequencing to detect new mutations and lineages of SARS-CoV-2, which is of utmost relevance for the monitoring efforts of emerging vaccine-escape SARS-CoV-2 mutants in the forthcoming post-vaccination era.
- Genomic sequencing of viruses found in wastewater provides complementary results to those of clinical laboratories, as has been demonstrated in various ways such as the confirmation of the initial detection of low number of reads on genomes from clinical specimens that was later confirmed in wastewater samples; the detection of amino acid substitutions in the spike protein weeks or months before their discovery in clinical samples; or the known amino acid substitutions in the spike protein detected for the first time in Spain.
- This technique provides complementary information for SARS-CoV-2 surveillance, allowing both the control of lineages including VOC and VOI already described and the detection and control of new emerging lineages.
- This data supports the hypothesis that the study of wastewater using high-throughput sequencing techniques is a useful and effective tool that can be implemented worldwide in support of public health for the epidemiological control of SARS-CoV-2.

CRediT authorship contribution statement

Alba Pérez-Cataluna: Investigation, Formal analysis, Writing – review & editing, Conceptualization. Álvaro Chiner-Oms: Investigation, Formal analysis, Writing – review & editing. Enric Cuevas-Ferrando: Investigation, Formal analysis, Writing – review & editing. Azahara Diaz-Reolid: Investigation, Formal analysis, Writing – review & editing. Irene Falco: Resources, Writing – review & editing. Walter Randazzo: Writing – review & editing. Ines Giron-Guzman: Resources, Writing – review & editing. Ana Allende: Writing – review & editing. Maria A. Bracho: Writing – review & editing. Inaki Comas: Funding acquisition, Writing – review & editing. Gloria Sanchez: Conceptualization, Funding acquisition, Writing – review & editing.

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.

Acknowledgements

This study was supported by projects “VIRIDIANA” (AGL2017-82909/ AEI/FEDER, UE) and MCEC-WATER (PID2020-116789RB-C42/ AEI/FEDER, UE) funded by Spanish Ministry of Science, Innovation and Universities; CSIC (202070E101), Generalitat Valenciana (Covid-19-SCI), the European Union – NextGenerationEU (SGL2103034), the COVID-19 wastewater surveillance project (VATar COVID19), funded by the Spanish Ministry for the Ecological Transition and the Demographic Challenge and the Spanish...
Ministry of Health, the Instituto de Salud Carlos III (Ref COV20/00140), ERC TB-RECONNECT 101001038 funded by The European Union – NextGenerationEU and the Lagoon project (PROMETEO/2021/044). EC-F is recipient of a predoctoral contract from the MICINN, Call 2018 and AP-C was supported by a postdoctoral fellowship (APOST/21/292). We acknowledge NILSA, FACSIA, and MITECO for authorizing the sampling. The authors thank Agustín Garrido Fernández, and Andrea Lopez de Mota for their technical support.

Supplementary materials
Supplementary material associated with this article can be found in the online version, at doi:10.1016/j.watres.2020.115942.

References
AAVV, A.D.E. 2018. Section 9510 D. Virus concentration by aluminum hydroxide. Ministry of Health, the Instituto de Salud Carlos III (Ref COV20/00140), ERC TB-RECONNECT 101001038 funded by The European Union – NextGenerationEU and the Lagoon project (PROMETEO/2021/044). EC-F is recipient of a predoctoral contract from the MICINN, Call 2018 and AP-C was supported by a postdoctoral fellowship (APOST/21/292). We acknowledge NILSA, FACSIA, and MITECO for authorizing the sampling. The authors thank Agustín Garrido Fernández, and Andrea Lopez de Mota for their technical support.

Supplementary materials
Supplementary material associated with this article can be found in the online version, at doi:10.1016/j.watres.2020.115942.

References
AAVV, A.D.E. 2018. Section 9510 D. Virus concentration by aluminum hydroxide. Ministry of Health, the Instituto de Salud Carlos III (Ref COV20/00140), ERC TB-RECONNECT 101001038 funded by The European Union – NextGenerationEU and the Lagoon project (PROMETEO/2021/044). EC-F is recipient of a predoctoral contract from the MICINN, Call 2018 and AP-C was supported by a postdoctoral fellowship (APOST/21/292). We acknowledge NILSA, FACSIA, and MITECO for authorizing the sampling. The authors thank Agustín Garrido Fernández, and Andrea Lopez de Mota for their technical support.

Supplementary materials
Supplementary material associated with this article can be found in the online version, at doi:10.1016/j.watres.2020.115942.

References
AAVV, A.D.E. 2018. Section 9510 D. Virus concentration by aluminum hydroxide. Ministry of Health, the Instituto de Salud Carlos III (Ref COV20/00140), ERC TB-RECONNECT 101001038 funded by The European Union – NextGenerationEU and the Lagoon project (PROMETEO/2021/044). EC-F is recipient of a predoctoral contract from the MICINN, Call 2018 and AP-C was supported by a postdoctoral fellowship (APOST/21/292). We acknowledge NILSA, FACSIA, and MITECO for authorizing the sampling. The authors thank Agustín Garrido Fernández, and Andrea Lopez de Mota for their technical support.

Supplementary materials
Supplementary material associated with this article can be found in the online version, at doi:10.1016/j.watres.2020.115942.

References
AAVV, A.D.E. 2018. Section 9510 D. Virus concentration by aluminum hydroxide. Ministry of Health, the Instituto de Salud Carlos III (Ref COV20/00140), ERC TB-RECONNECT 101001038 funded by The European Union – NextGenerationEU and the Lagoon project (PROMETEO/2021/044). EC-F is recipient of a predoctoral contract from the MICINN, Call 2018 and AP-C was supported by a postdoctoral fellowship (APOST/21/292). We acknowledge NILSA, FACSIA, and MITECO for authorizing the sampling. The authors thank Agustín Garrido Fernández, and Andrea Lopez de Mota for their technical support.

Supplementary materials
Supplementary material associated with this article can be found in the online version, at doi:10.1016/j.watres.2020.115942.

References
AAVV, A.D.E. 2018. Section 9510 D. Virus concentration by aluminum hydroxide. Ministry of Health, the Instituto de Salud Carlos III (Ref COV20/00140), ERC TB-RECONNECT 101001038 funded by The European Union – NextGenerationEU and the Lagoon project (PROMETEO/2021/044). EC-F is recipient of a predoctoral contract from the MICINN, Call 2018 and AP-C was supported by a postdoctoral fellowship (APOST/21/292). We acknowledge NILSA, FACSIA, and MITECO for authorizing the sampling. The authors thank Agustín Garrido Fernández, and Andrea Lopez de Mota for their technical support.

Supplementary materials
Supplementary material associated with this article can be found in the online version, at doi:10.1016/j.watres.2020.115942.

References
AAVV, A.D.E. 2018. Section 9510 D. Virus concentration by aluminum hydroxide. Ministry of Health, the Instituto de Salud Carlos III (Ref COV20/00140), ERC TB-RECONNECT 101001038 funded by The European Union – NextGenerationEU and the Lagoon project (PROMETEO/2021/044). EC-F is recipient of a predoctoral contract from the MICINN, Call 2018 and AP-C was supported by a postdoctoral fellowship (APOST/21/292). We acknowledge NILSA, FACSIA, and MITECO for authorizing the sampling. The authors thank Agustín Garrido Fernández, and Andrea Lopez de Mota for their technical support.

Supplementary materials
Supplementary material associated with this article can be found in the online version, at doi:10.1016/j.watres.2020.115942.

References
AAVV, A.D.E. 2018. Section 9510 D. Virus concentration by aluminum hydroxide. Ministry of Health, the Instituto de Salud Carlos III (Ref COV20/00140), ERC TB-RECONNECT 101001038 funded by The European Union – NextGenerationEU and the Lagoon project (PROMETEO/2021/044). EC-F is recipient of a predoctoral contract from the MICINN, Call 2018 and AP-C was supported by a postdoctoral fellowship (APOST/21/292). We acknowledge NILSA, FACSIA, and MITECO for authorizing the sampling. The authors thank Agustín Garrido Fernández, and Andrea Lopez de Mota for their technical support.

Supplementary materials
Supplementary material associated with this article can be found in the online version, at doi:10.1016/j.watres.2020.115942.

References
AAVV, A.D.E. 2018. Section 9510 D. Virus concentration by aluminum hydroxide. Ministry of Health, the Instituto de Salud Carlos III (Ref COV20/00140), ERC TB-RECONNECT 101001038 funded by The European Union – NextGenerationEU and the Lagoon project (PROMETEO/2021/044). EC-F is recipient of a predoctoral contract from the MICINN, Call 2018 and AP-C was supported by a postdoctoral fellowship (APOST/21/292). We acknowledge NILSA, FACSIA, and MITECO for authorizing the sampling. The authors thank Agustín Garrido Fernández, and Andrea Lopez de Mota for their technical support.

Supplementary materials
Supplementary material associated with this article can be found in the online version, at doi:10.1016/j.watres.2020.115942.