Heliyon 10 (2024) e35332

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Wastewater multiplex PCR amplicon sequencing revealed community transmission of SARS-CoV-2 lineages during the outbreak of infection in Chinese Mainland

Langjun Tang [1], Zhenyu Guo [1], Xiaoyi Lu, Junqiao Zhao, Yonghong Li [**], Kun Yang [*]

Department of Pharmaceutical & Biological Engineering, School of Chemical Engineering, Sichuan University, Chengdu, 610065, China

ARTICLE INFO

Keywords:
Multiplex PCR amplicon sequencing
Wastewater-based epidemiology
SARS-CoV-2
Deconvolution
Virus lineages

ABSTRACT

During the COVID-19, wastewater-based epidemiology (WBE) has become a powerful epidemic surveillance tool widely used worldwide. However, the development and application of this technology in Chinese Mainland are relatively lagging. Herein, we for the first time monitored the community circulation of SARS-CoV-2 lineages using WBE methods in Chinese Mainland. During the peak period of infection outbreak at the end of 2022, six precious sewage samples were collected from the manhole in the student dormitory area on Wangjiang Campus of Sichuan University. RT-qPCR revealed that the six sewage samples were all positive for SARS-CoV-2 RNA. Multiplex PCR amplicon sequencing of the sewage samples reflected the local transmission of SARS-CoV-2 variants. The results of two deconvolution methods indicate that the main virus lineages have clear evolutionary genetic correlations. Furthermore, the sampling time is consistent with the timeline of concern for these virus lineages, as well as the timeline of uploading the nucleic acid sequences from the corresponding lineages in Sichuan to the database. These results demonstrate the reliability of the sewage sequencing results. Multiplex PCR amplicon sequencing is by far the most powerful analytical tool of WBE, enabling quantitative detection of virus lineages transmission and evolution at the community level.

Environmental Implication

Wastewater-based epidemiology (WBE) is an important method of environmental surveillance. In this work, we reported the first case of monitoring community circulation of SARS-CoV-2 variant at lineage level using WBE methods in Chinese Mainland. Multiplex PCR amplicon sequencing is proved by far the most powerful WBE tool. Due to its dual potential of both qualitative and quantitative analysis, it is possible to determine at the community level when, where and to what extent virus variants are circulating. WBE and clinical testing complement each other, providing strong support for human response to the next pandemic of infectious disease.

[*] Corresponding author.
[**] Corresponding author.
E-mail addresses: liyonghong@scu.edu.cn (Y. Li), cookyoung@scu.edu.cn (K. Yang).

[1] Langjun Tang and Zhenyu Guo contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2024.e35332
Received 15 September 2023; Received in revised form 18 July 2024; Accepted 26 July 2024
Available online 27 July 2024

1. Introduction

Since the first detection of SARS-CoV-2 nucleic acid in sewage [1,2], the wastewater-based epidemiology (WBE) of SARS-CoV-2 has made great progress around the world [3]. The efforts made in the past two years include but are not limited to the explorations on sampling strategy [4,5], sample pretreatment [6,7], nucleic acid extraction [8], nucleic acid detection, decay kinetics of virus (nucleic acid) in sewage [9], correlation between sewage viral RNA concentration and community infection incidence [10] and community transmission of virus lineages [11]. The continuous breakthroughs in the above technology, methods and means have preliminarily established a theoretical framework for WBE in community monitoring of infectious diseases.

So far, the WBE for SARS-CoV-2 has developed four main application scenarios [12,13]. We summarize them as four "**W**" functions of WBE: to give early warning of the outbreak (**W**hen) [14,15], to track the trend of community infection (to **W**hat extent) [16], to pinpoint the infection hotspots (**W**here) [17,18], and to analyze the circulation and evolution of virus lineages (**W**hich variants) in the community [11]. Combining the four functions of WBE is expected to maximize its potential in epidemic monitoring.

Quantitative tracking the circulation and evolution of virus lineages in communities is currently at the forefront of WBE research, which has fully utilized the ***qualitative*** and ***quantitative*** capabilities of sewage surveillance. At the beginning, RT-PCR or RT-qPCR targeting specific mutations were used to track a few virus variants of interest [19,20]. Subsequently, it was developed to track mutation sites via high-throughput next-generation sequencing [21–23]. At present, the whole genome sequencing of sewage SARS-CoV-2 can comprehensively track the circulation of different virus lineages in the community [11]. There are currently three whole genome sequencing methods for viruses: multiplex PCR amplicon-based sequencing, hybrid capture-based sequencing, as well as ultra-high-throughput *meta*-transcriptomic sequencing [24]. Multiplex PCR amplicon sequencing, due to its highest enrichment efficiency, is very suitable for whole genome detection of trace microorganisms in complex matrices such as sewage samples. Current sequencing library construction methods include ARTIC [25], Swift [11] and ATOPlex [24]. The sequencing platforms involve MinION, Illumina, ONT-Nanopore and BGI-DNBSEQ. Up to now, only ***several*** literatures have reported ***deconvolution*** methods for quantifying the abundance of SARS-CoV-2 lineages in sewage [11,26].

The practice of WBE for SARS-CoV-2 in Chinese Mainland is relatively lagging due to the stringent epidemic control measures [27,28]. There were a few reports about the detection of SARS CoV-2 RNA from hospital wastewater only at the beginning of the epidemic [29–31]. However, in the Hong Kong region, close cooperation has been developed between researchers and the government, and a relatively complete sewage virus monitoring network has been established [32]. The technical methods have also been optimized and are basically becoming mature [6,8]. After the management of the disease was downgraded from Class A to Class B in China (on Jan. 8th' 2023), the first academic report on WBE of SARS-CoV-2 emerged from Chinese Mainland [28]. Herein, we will report the first detection of community circulation of SARS-CoV-2 lineages via WBE in Chinese Mainland.

The study route of this work is illustrated in Supplementary Fig. S1. Along with the comprehensive relaxation of epidemic prevention policies by the end of 2022, Chinese Mainland entered a peak period of infection outbreaks. During that period, sewage samples were collected in the student dormitory area of Wangjiang Campus of Sichuan University (Section 2.1). RT-qPCR was applied to detect the presence of SARS-CoV-2 nucleic acid in these sewage samples (Section 2.3). Multiplex PCR amplicon sequencing of the SARS-CoV-2 was performed on the positive samples (Section 2.4 and 2.5). Based on the deconvolution results of sequencing, the local circulation and evolution of SARS-CoV-2 variants was analyzed (Section 2.6). An alternative deconvolution method was proposed to quantify the abundance of SARS-CoV-2 lineages in sewage, and the result was compared with that of Freyja. To verify the reliability of WBE analysis, we specially tracked the database information about the circulation and evolution of SARS-CoV-2 lineages in Chinese Mainland at that time (Section 3.3). The novelty of this work should be that it is the first case of tracking the community circulation of SARS-CoV-2 variants *at lineage level* in Chinese Mainland, proving that multiplex PCR amplicon sequencing of sewage samples is up to now the most effective measure to quantitatively track the community transmission of viral lineages regardless of the sequencing platform used, which expands the application fields of WBE.

2. Materials and methods

2.1. Wastewater sampling

From the end of 2022 to the beginning of 2023, Chinese Mainland experienced a peak of SARS-CoV-2 (Omicron) infection. In less than three months, more than 80 % of the population was infected [33]. Starting from December 27th, untreated wastewater samples were collected every two days from sewage manholes in the dormitory area of Wangjiang campus of Sichuan University, Chengdu, China. Until January 6th, when the school had an early winter vacation, sampling was forced to be terminated due to closed campus management. The autosampler was programmed to collect 100 mL wastewater every 1 h over a 24-h period that were composited to provide 24-h composite samples. The sampling instrument came with a refrigeration system, and the sample temperature was maintained at 4 °C during the sampling process. Sewage samples were transported on ice to the laboratory immediately after collection and stored frozen at −40 °C for future detection. A total of 6 sewage samples were obtained during that period.

2.2. Virus concentration

To recover virus and viral RNA from sewage samples, a modified PEG precipitation method was applied [6]. Typically, the sewage samples were centrifuged at 1680×*g* for 5 min to remove large particles. After transferring the supernatant in a new 50-mL centrifuge tube, 3.50 g of PEG 8000 (to a final concentration of 100 g/L) and 0.79 g of NaCl (to a final concentration of 22.5 g/L) were dissolved in

35 mL of supernatant via gentle rotation. Thereafter, virus was precipitated statically overnight at 4 °C. The precipitated virus and viral RNA was recovered as a pellet by centrifugation at 14,000 g for 1 h at 4 °C.

2.3. Nucleic acid extraction and qPCR detection

Pellets deposited via centrifugation were extracted nucleic acid using MagicPure® Viral DNA/RNA Kit (TransGen Biotech Co., Ltd, Beijing, China) according to the manufacturer's instructions. To determine the relative abundance of SARS-CoV-2 in sewage, crAssphage was co-quantified as an endogenous reference biomarker [34]. Quantitative polymerase chain reaction (qPCR) was applied to quantify crAssphage using the PerfectStart® Green qPCR SuperMix (TransGen Biotech Co., Ltd, Beijing, China). Reverse transcription qPCR (RT-qPCR) was used to quantify SARS-CoV-2 with the TransScript® II Probe One-Step qRT-PCR SuperMix (TransGen Biotech Co., Ltd, Beijing, China). The primer pairs and probes targeting the genes of SARS-CoV-2 and the crAssphage are listed in Table 1. The protocols for qPCR and RT-qPCR are attached in the supplementary information (Supplementary documents 1 and 2, respectively).

The relative abundance of SARS-CoV-2 in sewage was calculated with the following formula (Eq. (1)) [36].

$$R_a = \frac{c_{0,t}}{c_{0,s}} = \frac{{\xi_s}^{C_{T,s}}}{{\xi_t}^{C_{T,t}}} \approx 2^{C_{T,s} - C_{T,t}} \tag{1}$$

Where ξ is the apparent amplification coefficient, which is equal to 2 under ideal conditions; C_T is the threshold cycle number; the subscript *s* indicates the gene of the reference biomarker crAssphage; and the subscript *t* refers to the target virus genes (ORF1ab or N).

2.4. ATOPlex multiplex PCR amplicon sequencing

To detect the whole genome of SARS-CoV-2 in sewage, we applied ATOPlex V3.1 multiplex PCR amplification sequencing. ATOPlex V3.1 adopts a double insurance design with dual primer coverage. Artificially synthesized DNA is spiked in during the multiplex PCR amplification as quality control (external control) for library construction and assisting quantification. Furthermore, the human glyceraldehyde-3-phosphate dehydrogenase gene (GAPDH) is also detected as quality control for sample processing and nucleic acid extraction, and as an internal reference for relative quantification of viral nucleic acid in each sample (Fig. 1).

The RNA multiplex PCR amplicon libraries were constructed by ATOPlex SARS-CoV-2 Full Length Genome Panel following the manufacture's protocol (MGI, Shenzhen, China). Briefly, sewage RNA extract (10 μL) was converted to the first-strand cDNA by reverse transcriptase with random hexamers (5′–NNNNNN–3′). The reverse transcription was performed in a Veriti 96 Well Thermocycler (Applied Biosystems, USA) using the following procedure: 5 min at 25 °C, 20 min at 42 °C, 5 min at 95 °C. The reverse transcript of each sample was mixed with a fixed copy number of artificially synthesized lambda genomic DNA (external control). Thereafter, the mixture was bisected into two equal parts (10 μL each) and two parallel PCR amplifications were performed using two independent primer pools (PCR Primer Pool 1 V3.1 and PCR Primer Pool 2 V3.1), respectively. The PCR primers in two primer pools include (1) primers dual-covering the whole genome of SARS-CoV-2, (2) primers targeting the spiked synthesized lambda DNA (external control) and (3) primers targeting the human GAPDH gene (internal control). The PCR amplifications were performed in a Veriti 96 Well Thermocycler (Applied Biosystems, USA) using the following procedure: 5 min at 37 °C, 5 min at 95 °C, 35 amplification cycles (consisting of 10 s at 95 °C, 1 min at 62 °C, 1 min at 58 °C and then 20 s at 72 °C), followed by a final elongation step at 72 °C for 1 min. The amplification products from two primer pools were mixed and purified with MGIEasy DNA Clean Beads. The amplification products were fragmentized, end-polished, A-tailed, and ligated with the adaptors.

PCR products of samples were pooled at equimolar and converted to single-stranded circular DNA (ssDNA) with the MGIEasy Circularization Kit (MGI, China). These circularized DNA went through rolling circle amplification to form billions of DNA nanoballs (DNBs). DNBs-based libraries were sequenced on DNBSEQ-E5 platform with paired-end 100 nt strategy.

2.5. Sequencing data analysis

After ATOPlex multiplex PCR amplicon sequencing, MGI MetargetCOVID V1.6 software is used for data analysis. The internal process was that low quality reads were filtered and then the clean reads were mapped against the SARS-CoV-2 reference genome (WuHan-Hu-1, NC_045512.2). The reads of mapped files contained primer will be trimmed by using an in-housed script. Then the SNPs

Table 1
Primer/probe sets targeting the genes of SARS-CoV-2 and crAssphage.

Organism	Target	Name	Primers and Probes	Product Size (bp)	T_m (°C)	Reference
SARS-CoV-2	ORF1ab	ORF1ab-F	5′-CCCTGTGGGTTTTACACTTAA-3′	119	58	[35]
		ORF1ab-P	FAM−CCGTCTGCGGTATGTGGAAAGGTTATGG−BHQ1			
		ORF1ab-R	5′-ACGATTGTGCATCAGCTGA-3′			
	N	N–F	5′-GGGGAACTTCTCCTGCTAGAAT-3′	99	58	
		N–P	FAM−TTGCTGCTGCTTGACAGATT− BHQ1			
		N–R	5′-CAGACATTTTGCTCTCAAGCTG-3′			
crAssphage	CPQ_056	056F1	5′-CAGAAGTACAAACTCCTAAAAAACGTAGAG-3′	126	62	[34]
		056R1	5′-GATGACCAATAAACAAGCCATTAGC-3′			

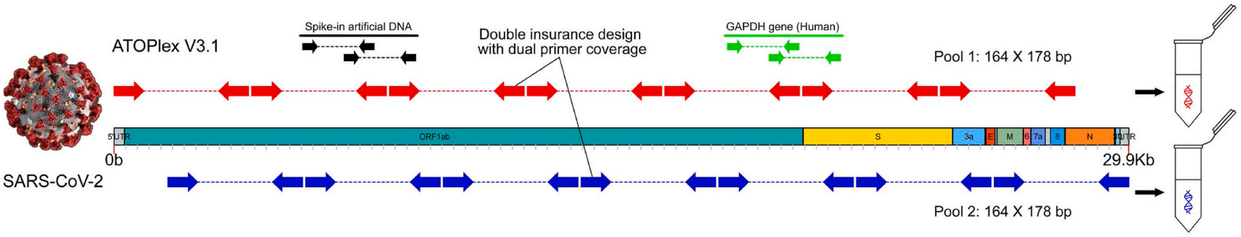

Fig. 1. Mechanism of ATOPlex sequencing. ATOPlex is a multiplex PCR amplicon-based high-throughput sequencing platform, which can detect the whole genome of trace microorganisms (especially viruses with smaller genomes) in complex samples. ATOPlex V3.1 adopts a double insurance design with dual primer coverage. Artificially synthesized DNA is spiked in during the multiplex PCR amplification as quality control for library construction and assisting quantification. Furthermore, the human glyceraldehyde-3-phosphate dehydrogenase gene (GAPDH) is also detected as quality control for sample processing and nucleic acid extraction, and also as an internal reference for relative quantification of viral nucleic acid in each sample.

(single-nucleotide polymorphisms), indels (insertions and deletions), MNPs (multi-nucleotide polymorphisms), and complex events (composite insertion and substitution events) were called based on the BAM file. After that, a consensus genome was obtained in term of the variant calling VCF file and the SARS-CoV-2 reference genome. Finally, the clades were provided through mutation identification, consensus genome quality checking, and phylogenetic placement and visualization to assign the lineages.

The sequencing data gave the reads of spike-in external control (artificially synthesized lambda genomic DNA), internal reference (GAPDH) and SARS-CoV-2. As the copy number of spike-in external control was known, the concentration of SARS-CoV-2 nucleic acid in each RNA extracts was calculated accordingly with the following formula (Eq. (2)).

$$C_{SARS-CoV-2} = 521.87 \times \left(\frac{SARS - CoV - 2\ reads}{Spike - in\ reads}\right)^{1.2301} (copies\,/\,mL) \tag{2}$$

Parameters 521.87 and 1.2301 were provided by the sequencing company derived from previous systematic experimental study. The relative abundance of SARS-CoV-2 nucleic acid against the internal reference (GAPDH) was calculated with the following equation (Eq. (3)).

$$R_{a_{,SARS-CoV-2/GAPDH}} = \frac{SARS - CoV - 2\ reads}{GAPDH\ reads} \tag{3}$$

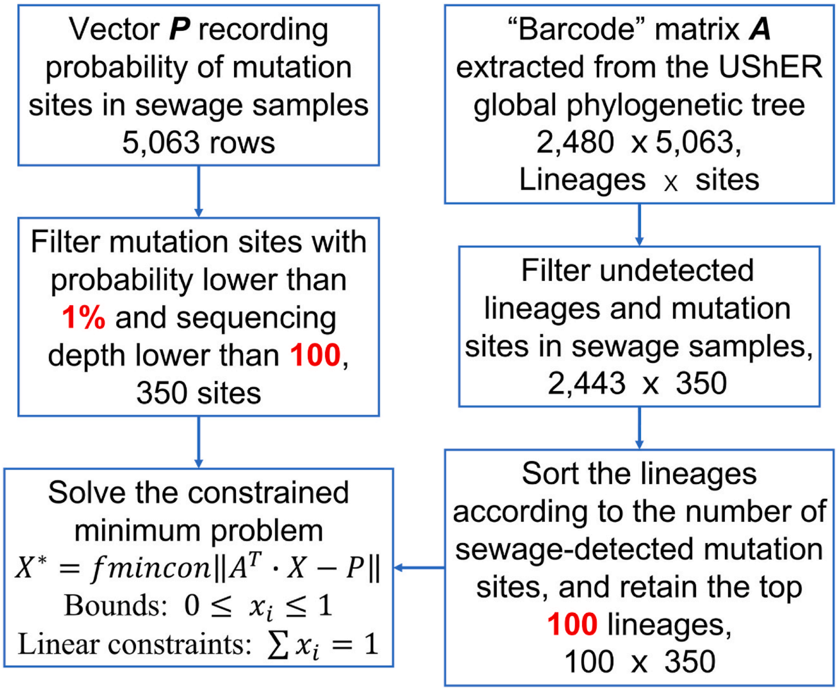
Vector P recording probability of mutation sites in sewage samples 5,063 rows
"Barcode" matrix A extracted from the UShER global phylogenetic tree 2,480 x 5,063, Lineages x sites
Filter mutation sites with probability lower than 1% and sequencing depth lower than 100, 350 sites
Filter undetected lineages and mutation sites in sewage samples, 2,443 x 350
Solve the constrained minimum problem $X^* = fmincon\|A^T \cdot X - P\|$ Bounds: $0 \le x_i \le 1$ Linear constraints: $\sum x_i = 1$
Sort the lineages according to the number of sewage-detected mutation sites, and retain the top 100 lineages, 100 x 350

Fig. 2. Diagram illustrating the deconvolution process of sewage ATOPlex sequencing data in this work.

2.6. Deconvolution of virus variants for sewage samples

The deconvolution of sequencing data provided the relative abundance of SARS-CoV-2 variants in sewage samples. To ensure the reliability of the results, two independent deconvolution methods were applied. The sequencing company MGI used Freyja for the deconvolution, which is based on a depth-weighted de-mixing method [11]. The relative abundance of SARS-CoV-2 variants/lineages in sewage samples was provided as an Excel spreadsheet. The second method was developed in in this work, which was amended from the first one. The deconvolution process was illustrated in Fig. 2. The mutation sites with probability more than 1 % were recorded when mapping the sequencing reads against the reference sequence, and those sites with sequencing depth low than 100 were filtered. The filtered mutation probability of each site was recorded in a vector $\boldsymbol{P}$ (Eq. (4)).

$$P = \begin{bmatrix} p_1 \\ \vdots \\ p_j \\ \vdots \\ p_n \end{bmatrix}, 0 \leq p_j \leq 1 \tag{4}$$

where $\boldsymbol{p_j}$ represents the overall mutation probability of site $\boldsymbol{j}$, which is a linear combination of mutation of each virus lineage at the mutation site. The “Barcode” matrix $\boldsymbol{A}$ containing lineage-defining mutations for known virus lineages was still in use (Eq. (5)).

$$A = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & a_{i,j} & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{bmatrix}, 0 \leq a_{i,j} \leq 1 \tag{5}$$

The matrix was obtained from the UShER global phylogenetic tree using the matUtils package [37], where the elements $\boldsymbol{a_{i,j}}$ denotes the mutation of virus lineage $\boldsymbol{i}$ at site $\boldsymbol{j}$. Since each virus lineage is defined by a combination of specific mutation sites, each row of the matrix represents one lineage (out of more than 2400 lineages included in the UShER global phylogenetic tree), and the distribution of individual mutations among different lineages is represented as columns. This matrix was also filtered to exclude those mutation sites (columns) and lineages (rows) undetected in all sewage samples. The relative abundance of virus lineages in sewage is specified with a vector $\boldsymbol{X}$ (Eq. (6)).

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_m \end{bmatrix}, x_i \geq 0 \ \& \ \sum_{i=1}^{m} x_i = 1 \tag{6}$$

Where $\boldsymbol{x_i}$ represents the relative abundance of virus lineage $\boldsymbol{i}$. The relative abundance vector $\boldsymbol{X}$ for each sewage sample can be obtained by solving the following constrained minimum problem (Eq. (7)).

$$X^* = fmincon\|A^T \bullet X - P\|$$

$$\text{Bounds} : 0 \leq x_i \leq 1$$

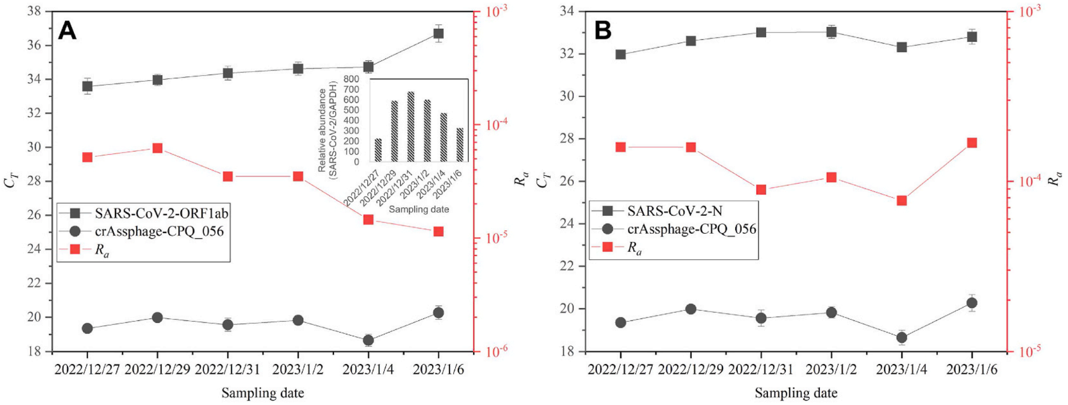

Fig. 3. Relative abundance of SARS-CoV-2 genes in sewage samples. The qPCR CT values of SARS-CoV-2 genes and the endogenous biomarker crAssphage are exhibited. The calculated relative abundances ($\boldsymbol{R_a}$) of ORF1ab and N gene are shown in panel A and B, respectively. The inset in panel A gives the relative abundance of SARS-CoV-2 nucleic acids calculated based on ATOPlex sequencing results. The internal reference for relative quantification is the human glyceraldehyde-3-phosphate dehydrogenase gene (GAPDH).

$$\text{Linear constraints}: \sum x_i = 1 \tag{7}$$

The problem was solved using the built-in function (*fmincon*) in MATLAB R2017a (MathWorks, United States). Please refer to the Supplementary MATLAB script "V_POWER.m" and the attached dataset "Mutation.xlsx".

2.7. Data availability

The sequencing data (clean FastQ reads) reported in this paper have been deposited in the Genome Sequence Archive [38] in National Genomics Data Center [39], China National Center for Bioinformation (CNCB)/Beijing Institute of Genomics (BIG), Chinese Academy of Sciences (GSA: CRA011084 or BioProject: PRJCA017073) that are publicly accessible at https://ngdc.cncb.ac.cn/gsa.

3. Results

3.1. Relative abundance of SARS-CoV-2 nucleic acid in sewage samples

When quantifying SARS-CoV-2 nucleic acid in the sewage samples, the abundance of the endogenous reference biomarker crAssphage was co-quantified. The relative abundance of SARS-CoV-2 nucleic acid was calculated to be around 10^{-5}-10^{-4} GC/crAssphage GC (Fig. 3.). The concentration of crAssphage in sewage was reported to be about 10^9 GC/L [40]. Thus, the concentration of SASR-CoV-2 RNA is about 10^4–10^5 GC/L. There seems to be no close linear correlation between the abundance of the two target genes (ORF1ab and N genes) of SARS-CoV-2 in sewage. The relative abundance of SARS-CoV-2 against the human GAPDH gene was calculated according to the ATOPlex sequencing reads via Eq. (2). The calculated relative abundance of SARS-CoV-2 nucleic acid ($\boldsymbol{R_{a,\ SARS\text{-}CoV\text{-}2/GAPDH}}$) in the sewage experienced a process of increasing first and then decreasing (insect of Fig. 3A).

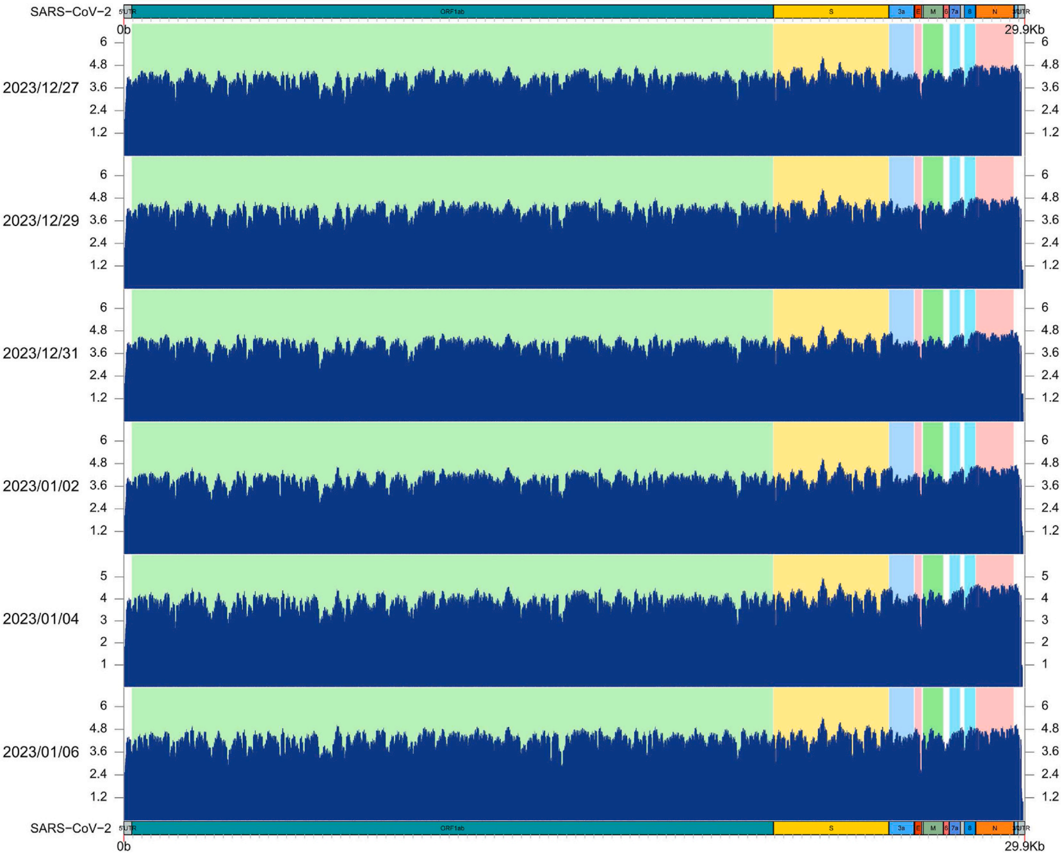

Fig. 4. SARS-CoV-2 genome coverage and sequencing depth of ATOPlex multiplex PCR amplicon sequencing for six sewage samples. The horizontal coordinate is the nucleic acid sequence position of the viral genome, and the vertical coordinate is the logarithm value of the sequencing depth ($\log_{10}$). The genome coverage is good and the sequencing depth is high. Except for 3′-UTR and 5′-UTR, the average sequencing depth of other regions is more than 1000 x.

3.2. Multiplex PCR amplicon sequencing reflects variant prevalence in sewage

ATOPlex multiplex PCR amplicon sequencing gave good genome coverage and high sequencing depth for all six sewage samples (Fig. 4). Most of the sequencing reads (≥90 %) were from SARS-CoV-2 genome (Table 2). Except for 3′-UTR and 5′-UTR, the average sequencing depth of most other genome regions is more than 1000 x (Supplementary dataset 1). The 100 x coverages for six sewage samples are overall greater than 98.5 % (Table 2). Except for two samples (2023/1/4 and 2023/1/6) with amplicon drop off in the E gene (the last two panels in Fig. 4), the 100 x coverage of the SARS-CoV-2 genome coding region exceeded 99 % (Supplementary dataset 1). We also found that the sequencing coverage/depth plots of different sewage sample exhibited similar "waveform" (Fig. 4). This reflects the ability of the two primer pools to cover the entire genome sequence of SARS-CoV-2 and the bias of PCR amplification. Overall, the coverage of the viral genome by the multiplex PCR is good.

Sample deconvolution gives the relative abundance of SARS-CoV-2 lineages in sewage samples. Two deconvolution methods give similar results (Fig. 5A and B). The main virus variants (prevalence rate>1 %) in sewage samples exhibit clear evolutionary genetic correlations (Fig. 5C). They all evolved from the BA. 5.2 lineage through a series of mutations. During the process of evolution, three main sublineages were formed, namely BA.5.2.49, BA.5.2.48, and BF.7.14. Further evolution of each sublineage has led to the emergence of new variants. The results obtained by the two deconvolution methods differ at the haplotype level but are highly consistent at the sublineage level (Fig. 5A and B).

3.3. Wastewater reveals local community transmission of SARS-CoV-2 variants

According to the sequencing data, the probability and sequencing depth of main amino acid (AA) mutations across all six sewage samples were recorded and shown as heatmaps in Fig. 6. To prove the reliability of the sewage sequencing data, we retraced the pango-designation issues of these virus variants/lineages detected in these sewage samples on the website https://cov-lineages.org/lineage_list.html and recorded their information in Supplementary Table S2. The AA mutation sites of these virus variants/lineages are exhibited in Supplementary Fig. S2. Only the mutation frequencies of mutation sites in ORF1ad and S gene with relatively concentrated mutation are shown. By comparing Fig. 6A and Fig. S2, it was found that these main AA mutation sites were also detected in these sewage samples via the sequencing data (also refer to supplementary dataset 2 for detailed information). The main transmission sites ("Most common countries") of these SARS-CoV-2 variants/lineages are all in China. Issue-open dates of these lineages were concentrated on December 26 and 27, which were around our sampling date. We also found that 26 sequences of BA.5.2.49 were ever uploaded from Sichuan on Dec 29th, 2022 and two more sequences were uploaded from Sichuan on Jan 8th, 2023 (https://github.com/cov-lineages/pango-designation/issues/1480). The branch of lineage BF.7.14.7 contains 11 out of 34 (33 %) Sichuan sequences as of Jan 9th, 2023 (https://github.com/cov-lineages/pango-designation/issues/1672). These facts prove that the sewage viral sequencing data reveals local community transmission of SARS-CoV-2 variants.

4. Discussion

During the multiplex PCR amplification, ***fixed copy number*** of artificially synthesized lambda genomic DNA was spiked in the reaction mixture as ***external control***. The endogenous biomarker human GAPDH gene was also detected as ***internal reference***. The spike-in external control can realize the absolute quantification of SARS-CoV-2 nucleic acid in the sewage RNA extracts. However, RNA extraction may be affected by the physicochemical property of the sewage. The concentration of viral nucleic acid in RNA extracts may not reflect the true contents of the virus in sewage samples. The relative abundance of the viral nucleic acid against the endogenous

Table 2
Basic information of ATOPlex sequencing results.

Sample ID	GAPDH reads	Spike-in control reads	SARS-CoV-2 reads	SARS-CoV-2 (copies/mL)	Relative abundance SARS-CoV-2/GAPDH	100 x Coverage (%)	Average depth	Total mutation	Lineage ID
2022/12/27	3514	72585	791966	9868.1	225.37	99.17 %	2124.08	77	BA.5.2
2022/12/29	1467	95434	867871	7887.21	591.60	99.42 %	2303.02	80	BA.5.2
2022/12/31	1029	54149	697512	12103.91	677.85	99.25 %	1888.88	80	BA.5.2
2023/1/2	940	44298	563937	11929.82	599.93	98.81 %	1510.35	80	BA.5.2
2023/1/4	1032	56414	486036	7379.93	470.97	98.60 %	1288.6	77	BA.5.2
2023/1/6	3706	108615	1204748	10069.89	325.08	99.09 %	3166.26	77	BA.5.2

GAPDH: The internal reference gene encoding human glyceraldehyde-3-phosphate dehydrogenase.

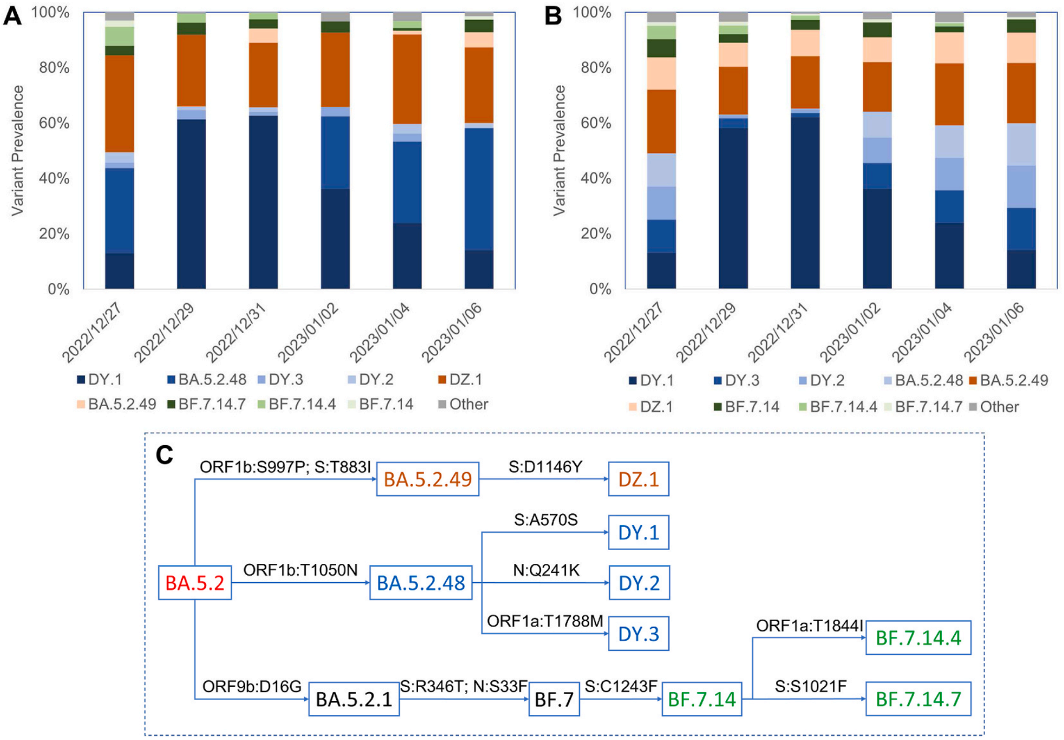

Fig. 5. The prevalence (%) of virus variants/lineages in sewage samples. The results obained with Freyja and our method were ehibited in panels **A** and **B**, respectively. The main virus variants (Prevalence%>1 %) in sewage samples exhibit clear evolutionary genetic correlations. They all evolved from the BA. 5.2 lineage through a series of mutations. During the process of evolution, three main sublineages were formed, namely BA.5.2.49, BA.5.2.48, and BF.7.14 (Panel **C**). The virus variants of the same sublineages are indicated with same colors of different saturations in panels A and B. Blue represents the sublineage BA.5.2.48, brown represents BA.5.2.49, and green represents BF.7.14. In addition, the sampling time is very consistent with the timeline of uploading the nucleic acid sequences of the corresponding virus variants/lineages from Sichuan to the database (GISAID, https://gisaid.org/) (Supplementary Table S2). This indicates the reliability of the sewage sequencing results.

biomarker can reflect the true viral load in wastewater. At the beginning of its development, ATOPlex was mainly used for the detection of clinical samples, so human GAPDH gene was adopted as the internal reference [24]. For sewage samples, commonly used endogenous biomarkers such as human gut specific bacteriophage (crAssphage) [41] and pepper mild mottle virus (PMMoV) [42] may be more suitable as internal references. The RT-qPCR determined abundances of different genes (ORF1ab and N) are not consistent with each other (Fig. 3A and B), and also differ from the abundance determined by ATOPlex sequencing (inset of Fig. 3A). It has been suggested that RT-qPCR may give false negative detection for SARS-CoV-2 [43], which can be attribute to degradation of SARS-CoV-2 RNA in the wastewater matrix [9]. In addition, the accuracy of RT-qPCR quantification is also affected by mutations in the viral genome sequence targeted by primers/probes [44,45], especially for samples such as sewage that may contain multiple virus variants. Therefore, a quantification method that targets the whole genome of SARS-CoV-2 is valuable for WBE [46]. The ATOPlex sequencing results indicate that the relative abundance of SARS-CoV-2 RNA in wastewater increases first and then decreases over time (inset of Fig. 3A), which seems reasonable.

The design of the primer set is crucial for improving the sequencing coverage and depth of multiplex PCR amplicon sequencing [47]. The amplicon droop-offs due to primer dimers and primer mismatch caused by virus genome mutations are problems that needs to be addressed during primer set design and optimization [47]. To improve the coverage of sequencing, the primer set is designed to tile amplicons across entire sequence of the published reference SARS-CoV-2 genome NC_045512.2 [48]. The primer set is divided into two separate subsets (Pools 1 and 2), and primer pairs targeting adjacent regions are separated in alternate pools so that overlap of PCR fragments occurs between pools but not within [25,49]. The ATOPlex SARS-CoV-2 full-length genome panel has evolved to the third version. Its previous versions have shown considerable application potential in the detection of input virus variants [50], quantification of sewage SARS-CoV-2 RNA [46], and resolution of sewage virus lineages [51,52]. ATOPlex V3.1 also adopts a double insurance design with dual primer coverage (Fig. 1). Quantification of the abundance of virus lineages in sewage places higher demands on sequencing coverage and depth. To accurately analyze virus lineages with relative prevalence exceeding 1 %, the sequencing depth should reach at least 1000 x. The sequencing coverage (100 x coverages >98.5 %) and depth (more than 1000 x on average) of each sewage sample in this study were both of the highest levels to date with little coverage bias (Supplementary dataset 1), basically meeting the above requirements. Furthermore, this work is one of the few studies that quantified virus lineage abundance in sewage by deconvolution [11,26]. We also provide a deconvolution candidate scheme. The DNB-based sequencing method, which obtained the high-fidelity sequencing library through rolling circle amplification, ensured the accuracy of sequencing. Considering that the

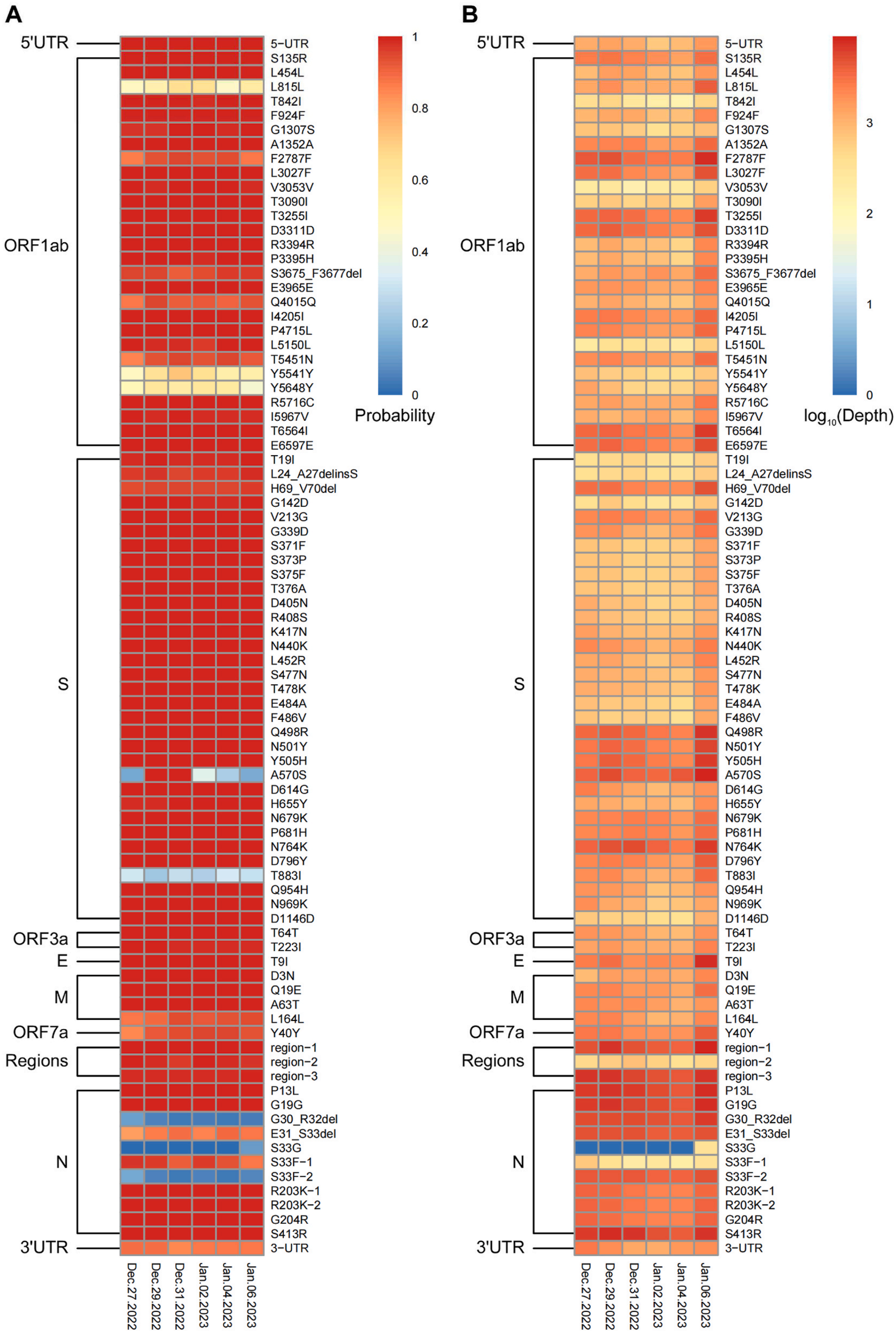

Fig. 6. Heatmap showing the probability (A) and sequencing depth (B) of main AA mutations (n = 85) across all six sewage samples. Please refer to Supplementary dataset 2 for detailed information. The mutation S33G of N gene was only detected in the sewage sample on Jan. 6th, 2023.

coverage and depth of sequencing are mainly determined by the design of primer sets, during the deconvolution process, only mutation points with sequencing depths below 100 were filtered, without using the depth-weighted method [11]. Nevertheless, the two deconvolution methods gave highly consistent results, especially at the level of sublineage, where the prevalence values of three main sublineages given by two methods (i.e., BA.5.2.49, BA.5.2.48, and BF.7.14, Fig. 5C) were quite close (comparing the results in Fig. 5A and B).

Six sewage samples contain three major SARS-CoV-2 variants, while BA.5.2.48 is absent from the samples of 2022/12/29 and 2022/12/31. We obtained sewage samples from a manhole in the student dormitory area. The community monitored by the sewage manhole is relatively small, with only several student dormitory buildings. Therefore, the secretions of infected individuals contained in these sewage samples have a significant randomness. There may be a situation where the sewage sample collected on a certain day lacks the secretion of an infected person. However, even so, only six sewage samples may already contain important information about the main prevalent SARS-CoV-2 lineages in the entire region at that time.

Multiplex PCR amplicon sequencing has both quantitative and qualitative potential and is currently the most powerful WBE tool. Overall, our study is still a retrospective one. The premise for defining a SARS-CoV-2 lineage is that the virus variant should have the potential of ongoing transmission, which means it has achieved a certain scale of community transmission when being named [53]. When new SARS-CoV-2 variants is detected in sewage, the corresponding lineages may have not yet been named. As in the case of this study, the sampling time for sewage spans from Dec. 27, 2022 to Jan. 6, 2023, and the formal name of main SARS-CoV-2 lineages detected in sewage had not yet been designated during that period. Multiple PCR amplicon sequencing should enable the real-time tracking of the transmission of known virus lineages imported into a new community. In addition, there is still room for optimization in the deconvolution method of sewage samples. Different sequencing methods may require targeted deconvolution methods. However, it is worth noting that any deconvolution method cannot distinguish recombination events between virus variants. Therefore, in summary, clinical testing and sewage testing are complementary in community monitoring of infectious diseases. WBE cannot completely replace clinical testing, but in terms of large-scale monitoring, WBE has shown unique advantages over clinical testing. In the long run, population growth and global warming are likely to exacerbate human–pathogen interactions, combined with the continuous evolution of pathogens, suggesting an urgent need for continuous technological advancements to keep up with these changes and provide more effective surveillance. When the next new threat arises, a universal early-warning system that includes sewage pathogen variant monitoring is crucial for protecting the public and relieving the pressure on healthcare system [54].

5. Conclusion

During the outbreak of infection in Chinese Mainland at the end of 2022, six sewage samples collected from the student dormitory area on Wangjiang Campus of Sichuan University were positive for SARS-CoV-2 RNA via RT-qPCR detection. The abundance of the viral nucleic acid in these sewage samples is approximately 10^4–10^5 GC/L. All six virus RNA positive sewage samples were subjected to ATOPlex multiplex PCR amplicon sequencing. The sequencing data were deconvoluted with two alternative methods. Both deconvolution methods give similar results, depicting the circulation and evolution of SARS-CoV-2 variants in the community, which exhibits high consistence with the timeline for locally uploading gene sequences of relevant virus variants to the database. This work proved that multiplex PCR amplicon sequencing is up to now the most powerful WBE tool, which can quantitatively reveal the community transmission of virus variants regardless of the sequencing platform used.

Data availability statement

Data will be made available on request. The sequencing data (clean FastQ reads) reported in this paper have been deposited in the Genome Sequence Archive [38] in National Genomics Data Center [39], China National Center for Bioinformation (CNCB)/Beijing Institute of Genomics (BIG), Chinese Academy of Sciences (GSA: CRA011084 or BioProject: PRJCA017073) that are publicly accessible at https://ngdc.cncb.ac.cn/gsa. The sequences of the primer pool (V2.0) for ATOPlex sequencing are available on online [24] (https://github.com/MGI-tech-bioinformatics/SARS-CoV-2_Multi-PCR_v1.0/blob/master/database/nCoV.primer.2.0.xls).

CRediT authorship contribution statement

Langjun Tang: Writing – review & editing, Visualization, Methodology, Investigation, Data curation. **Zhenyu Guo:** Writing – review & editing, Visualization, Data curation. **Xiaoyi Lu:** Writing – review & editing. **Junqiao Zhao:** Writing – review & editing. **Yonghong Li:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Kun Yang:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

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

We thank the financial supports from the National Natural Science Foundation of China (No. 22176133, No. 21677104). We would like to express our gratitude to Mr. Bo Zhang at MGI for his assistance in sequencing data processing.

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e35332.

Nomenclature

A Barcode matrix containing lineage-defining mutations for known SARS-CoV-2 lineages
$a_{i,j}$ the mutation of virus lineage $\boldsymbol{i}$ at site $\boldsymbol{j}$
R_a relative abundance of SARS-CoV-2 gene in sewage (against crAssphage CPQ_056)
$R_{a,SARS\text{-}CoV\text{-}2/GAPDH}$ relative abundance of SARS-CoV-2 genome in sewage (against GAPDH)
$C_{SARS\text{-}CoV\text{-}2}$ concentration of SARS-CoV-2 genome in sewage nucleic acid extracts
C_T cycle threshold
c_0 initial copy number of the target gene during qPCR amplification (GC/reaction)
P vectors recording the probability of each mutation site
p_j the overall probability of mutation site $\boldsymbol{j}$
t hydraulic retention time (HRT) of virus/biomarker in sewage (h)
$t_{1/2}$ half-life of virus in the sewage (h)
X the vector indicating relative abundance of virus lineages
X^* the optimized relative abundance vector
x_i the relative abundance of virus lineage $\boldsymbol{i}$

Greek letters
ξ apparent amplification coefficient, equaling 2 under ideal condition

Subscripts
i virus lineage $\boldsymbol{i}$
j mutation site $\boldsymbol{j}$
s parameters in related to the reference biomarker
t parameters in related to the target viral gene

References

[1] W. Lodder, A.M. de Roda Husman, SARS-CoV-2 in wastewater: potential health risk, but also data source, Lancet Gastroenterol Hepatol 5 (6) (2020) 533–534.
[2] W. Ahmed, et al., First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: a proof of concept for the wastewater surveillance of COVID-19 in the community, Sci. Total Environ. 728 (2020) 138764.
[3] S. Ciannella, C. Gonzalez-Fernandez, J. Gomez-Pastora, Recent progress on wastewater-based epidemiology for COVID-19 surveillance: a systematic review of analytical procedures and epidemiological modeling, Sci. Total Environ. 878 (2023) 162953.
[4] L. Haak, et al., Spatial and temporal variability and data bias in wastewater surveillance of SARS-CoV-2 in a sewer system, Sci. Total Environ. 805 (2022) 150390.
[5] C. Schang, et al., Passive sampling of SARS-CoV-2 for wastewater surveillance, Environ. Sci. Technol. 55 (15) (2021) 10432–10441.
[6] X. Zheng, et al., A rapid, high-throughput, and sensitive PEG-precipitation method for SARS-CoV-2 wastewater surveillance, Water Res. 230 (2023) 119560.
[7] M.A.I. Juel, et al., Performance evaluation of virus concentration methods for implementing SARS-CoV-2 wastewater based epidemiology emphasizing quick data turnaround, Sci. Total Environ. 801 (2021) 149656.
[8] X. Zheng, et al., Comparison of virus concentration methods and RNA extraction methods for SARS-CoV-2 wastewater surveillance, Sci. Total Environ. 824 (2022) 153687.
[9] W. Ahmed, et al., Decay of SARS-CoV-2 and surrogate murine hepatitis virus RNA in untreated wastewater to inform application in wastewater-based epidemiology, Environ. Res. 191 (2020) 110092.
[10] X. Li, et al., Correlation between SARS-CoV-2 RNA concentration in wastewater and COVID-19 cases in community: a systematic review and meta-analysis, J. Hazard Mater. 441 (2023) 129848.
[11] S. Karthikeyan, et al., Wastewater sequencing reveals early cryptic SARS-CoV-2 variant transmission, Nature 609 (7925) (2022) 101–108.
[12] F. Wu, et al., Making waves: wastewater surveillance of SARS-CoV-2 in an endemic future, Water Res. 219 (2022) 118535.
[13] X. Zheng, et al., Quantification of SARS-CoV-2 RNA in wastewater treatment plants mirrors the pandemic trend in Hong Kong, Sci. Total Environ. (2022) 157121.
[14] W. Randazzo, et al., SARS-CoV-2 RNA in wastewater anticipated COVID-19 occurrence in a low prevalence area, Water Res. 181 (2020) 115942.
[15] G. Medema, et al., Presence of SARS-coronavirus-2 RNA in sewage and correlation with reported COVID-19 prevalence in the early stage of the epidemic in The Netherlands, Environ. Sci. Technol. Lett. 7 (7) (2020) 511–516.
[16] J. Peccia, et al., Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics, Nat. Biotechnol. 38 (10) (2020) 1164–1167.
[17] Y. Deng, et al., Use of sewage surveillance for COVID-19 to guide public health response: a case study in Hong Kong, Sci. Total Environ. 821 (2022) 153250.
[18] W.Q. Betancourt, et al., COVID-19 containment on a college campus via wastewater-based epidemiology, targeted clinical testing and an intervention, Sci. Total Environ. 779 (2021) 146408.

[19] K. Yaniv, et al., RT-qPCR assays for SARS-CoV-2 variants of concern in wastewater reveals compromised vaccination-induced immunity, Water Res. 207 (2021) 117808.
[20] T.E. Graber, et al., Near real-time determination of B.1.1.7 in proportion to total SARS-CoV-2 viral load in wastewater using an allele-specific primer extension PCR strategy, Water Res. 205 (2021) 117681.
[21] M. Avgeris, et al., Novel nested-seq approach for SARS-CoV-2 real-time epidemiology and in-depth mutational profiling in wastewater, Int. J. Mol. Sci. 22 (16) (2021).
[22] A. Crits-Christoph, et al., Genome sequencing of sewage detects regionally prevalent SARS-CoV-2 variants, mBio 12 (1) (2021).
[23] D.S. Smyth, et al., Tracking cryptic SARS-CoV-2 lineages detected in NYC wastewater, Nat. Commun. 13 (1) (2022) 635.
[24] M. Xiao, et al., Multiple approaches for massively parallel sequencing of SARS-CoV-2 genomes directly from clinical samples, Genome Med. 12 (1) (2020) 57.
[25] K. Itokawa, et al., Disentangling primer interactions improves SARS-CoV-2 genome sequencing by multiplex tiling PCR, PLoS One 15 (9) (2020) e0239403.
[26] F. Amman, et al., Viral variant-resolved wastewater surveillance of SARS-CoV-2 at national scale, Nat. Biotechnol. 40 (12) (2022) 1814–1822.
[27] X. Zhang, et al., Detection of the SARS-CoV-2 delta variant in the transboundary rivers of yunnan, China, ACS ES T Water 2 (12) (2022) 2367–2377.
[28] G. Ou, et al., Wastewater surveillance and an automated robot: effectively tracking SARS-CoV-2 transmission in the post-epidemic era, Natl. Sci. Rev. 10 (6) (2023) nwad089.
[29] D. Zhang, et al., Potential spreading risks and disinfection challenges of medical wastewater by the presence of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) viral RNA in septic tanks of Fangcang Hospital, Sci. Total Environ. 741 (2020) 140445.
[30] J. Wang, et al., SARS-CoV-2 RNA detection of hospital isolation wards hygiene monitoring during the Coronavirus Disease 2019 outbreak in a Chinese hospital, Int. J. Infect. Dis. 94 (2020) 103–106.
[31] L. Zhao, et al., Environmental surveillance of SARS-CoV-2 RNA in wastewater systems and related environments in Wuhan: april to May of 2020, J. Environ. Sci. (China) 112 (2022) 115–120.
[32] W.Y. Ng, et al., The city-wide full-scale interactive application of sewage surveillance programme for assisting real-time COVID-19 pandemic control - a case study in Hong Kong, Sci. Total Environ. 875 (2023) 162661.
[33] D. Fu, et al., Preplanned studies: effectiveness of COVID-19 vaccination against SARS-CoV-2 omicron variant infection and symptoms-China, december 2022-february 2023, China CDC Weekly 5 (17) (2023) 369–373.
[34] E. Stachler, et al., Quantitative CrAssphage PCR assays for human fecal pollution measurement, Environ. Sci. Technol. 51 (16) (2017) 9146–9154.
[35] Prevention, N.I.f.V.D.C.a., Specific Primers and Probes for Detection 2019 Novel Coronavirus, 2020.
[36] J. Wu, et al., Technical framework for wastewater-based epidemiology of SARS-CoV-2, Sci. Total Environ. 791 (2021) 148271.
[37] Y. Turakhia, et al., Ultrafast Sample placement on Existing tRees (UShER) enables real-time phylogenetics for the SARS-CoV-2 pandemic, Nat. Genet. 53 (6) (2021) 809–816.
[38] T. Chen, et al., The genome sequence archive family: toward explosive data growth and diverse data types, Dev. Reprod. Biol. 19 (4) (2021) 578–583.
[39] Members, C.-N. and Partners, Database resources of the national genomics data center, China national center for bioinformation in 2022, Nucleic Acids Res. 50 (D1) (2022) D27–D38.
[40] M.A. Sabar, R. Honda, E. Haramoto, CrAssphage as an indicator of human-fecal contamination in water environment and virus reduction in wastewater treatment, Water Res. 221 (2022) 118827.
[41] M.L. Wilder, et al., Co-quantification of crAssphage increases confidence in wastewater-based epidemiology for SARS-CoV-2 in low prevalence areas, Water Res. X 11 (2021) 100100.
[42] P.M. D'Aoust, et al., Catching a resurgence: increase in SARS-CoV-2 viral RNA identified in wastewater 48 h before COVID-19 clinical tests and 96 h before hospitalizations, Sci. Total Environ. 770 (2021) 145319.
[43] Z. Li, et al., Development and clinical application of a rapid IgM-IgG combined antibody test for SARS-CoV-2 infection diagnosis, J. Med. Virol. 92 (9) (2020) 1518–1524.
[44] R. Challen, et al., Risk of mortality in patients infected with SARS-CoV-2 variant of concern 202012/1: matched cohort study, BMJ 372 (2021) n579.
[45] N.G. Davies, et al., Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7, Nature 593 (2021) 270–274.
[46] G.F. Ni, et al., Novel multiplexed amplicon-based sequencing to quantify SARS-CoV-2 RNA from wastewater, Environ. Sci. Technol. Lett. 8 (8) (2021) 683–690.
[47] A.W. Lambisia, et al., Optimization of the SARS-CoV-2 ARTIC network V4 primers and whole genome sequencing protocol, Front. Med. 9 (2022) 836728.
[48] F. Wu, et al., A new coronavirus associated with human respiratory disease in China, Nature 579 (7798) (2020) 265–269.
[49] J. Quick, et al., Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples, Nat. Protoc. 12 (6) (2017) 1261–1276.
[50] W. Ahmed, et al., Detection of the Omicron (B.1.1.529) variant of SARS-CoV-2 in aircraft wastewater, Sci. Total Environ. 820 (2022) 153171.
[51] W. Ahmed, et al., RT-qPCR and ATOPlex sequencing for the sensitive detection of SARS-CoV-2 RNA for wastewater surveillance, Water Res. 220 (2022) 118621.
[52] Y. Wang, et al., Detection of SARS-CoV-2 variants of concern with tiling amplicon sequencing from wastewater, ACS EST Water 2 (2022) 2185–2193.
[53] A. Rambaut, et al., A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology, Nat Microbiol 5 (2020) 1403–1407.
[54] M.B. Diamond, et al., Wastewater surveillance of pathogens can inform public health responses, Nat Med 28 (10) (2022) 1992–1995.