

Technical framework for wastewater-based epidemiology of SARS-CoV-2 based on relative quantification via qPCR

Jinyong Wu

Department of Pharmaceutical & Biological Engineering, School of Chemical Engineering, Sichuan University

Zizheng Wang

West China School of Stomatology, Sichuan University

Yufei Lin

Sichuan University

Lihua Zhang

Sichuan University

Jing Chen

Department of Pharmaceutical & Biological Engineering, School of Chemical Engineering, Sichuan University

Panyu Li

Department of Pharmaceutical & Biological Engineering, School of Chemical Engineering, Sichuan University

Wenbin Liu

Department of Pharmaceutical & Biological Engineering, School of Chemical Engineering, Sichuan University

Yabo Wang

Department of Pharmaceutical & Biological Engineering, School of Chemical Engineering, Sichuan University

Changhong Yao

Department of Pharmaceutical & Biological Engineering, School of Chemical Engineering, Sichuan University

Kun Yang (✉ cookyoung@scu.edu.cn)

Department of Pharmaceutical & Biological Engineering, School of Chemical Engineering, Sichuan University <https://orcid.org/0000-0001-5300-8119>

Keywords: Wastewater-based epidemiology, COVID-19, Relative quantification, RT-qPCR, Modeling, Decay kinetics

Posted Date: December 22nd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-128176/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Technical framework for wastewater-based epidemiology of SARS-CoV-2 based on relative quantification *via* qPCR

Jinyong Wu¹, Zizheng Wang², Yufei Lin¹, Lihua Zhang¹, Jing Chen¹, Panyu Li¹,

Wenbin Liu¹, Yabo Wang¹, Changhong Yao¹, Kun Yang^{1*}

1. Department of Pharmaceutical & Biological Engineering, School of Chemical Engineering, Sichuan University, Chengdu 610065, China.
2. West China School of Stomatology, Sichuan University, Chengdu 610041, China.

*Correspondence to:

Kun Yang

cookyoung@scu.edu.cn

Highlights

1. Technical framework for wastewater-based epidemiology of SARS-CoV-2 is established.
2. The epidemic burden of community is determined *via* qPCR relative quantification.
3. Per capita viral shedding and sewage production affect the detectability.
4. Ideal detectability is one single infected person among up to 400,000 population.
5. The decay of SARS-CoV-2 RNA in wastewater is not the key factor.

ABSTRACT

The global pandemic of coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has greatly affected people's normal production, living, social and economic activities. Wastewater-based epidemiology (WBE) is expected to become a powerful tool to monitor the dissemination of SARS-CoV-2 at the community level, which has attracted the attention of scholars all over the world. However, there is not yet a standard protocol to guide its implementation. In this paper, we attempt to propose a technical framework of relative quantification for determining the virus abundance in wastewater and estimating the infection rate in corresponding communities, which is expected to achieve horizontal and vertical comparability in virtue of human-specific biomarkers as internal references. A comprehensive theoretical framework for relative quantification of viruses by qPCR is provided and discussed in detail. Critical factors affecting the virus detectability in wastewater and the estimation of infection rate include virus concentration methods, lag-period, per capita virus shedding amount, sewage generation rate, temperature-related decay kinetics of virus/biomarkers in wastewater, and hydraulic retention time (HRT), etc. Theoretical simulation shows that the main factors affecting the detectability of virus in sewage are per capita virus shedding amount and sewage generation rate. While the decay of SARS-CoV-2 in sewage is a relatively slow process, which has limited impact on its detection. Under the ideal condition of high per capita virus shedding amount and low sewage generation rate, WBE can give early warning for single infected person among 400,000 people.

Keywords: Wastewater-based epidemiology, COVID-19, Relative quantification, RT-qPCR, Modeling, Decay kinetics.

1. Introduction

The ongoing global pandemic of coronavirus disease 2019 (COVID-19) is caused by a novel severe acute respiratory syndrome coronavirus (SARS-CoV-2) [1]. It has been a Public Health Emergency of International Concern, which has caused confirmed infection of 59,759,494 and death of 1,409,252 (Coronavirus Disease Situation Reports, World Health Organization, accessed 28 Oct. 2020, <https://go.nature.com/3aahjbg>). In recent studies, the RNA of SARS-CoV-2 has been detected in the stool of symptomatic, asymptomatic, pre-symptomatic and even post-symptomatic infected individuals [2-11]. The fecal SARS-CoV-2 RNA detection rate ranged from 15 to 83% among infected patients. The load of SARS-CoV-2 in feces is in the range 5×10^3 – $10^{7.6}$ RNA copies/mL [12]. Meanwhile, researchers all over the world have also detected the RNA of SARS-CoV-2 from the local sewage [13-23]. Thus, wastewater-based epidemiology (WBE) of SARS-CoV-2 which surveys sewage for virus RNA to inform epidemiological monitoring of COVID-19 has been proposed as the most promising method to determine the burden of undiagnosed infections at the population level [24-27]. WBE, defined as the determination of per capita mass loads of analyte in sewage, provides population-scale information on human activity within catchment boundaries [28]. It has been theorized for two decades [29] and was initially implemented for estimating the community drug abuse [30]. Thereafter its application field expands to community-level assessments on per capita consumption, use, exposure, or release of various chemical or biological agents (Figure 1) [28].

At present, reverse transcription quantitative polymerase chain reaction (RT-

qPCR) is commonly used to detect the RNA of SARS-CoV-2 in sewage. Most of them adopt absolute quantification, and virus gene standard (normally a plasmid with SARS-CoV-2 genes) is needed to construct standard curve [22]. Absolute quantification gives the real concentration of virus RNA in sewage, which can be correlated with local clinical observations [17, 22, 31]. While other studies only reported the value of cycle threshold (C_T) [16] or even just gave qualitative results [19]. Such data provide limited information for monitoring actual infection rate in the community. Various factors affect the accuracy of absolute quantitative estimation of community infection rate or hinder the lateral comparison between different regions. For example, in the combined sewage system, the diluting effect of rainfall will affect the concentration of virion in sewage. In addition, the individual sewage generation rate is in the range of 50-500 L/person/day, which fluctuates with seasons and affects the concentration of virus in the sewage. Furthermore, in order to estimate the infection rate in a community, it is necessary to accurately measure the community population, i.e. the population normalization for the data [32]. This is a challenging task for areas with high population mobility. Several studies have tried to quantify some population-related biomarkers, such as human ribonuclease P [14], creatinine, urea, benzotriazole, caffeine etc. [15, 17] to achieve this goal. Surprisingly, so far, no research has adopted relative quantification using a human-specific biomarker as internal reference.

In this paper, we try to provide a better alternative for monitoring the SARS-CoV-2 in sewage, adopting the relative quantification with qPCR. Based on this concept, a self-contained technical framework is to be proposed by incorporating a human-specific

fecal biomarker as internal reference during the quantification. The nucleic acid property of the chosen biomarkers confers themselves the possibility of being quantifiable with the same processes and platforms as the viral nucleic acid. The promising biomarkers that can be used as internal reference in the *relative quantification* with qPCR are recommended. The critical stages and key points involved in the WBE of SARS-CoV-2 are evaluated. The decay kinetics of biomarkers in sewage, which is one of important factors affecting the detection sensitivity, is studied in a mimic mode in virtue of qPCR.

2. Theory and methodology

2.1. Relative quantification via qPCR

Two popular operation modes of qPCR are available for the quantification of genetic RNA or DNA: SYBR green and TaqMan. In this study, we prefer using the TaqMan mode due to its improving specificity. During PCR amplification, the fluorescence signal increment at the Nth amplification cycle (ΔR_n) can be expressed as:

$$\Delta R_n = \phi c_0 \xi^N \quad (1)$$

where ϕ is the luminescent intensity of unit fluorescence molecule; c_0 is the initial copy number of the gene to be amplified; ξ is the apparent amplification coefficient, which is 2 under ideal conditions.

By selecting the characteristic genes of biomarkers related to population (such as characteristic genes of human feces-associated bacteria or bacteriophage [33]) as reference genes, the relative quantification of the target virus genes can be realized. The

same threshold of fluorescence signal (ΔR_T) is set for both reference and target genes.

$$\Delta R_T = \phi_t c_{0,t} \xi_t^{C_{T,t}} = \phi_r c_{0,s} \xi_s^{C_{T,s}} \quad (2)$$

In the case of using the same fluorescent reporter molecule (i.e. $\phi_t = \phi_r$), the relative abundance of the target virus gene (R_a) can be expressed as:

$$R_a = \frac{c_{0,t}}{c_{0,s}} = \frac{\xi_s^{C_{T,s}}}{\xi_t^{C_{T,t}}} \approx 2^{C_{T,s} - C_{T,t}} \quad (3)$$

Where C_T is the threshold cycle number; the subscript s indicates the reference gene; and the subscript t refers to the target virus gene.

For the population in a certain sewage catchment, the relative abundance of virus gene in the sewage is proportional to the infection rate (I_r) in the catchment, as shown in Eq. (4):

$$R_a = \frac{S_{r,t} \cdot \eta_t \cdot r_t}{S_{r,s} \cdot \eta_s \cdot r_s} \cdot I_r \quad (4)$$

Where the infection rate $I_r = \frac{P_I}{P_T}$, is the ratio of the number of infected persons (P_I) to the total population (P_T) in the sewage catchment; S_r is the personal shedding amount of biomarkers (copies/person/day); η is the decay rate of biomarkers in the sewage related to temperature and hydraulic retention time (HRT); and r is the recovery rate of biomarkers. Similarly, the subscript s indicates the reference gene; and the subscript t refers to the target virus gene. Then the relative abundance of target virus gene in the feces sample of infected person ($R_{a,0}$) can be expressed as:

$$R_{a,0} = \frac{S_{r,t}}{S_{r,s}} \quad (5)$$

Then, the infected rate can be calculated with the following equation.

$$I_r = \frac{R_{a,0} \cdot \eta_s \cdot r_s}{R_{a,0} \cdot \eta_t \cdot r_t} \quad (6)$$

2.2. *In silico* performance evaluation of primer-probe sets

SARS-CoV-2 is an enveloped positive-sense single-stranded RNA virus, showing a higher propensity for mutation [34, 35]. The inclusivity of the primer-probe sets currently used in RT-qPCR needs to be evaluated. Over 5,000 SARS-CoV-2 genome sequences were downloaded from GISAID's EpiCoV™ database in early April. After filtering out the sequences with poor quality, the remaining 3986 sequences were used for in silico analysis of present primer-probe sets. The genome sequence of bat coronavirus RaTG13, which has an identity of 96.2% with the genome of SARS-CoV-2, was also included [1]. All the downloaded genome sequences were clustered with CD-HIT-EST using a sequence identity cut-off of 1.0 to group genomes with identical sequence into one cluster [36]. Representative genome sequences (n=2257) of all clusters were aligned with the multiple sequence alignment program MAFFT [37]. The mutation rate of each site in the target binding region of every primer/probe was recorded, respectively. The primer-probe sets with less mutation rate in the targeting region is considered to be more inclusive. The exclusivity of primer-probe sets was also determined by assessing their ability to distinguish SARS-CoV-2 from RaTG13. The primer pairs with high inclusivity and high exclusivity were further evaluated *via* Primer-BLAST at the website of national center for biotechnology information (NCBI).

2.3. Virus detectability in sewage samples

The concentration of biomarkers corresponding to the reference gene (population-related bacteria or bacteriophages) is relatively high in the sewage. Therefore, it is expected that there will not be a situation in which they cannot be

detected. The detection limit of the relative quantification is the same as that of the absolute one, which is determined by the concentration of virus RNA in the sewage.

In this work, the sensitivity of the whole method is expressed in terms of the scale of a Population containing a Single Infected Person (PSIP), among which the SARS-CoV-2 RNA shedding off the infected person is detectable in the sewage samples collected from the catchment of the same population. The higher the value, the more sensitive the detection is. The SARS-CoV-2 loading to sewage was estimated using recently reported its excretion rate (E_r) in human stool (copies/g feces), and assuming a fecal load (M_f) in the range of 100–400 g feces/day/person. Then, the PSIP can be expressed as:

$$PSIP = \frac{E_r \cdot M_f}{1000 \cdot LOD \cdot V_s} \quad (7)$$

Where LOD means the lowest limitation of detection, which is set to be 1 copy/mL sewage [38]; V_s is the per capita sewage generation rate (50-500 L/person/day).

2.4. Decay kinetics of biomarker in Sewage

It worth noting that the decay kinetics of biomarkers in wastewater is an important factor to be considered whether it is relative or absolute quantification. The biomarkers (reference microbe or the virus) will decay exponentially with time in sewage of given composition [27, 39].

$$n = n_0 e^{-kt} = n_0 \left(\frac{1}{2}\right)^{\frac{t}{t_{1/2}}} \quad (8)$$

Where n_0 is the initial amount of biomarkers discharged into the sewage system; n is the amount after hydraulic retention time (HRT) of t ; k is the first order decay rate

constant; and $t_{1/2}$ the half-life (i.e. T_{50} the time for half decay).

The decay rate η can be expressed as:

$$\eta = \frac{n}{n_0} = \left(\frac{1}{2}\right)^{\frac{t}{t_{1/2}}} \quad (9)$$

According to Eq. (8)

$$k = \frac{\ln(2)}{t_{1/2}} \quad (10)$$

$$T_{90} = \frac{\ln(10)}{k} \quad (11)$$

$$T_{99} = \frac{\ln(100)}{k} \quad (12)$$

Where T_{90} is the time required for 1 log10 decay, and T_{99} the time for 2 log10 decay.

The decay of biomarkers in sewage is closely related to the temperature, which is expressed with the Arrhenius equation.

$$\ln k = \frac{-E_a}{RT} + C \quad (13)$$

Where E_a is the activation energy of the decay reaction (J/mol); R is the ideal gas constant (8.314 J/mol/K); T the absolute temperature (K); and C the constant related to standard state setting. Therefore, the decay rate constant and half-life at different sewage temperatures can be calculated using Eqs. (14) and (15).

$$k_2 = k_1 \times e^{\frac{E_a(T_2 - T_1)}{RT_1 T_2}} \quad (14)$$

$$t_{1/2,2} = \frac{t_{1/2,1}}{e^{\frac{E_a(T_2 - T_1)}{RT_1 T_2}}} \quad (15)$$

Where k_1 is the decay rate constant at temperature T_1 ; k_2 the decay rate constant at temperature T_2 ; $t_{1/2,1}$ is half-time at temperature T_1 ; $t_{1/2,2}$ is half-time at temperature T_2 .

The decay kinetics of the virus in the sewage also affects the detection sensitivity.

According to the linear correlation between $\ln k$ and $1/T$ [Eq. (13)], the decay activation energy (E_a) of SARS-CoV-2 RNA in sewage can be deduced using the reported decay rate constant at different temperatures [39]. The decay rate constant at each temperature is calculated by introducing E_a into Eq. (14). The impact of temperature and HRT on detection sensitivity is evaluated.

3. Results and discussion

3.1. Critical stages and key points

The critical stages and key points involved in the wastewater-base epidemiology of SARS-CoV-2 are summarized in Table 1. Many factors need to be considered in both relative and absolute quantification, while relative quantification exhibits more advantages over absolute quantification. The ultimate purpose of determining the virus content in sewage is to estimate the infection rate at population level. The results from absolute quantification are sensitive to the fluctuation of sewage generation rate (V_s) and the dilution effect of rainfall events. As the reference and target biomarker will be diluted to the same extent when sewage generation rate increases or the rainfall event occurs, the relative quantification will not be affected too much. In our previous work, qPCR was used to quantify the prevalence of enteropathogenic *E. coli* (EPEC) in sewage successfully. The fluctuation of absolute quantification data spanned two orders of magnitude over a period of one year, while relative quantification data changed by no more than an order of magnitude [40]. The relative quantitative data more accurately reflected the changes of EPEC in the sewage. In the relative quantification process, it

is not necessary to accurately record the volume of samples used for virus recovery and nucleic acid extraction, nor to accurately determine the volume of the final nucleic acid extract. In addition, there is no need of the target gene standard to construct the standard curve. Even so, in order to ensure the reproducibility of the results and the standardization of the methods, essential information such as sample processing procedure, qPCR protocol and data analysis method etc. should be provided when reporting, as recommended in previously published guidance [41].

One of the major challenges in the practice of WBE on the community monitoring of COVID-19 is the establishment of standardized methods and procedures. Among them, the recovery of virus from sewage samples is the first and most important step. Table 2 summarizes present studies reporting the detection of SARS-CoV-2 in wastewater. As shown in the table, at present, the main virus concentration methods for SARS-CoV-2 from sewage samples are PEG precipitation, electronegative membrane adsorption and ultrafiltration etc. These methods of concentrating nonenveloped virus from sewage are more targeted at the recovery of virus particles from the sewage supernatant which has been pretreated to remove suspended solids and bacteria. However, as an enveloped virus, SARS-CoV-2 is different from nonenveloped virus in nature, which will affect its partition behavior in sewage. Therefore, the concentration method should be adjusted accordingly [42]. According to our knowledge and reported studies, the virus adsorbed on the sewage particles cannot be neglected and may even account for the major portion of the SARS-CoV-2 in the sewage [17, 43]. By far the highest value of virus content ever detected in sewage-related samples is from the

primary sewage sludge, and the value is as high as 4.6×10^8 RNA copies/L [14]. Simultaneous recovery of virions from both liquid and solid phases also gave higher detection sensitivity (higher PSIP values) as shown in Table 2 [19, 44]. However, we do not recommend directly using the virus RNA abundance in the primary sewage sludge to estimate the community infection rate. The HRT of sludge is longer than that of wastewater, resulting in the accumulation of virus in sludge. If one wants to use the abundance of sludge virus RNA to estimate the community infection rate, the retention time of sludge or the residence time distribution need to be considered. Accordingly, we suggest that virus RNA of both solid and liquid phase of sewage samples should be considered in wastewater surveillance. Effective recovery of virus from sewage sample will largely prompt the detectability of the SARS-CoV-2 RNA.

According to Eq. (6), the infection rate of communities in the sewage catchment can be calculated from the relative abundance of the target virus gene. The premise is to determine the average relative abundance of the target virus gene in the feces of the infected person. It has been reported that not all infected people are positive for RNA of SARS-CoV-2 in their feces, virus shedding amount in the feces varies from person to person, which also fluctuates throughout the infection period [8, 10]. In order to incorporate these factors into consideration, the range of infection rate is ought to be estimated *via* such as Monte Carlo model [18]. In addition, the decay kinetics and recovery efficiency of viruses and reference biomarkers from sewage are also factors that need to be considered, which will be elaborated in the following sections. Under the relatively loose experimental operation standard, the relative abundance data

obtained by different laboratories can be compared horizontally. In addition, the relative abundance data obtained from sewage catchments of different population sizes can be directly used to compare the severity of epidemics among communities. The results can be used to guide regional policy adjustment for either restart of economies or continuous lockdown. For a specific community within a sewage catchment, the parameters (η_s , η_t , r_s and r_t) in Eq. (6) can be considered as constants when using fixed sampling mode, virus concentration method, nucleic acid extraction method and standardized qPCR relative quantification operation. The equation can be simplified to

$$I_r = K \frac{R_a}{R_{a,0}} \quad (16)$$

If accurate clinical diagnosis data are available, the constant K can be accurately estimated by correlating the actual infection rates with the relative abundances of virus RNA in sewage. With sufficient basic data, the method of WBE can even be combined with machine learning to further improve the prediction accuracy of community infection rate.

Besides the above key points which may affected the performance of the quantification, biosafety related factors also need to be considered. Scholars in Italy and Germany have proved that the infectivity of SARS-CoV-2 in sewage is limited [15, 17]. However, the possibility of fecal-oral transmission cannot be completely ruled out. Furthermore, it seems that SARS-CoV-2 can be effectively removed in the sewage treatment process. Scholars from China, Spain, India, and Italy have detected the RNA of SARS-CoV-2 in the influent of the sewage treatment plant, but not in the effluent [15, 16, 20, 45]. That is to say, the location of sampling may affect the detectability of

the virus. Personal protection (wearing N95 mask, gloves, and goggles etc.) during sampling and pasteurization of samples (60 °C, 30-90 min) before virus concentration are essential for biosafety reasons.

3.2. *Performance of primer-probe sets*

A total of 18 currently used primer-probe sets were evaluated in this work (Supplementary Table S1). Their positions in the SARS-CoV-2 genome are illustrated in Supplementary Figure S1. The mutation rate of SARS-CoV-2 genome at the probe/primer targeting regions is shown as a heatmap in Figure 2A. The original data are available as online supplementary datasets (Supplementary Dataset S1 and S2). Four primer-probe sets with high inclusivity (low mutation rate at their targeting regions, marked in green in Figure 2A) and high exclusivity (more mismatches with the sequence of RaTG13, Figure 2B) were recommended (Table 3). Sequences of some primers were amended according to the sequence alignment and/or Primer-BLAST analysis, and the primers were renamed accordingly. Although we have carefully evaluated these primer-probe sets, it was still found that some mutant virus sequences in the database, especially those uploaded after April, could not perfectly match the primer/probe sequences. Therefore, we suggest that two or three primer-probe sets should be used simultaneously in practice to ensure the accuracy of detection results.

3.3. *Reference biomarkers*

To realize the relative quantification of SARS-CoV-2, a population-related reference biomarker needs to be chosen. There are numerous assays available for the

quantitative assessment of human fecal pollution [46], which can meet this demand. Those qPCR-based microbial source tracking (MST) methods targeting mitochondrial DNA, rRNA and functional genes involved in microorganism-human interaction are favorable candidates. In Table 4, we recommend some promising qPCR assays that have been proved to be effective for the detection of human-associated MST markers. The genes from the three main sources can be used as reference genes for the relative quantification: genes from human hosts (mitochondria) [47], genes from human-associated bacteria (*Bacteroidales*) [48] and bacteriophages (crAssphage) [33]. As shown in Table 3, the MST method targeting the reference gene should be of high enough sensitivity (true-positive rate) and specificity (true-negative rate). Furthermore, the reference gene must also be of a high enough content in feces or waste, so as to ensure that it can still be detected in the sewage even at a high dilution level.

The validity of reference biomarkers (reference genes) can be verified by comparing the abundance of different reference biomarkers. The premise of this self-consistent mode of the validation is that all reference biomarkers are all population related. Dilution will not affect the relative abundance between different reference biomarkers. Therefore, the relative abundance will not fluctuate dramatically over time. The result can also support the feasibility of relative quantification of SARS-CoV-2. In addition, reasonably selected reference gene itself is a good internal control of the whole method, which can exclude the influence of concentration efficiency, inhibiting substance of PCR and other factors to a certain extent.

3.4. Factors affecting the detectability

First, personal virus shedding amount and per capita sewage production will directly affect the virus concentration in the sewage, thus affecting the detectability. The virus shedding varies from person to person. Even for the same infected person, the virus emission also changes during the whole infection period. The personal sewage production fluctuates with seasons. According to previous reports, the load of SARS-CoV-2 in feces is in the range of 10^3 - 10^8 RNA copies/g [8, 12, 24]. We assume the fecal load (M_f) in the range of 100–400 g feces/person/day and the sewage generation rate in the range of 100-400 L/person/day [27]. The detection sensitivity (PSIP) is calculated with Eq. (7) as high as 4×10^5 (Figure 3), which means that one single infected person is detectable among a population of 4×10^5 at ideal conditions with the highest load of virus in feces (10^8 RNA copies/g), the highest fecal load (400 g feces/day/person) and the lowest sewage generation rate (100 L/person/day). With the decrease of personal virus shedding amount and the increase of sewage generation rate, the detection sensitivity gradually decreased to $PSIP=0.25$, which means that in the case of extremely low virus shedding amount (1×10^5 RNA copies/day/person) and high sewage production (400 L/person/day), the detection of sewage from one single infected person may also be negative. However, in practice, when encountering an infected person with such a low virus shedding amount, it may be caused by the following two situations: one is that the person has just been infected, and the virus has not yet been copied in large quantities; the other is that the patient has recovered from the disease and the virus shedding has been reduced to a very low level. The above two types of infection do limited harm to the community, so even if they are not detected, they will not have a

great impact on the disease surveillance in the community.

On the other hand, the detectability is also affected by the decay kinetics of virus and the HRT. The decay of virus may be due to biological and chemical activities in sewage [49]. Pasteurization can delay the decay kinetics of virus in sewage *via* eliminating bacterial extracellular enzyme activity and protozoan or metazoan predation [42]. Recently, Ahmed et al. reported the decay kinetics of SARS-CoV-2 in pasteurized and unpasteurized sewage [39]. We extracted the decay kinetics data of SARS-CoV-2 RNA in unpasteurized sewage from Table 3 in their paper. According to Eq. (11), plotting $\ln k$ versus $1/T$ gave a straight line with a slope of $-E_a/R$. Thus, the value of E_a was determined to be 27110.3 (J/mol). As shown in Figure 3C and D, the values of k , $t_{1/2}$ and T_{90} in the temperature range of 0 to 40 °C were also calculated with Eqs. (14), (15) and (11), respectively. If the virus shedding amount of 400×10^8 copies/person/day and sewage generation rate of 100 L/person/day are set, the detection sensitivity (PSIP) varies between 2.94×10^5 and 4×10^5 , as the HRT and sewage temperature change (Figure 3B). The calculated values of T_{90} at different temperatures are in good agreement with the experimental ones ($T_{90,exp}$ in Figure 3D). The HRT (~ 24 h) of the virus in sewage is relatively short in comparison to its half-life (2.25-10.36 days). Overall, the decay of virus has little effect on its detectability (Figure 3B). However, as an important part of the technical system of WBE, the decay kinetics of pathogenic microorganisms (or their RNA/DNA) in sewage should be systematically evaluated. If necessary, the main chemical and biological factors affecting their persistence in the sewage should be determined.

The above simulation results show that the most important factor affecting the detectability of SARS-CoV-2 RNA in sewage is the virus shedding amount off the infected person. In this work, we only discussed the sewage load of virus *via* feces of the infected person. But in fact, other virus-containing secretions of the patient may also be discharged into the sewage system, such as sputum. The virus content in sputum is as high as 2.35×10^9 RNA copies/mL [8]. Therefore, the shedding amount of virus in the sewage from infected persons should be higher than what we have estimated. The previous simulation study has shown that the detection of one infected case among a population of 2,000,000 is theoretically feasible [27]. However, according to our estimation, this population should be in the order of 100,000 among which the detectability of single infected person is more promising. The experimental study also has proved that sewage sample started to give positive RT-qPCR signal of virus genes when the observed COVID-19 prevalence was around or even below one case in 100,000 people [22]. According to the calculations, under optimal conditions, one infected person can be detected in a population of 400,000 by the means of WBE. For Chengdu, a city with 16 million residents, a reasonable layout of 40 to 160 sampling points will be sufficient for the entire city's epidemic surveillance of SARS-CoV-2. Such a small-scale of detection (40-160 samples) can even be routinely completed in our laboratory every day to implement the real-time monitoring of the epidemic situation for the entire city.

3.5. *Mimic study on the decay kinetics of biomarkers*

The decay kinetics of biomarkers (target or reference genes) in sewage can be studied *via* microcosm experiments in the laboratory [39]. Freshly collected sewage samples will be spiked with certain number of biomarkers (bacteria, bacteriophages, or inactivated SARS-CoV-2 virions) and incubated at certain temperature. QPCR will be still used to track the changes of biomarkers in the sewage.

$$\frac{c_{i,t}}{c_{i,i}} = \left(\frac{1}{2}\right)^{\frac{t}{t_{1/2}}} = \xi^{C_{T,i}-C_{T,t}} \approx 2^{C_{T,i}-C_{T,t}} \quad (17)$$

$$C_{T,t} = \frac{\ln 2 \cdot t}{\ln \xi \cdot t_{1/2}} + C_{T,i} \approx \frac{t}{t_{1/2}} + C_{T,i} \quad (18)$$

Where $c_{i,t}$ is the concentration of biomarker i in the sewage at time t (RNA or DNA copies/mL sewage), $c_{i,i}$ is the initial concentration of biomarker i (RNA or DNA copies/mL sewage), $C_{T,i}$ the threshold cycle number for the initial sewage sample, $C_{T,t}$ the threshold cycle number for the sewage sample at time t . Plotting $C_{T,t}$ versus t will give a straight line with a slope of $1/t_{1/2}$ (Eq. (17)). Thus, one can obtain the half-life of biomarkers at different temperature. Then plotting $\ln k$ versus $1/T$ will also give a straight line with a slope of $-E_a/R$ (Eq. (12)). Introducing E_a into Eq. (15) can calculate the half-life ($t_{1/2}$) at any temperature.

Considering the pathogenicity of SARS-CoV-2, we can alternatively use the enveloped virus surrogates to study the decay kinetics in sewage and mimic its recovery from sewage samples. Typical surrogates are murine hepatitis virus (MHV) and Pseudomonas phage $\phi 6$ [42]. According to the detected temperature of sewage samples, introducing corresponding $t_{1/2}$ and the estimated HRT (t) into Eqs. (9) and (6) will give a reasonable estimation of the real infection rate in the sewage catchment with the experimentally determined relative abundance of SARS-CoV-2. In addition, if the

bacteriophage is chosen as the reference biomarker, its decay behavior in sewage may be similar with SARS-CoV-2, due to their similar properties [50].

3.6. *Summary of the technical framework*

The technical framework of WBE for SARS-CoV-2 is illustrated in Figure 4. Five main steps are involved: sampling, virus concentration, RNA/DNA extraction, RT-qPCR/qPCR, and data/results reporting. According to the different research purposes and local epidemic situations, different sampling strategies may be adopted. In areas where the COVID-19 is still exploding, we can not only track the change of infection rate in a community over time, but also compare the epidemic burden of different communities. For the former, the strategy of repeated sampling at the same sampling point on different dates (temporal) will be adopted, which can be used to track the changing trend of community epidemic. While for the latter, samples will be collected in different communities (spatial) on the same dates, which can realize the horizontal comparison of epidemic burden among different communities and provide regional guidance for keeping isolation or restarting the economy. If accurate community infection data are available, both strategies can eventually establish a quantitative relationship between the abundance of sewage virus RNA and community infection rate. In a low prevalence area where the epidemic has been under control, reasonable layout of sampling points can realize real-time monitoring of the whole area and prevent the resurgence of the epidemic caused by imported asymptomatic infected persons or virus contaminated cold chain foods. Once a positive signal is detected at a certain sampling

point, we can add secondary sampling points in the corresponding area or conduct regional full staff screening to track down the infected persons. Regardless of the sampling strategy, we recommend using 24-hour flow-weighted-composite samples for testing. For the steps of virus concentration and RNA/DNA extraction, we recommend recovering virus from both solid and liquid phase and extracting both RNA and DNA from concentrated samples, because of the adsorption tendency of enveloped virus on solid particles and the DNA property of the reference gene. In the detection step, the RT-qPCR and qPCR will be applied for the quantification of viral gene and the reference gene, respectively. When reporting the data, it is necessary to clarify some important matters that affect the interpretation of the results, such as the lag-period between WBE and clinic data, personal virus shedding amount and sewage generation rate, and the decay kinetics and HRT of biomarkers (virus RNA and reference gene) etc.

4. Conclusion

- By incorporating human-specific fecal biomarkers (genes) as internal reference of qPCR, a technical framework of WBE based on relative quantification has been established for monitoring the dissemination of SARS-CoV-2.
- Many factors that may affect virus detectability and estimation of community infection rate need to be considered. Such as lag-period, personal virus shedding amount, sewage generation rate, decay kinetics of biomarker/virus, temperature, and HRT, which should be announced at the reporting of the results.
- Inclusivity of present primer-probe sets was evaluated. The primer-probe sets with

good specificity and inclusivity are recommended.

- Recovery and concentration methods for the enveloped virus SARS-CoV-2 need to be optimized. Virions in both the liquid and solid phases of sewage should be recovered simultaneously.
- The technical framework established in this paper can not only be used to monitor the current COVID-19 pandemic, but also monitor other viruses in sewage, and carry out community epidemiological monitoring of corresponding diseases or provide early warning for the next human pandemic.

Author contributions

J.W. and K.Y. designed the study. J.W. and K.Y. derived all the theoretical models in this paper. Z.W. and K.Y. conducted the sequence alignment and the evaluation of primer/probe sets. J.W., Z.W. and K.Y. constructed the Figures in the paper. Y.L., L.Z. and K.Y. collected important information from published papers and constructed all tables. J.W. and K. Y. analyzed the data and wrote the first draft. J.C., P.L., W.L., Y.W., and C.Y. participated in the discussion of the project, read the manuscript, and gave constructive suggestions for revision. All authors read and approved 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.

Acknowledgements

We thank the financial supports from the National Natural Science Foundation of

China (No. 21677104). We gratefully acknowledge the authors, originating and submitting laboratories of the sequences from GISAID's EpiFlu™ Database on which this research is based. The list is detailed in the attached acknowledgement table (gisaid_cov2020_acknowledgement_table.xls). Finally, the author (Kun Yang) would like to express his sincere thanks to his longtime friend, Dr. Zheng Yang for the English editing of our articles. At this special time of global pandemic, the author (K.Y.) hopes that he and his family are all well throughout the epidemic situation of COVID-19.

Additional information

- Supplementary Table S1. Information of RT-qPCR primer–probe sets analyzed in this work.
- Supplementary Figure S1. Illustration of the positions of RT-qPCR primer–probe sets in the SARS-CoV-2 genome.
- Supplementary Dataset S1. Statistics of numbers of point mutation sequences (among analyzed 2257 sequences) in the target binding region of primer/probe sets.
- Supplementary dataset S2. Calculated mutation rate for each site within the target binding region of primers/probes (n=2257).
- Acknowledgement table for downloaded genome sequences from GISAID EpiCoV database.

References

1. Zhou, P., et al., *A pneumonia outbreak associated with a new coronavirus of probable bat origin*. Nature, 2020. **579**(7798): p. 270-273.
2. Holshue, M.L., et al., *First Case of 2019 Novel Coronavirus in the United States*. N Engl J Med, 2020. **382**(10): p. 929-936.
3. Gao, Q.Y., Y.X. Chen, and J.Y. Fang, *2019 Novel coronavirus infection and gastrointestinal tract*. J Dig Dis, 2020. **21**(3): p. 125-126.
4. Cai, J., et al., *A Case Series of children with 2019 novel coronavirus infection: clinical and epidemiological features*. Clin Infect Dis, 2020.
5. Tang, A., et al., *Detection of Novel Coronavirus by RT-PCR in Stool Specimen from Asymptomatic Child, China*. Emerg Infect Dis, 2020. **26**(6): p. 1337-1339.
6. Zhang, J., S. Wang, and Y. Xue, *Fecal specimen diagnosis 2019 novel coronavirus-infected pneumonia*. J Med Virol, 2020. **92**(6): p. 680-682.
7. Zhang, W., et al., *Molecular and serological investigation of 2019-nCoV infected patients: implication of multiple shedding routes*. Emerg Microbes Infect, 2020. **9**(1): p. 386-389.
8. Wolfel, R., et al., *Virological assessment of hospitalized patients with COVID-2019*. Nature, 2020. **581**(7809): p. 465-469.
9. Bogler, A., et al., *Rethinking wastewater risks and monitoring in light of the COVID-19 pandemic*. Nature Sustainability, 2020.
10. Zhang, N., et al., *Comparative study on virus shedding patterns in nasopharyngeal and fecal specimens of COVID-19 patients*. Sci China Life Sci, 2020.
11. Wu, Y., et al., *Prolonged presence of SARS-CoV-2 viral RNA in faecal samples*. Lancet Gastroenterol Hepatol, 2020. **5**(5): p. 434-435.
12. Foladori, P., et al., *SARS-CoV-2 from faeces to wastewater treatment: What do we know? A review*. Sci Total Environ, 2020. **743**: p. 140444.
13. Zhang, D., 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, 2020. **741**: p. 140445.
14. Peccia, J., et al., *Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics*. Nat Biotechnol, 2020.
15. Rimoldi, S.G., et al., *Presence and infectivity of SARS-CoV-2 virus in wastewaters and rivers*. Sci Total Environ, 2020. **744**: p. 140911.
16. Kumar, M., et al., *First proof of the capability of wastewater surveillance for COVID-19 in India through detection of genetic material of SARS-CoV-2*. Sci Total Environ, 2020. **746**: p. 141326.
17. Westhaus, S., et al., *Detection of SARS-CoV-2 in raw and treated wastewater in Germany - Suitability for COVID-19 surveillance and potential transmission risks*. Sci Total Environ, 2020. **751**: p. 141750.
18. Ahmed, W., 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, 2020. **728**: p. 138764.
19. La Rosa, G., et al., *First detection of SARS-CoV-2 in untreated wastewaters in Italy*. Sci

- Total Environ, 2020. **736**: p. 139652.
20. Randazzo, W., et al., *SARS-CoV-2 RNA in wastewater anticipated COVID-19 occurrence in a low prevalence area*. Water Res, 2020. **181**: p. 115942.
 21. Sherchan, S.P., et al., *First detection of SARS-CoV-2 RNA in wastewater in North America: A study in Louisiana, USA*. Sci Total Environ, 2020. **743**: p. 140621.
 22. Medema, G., 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., 2020. **7**(7): p. 511–516.
 23. Haramoto, E., et al., *First environmental surveillance for the presence of SARS-CoV-2 RNA in wastewater and river water in Japan*. Sci Total Environ, 2020. **737**: p. 140405.
 24. Kitajima, M., et al., *SARS-CoV-2 in wastewater: State of the knowledge and research needs*. Sci Total Environ, 2020. **739**: p. 139076.
 25. Daughton, C.G., *Wastewater surveillance for population-wide Covid-19: The present and future*. Sci Total Environ, 2020. **736**: p. 139631.
 26. Bivins, A., et al., *Wastewater-Based Epidemiology: Global Collaborative to Maximize Contributions in the Fight Against COVID-19*. Environ Sci Technol, 2020. **54**(13): p. 7754-7757.
 27. Hart, O.E. and R.U. Halden, *Computational analysis of SARS-CoV-2/COVID-19 surveillance by wastewater-based epidemiology locally and globally: Feasibility, economy, opportunities and challenges*. Sci Total Environ, 2020. **730**: p. 138875.
 28. Choi, P.M., et al., *Wastewater-based epidemiology biomarkers: Past, present and future*. Trac-Trends in Analytical Chemistry, 2018. **105**: p. 453-469.
 29. Daughton, C.G., *Emerging pollutants, and communicating the science of environmental chemistry and mass spectrometry: pharmaceuticals in the environment*. J Am Soc Mass Spectrom, 2001. **12**(10): p. 1067-76.
 30. Zuccato, E., et al., *Estimating community drug abuse by wastewater analysis*. Environ Health Perspect, 2008. **116**(8): p. 1027-32.
 31. Nemudryi, A., et al., *Temporal detection and phylogenetic assessment of SARS-CoV-2 in municipal wastewater*. Cell Rep Med, 2020: p. 100098.
 32. Polo, D., et al., *Making waves: Wastewater-based epidemiology for COVID-19 - approaches and challenges for surveillance and prediction*. Water Res, 2020. **186**: p. 116404.
 33. Stachler, E., et al., *Quantitative CrAssphage PCR Assays for Human Fecal Pollution Measurement*. Environ Sci Technol, 2017. **51**(16): p. 9146-9154.
 34. Tang, X., et al., *On the origin and continuing evolution of SARS-CoV-2*. National Science Review, 2020. **7**: p. 1012–1023.
 35. Rambaut, A., et al., *A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology*. Nat Microbiol, 2020.
 36. Huang, Y., et al., *CD-HIT Suite: a web server for clustering and comparing biological sequences*. Bioinformatics, 2010. **26**(5): p. 680-2.
 37. Katoh, K., J. Rozewicki, and K.D. Yamada, *MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization*. Brief Bioinform, 2019. **20**(4): p. 1160-1166.
 38. Vogels, C.B.F., et al., *Analytical sensitivity and efficiency comparisons of SARS-CoV-2 RT-*

- qPCR primer-probe sets*. Nature Microbiology, 2020. **5**(10): p. 1299-+.
39. Ahmed, W., 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, 2020: p. 110092.
 40. Yang, K., E. Pagaling, and T. Yan, *Estimating the prevalence of potential enteropathogenic Escherichia coli and intimin gene diversity in a human community by monitoring sanitary sewage*. Appl Environ Microbiol, 2014. **80**(1): p. 119-27.
 41. Bustin, S.A., et al., *The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments*. Clin Chem, 2009. **55**(4): p. 611-22.
 42. Ye, Y., et al., *Survivability, Partitioning, and Recovery of Enveloped Viruses in Untreated Municipal Wastewater*. Environ Sci Technol, 2016. **50**(10): p. 5077-85.
 43. Balboa, S., et al., *The fate of SARS-CoV-2 in wastewater treatment plants points out the sludge line as a suitable spot for incidence monitoring*. medRxiv, 2020. **2020.05.25.20112706**.
 44. Wurtzer, S., et al., *Evaluation of lockdown impact on SARS-CoV-2 dynamics through viral genome quantification in Paris wastewaters*. medRxiv, 2020. **2020.04.12.20062679**.
 45. Wang, J., 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, 2020. **94**: p. 103-106.
 46. Harwood, V.J., et al., *Microbial source tracking markers for detection of fecal contamination in environmental waters: relationships between pathogens and human health outcomes*. FEMS Microbiol Rev, 2014. **38**(1): p. 1-40.
 47. Schill, W.B. and M.V. Mathes, *Real-time PCR detection and quantification of nine potential sources of fecal contamination by analysis of mitochondrial cytochrome b targets*. Environ Sci Technol, 2008. **42**(14): p. 5229-34.
 48. Shanks, O.C., et al., *Performance of PCR-based assays targeting Bacteroidales genetic markers of human fecal pollution in sewage and fecal samples*. Environ Sci Technol, 2010. **44**(16): p. 6281-8.
 49. Mohamed Hamouda, et al., *Wastewater surveillance for SARS-CoV-2: Lessons learnt from recent studies to define future applications*. Science of the Total Environment, 2020.
 50. McMinn, B.R., N.J. Ashbolt, and A. Korajkic, *Bacteriophages as indicators of faecal pollution and enteric virus removal*. Lett Appl Microbiol, 2017. **65**(1): p. 11-26.
 51. Michael-Kordatou, I., P. Karaolia, and D. Fatta-Kassinos, *Sewage analysis as a tool for the COVID-19 pandemic response and management: the urgent need for optimised protocols for SARS-CoV-2 detection and quantification*. J Environ Chem Eng, 2020. **8**(5): p. 104306.
 52. Mlejnkova, H., et al., *Preliminary Study of Sars-Cov-2 Occurrence in Wastewater in the Czech Republic*. Int J Environ Res Public Health, 2020. **17**(15).
 53. Guerrero-Latorre, L., et al., *SARS-CoV-2 in river water: Implications in low sanitation countries*. Sci Total Environ, 2020. **743**: p. 140832.
 54. Arora, S., et al., *Sewage surveillance for the presence of SARS-CoV-2 genome as a useful wastewater based epidemiology (WBE) tracking tool in India*. medRxiv, 2020. **2020.06.18.20135277**.
 55. Or, I.B., et al., *Regressing SARS-CoV-2 sewage measurements onto COVID-19 burden in*

- the population: a proof-of-concept for quantitative environmental surveillance.* medRxiv, 2020. **2020.04.26.20073569**.
56. La Rosa, G., et al., *SARS-CoV-2 has been circulating in northern Italy since December 2019: Evidence from environmental monitoring.* Sci Total Environ, 2021. **750**: p. 141711.
 57. Hata, A., et al., *Detection of SARS-CoV-2 in wastewater in Japan during a COVID-19 outbreak.* Science of the Total Environment, 2020.
 58. Goncalves, J., et al., *Detection of SARS-CoV-2 RNA in hospital wastewater from a low COVID-19 disease prevalence area.* Sci Total Environ, 2020: p. 143226.
 59. Kocamemi, B.A., et al., *SARS-CoV-2 Detection in Istanbul Wastewater Treatment Plant Sludges.* medRxiv, 2020. **2020.05.12.20099358**.
 60. Kocamemi, B.A., et al., *First Data-Set on SARS-CoV-2 Detection for Istanbul Wastewaters in Turkey.* medRxiv, 2020. **2020.05.03.20089417**.
 61. Wu, F., et al., *SARS-CoV-2 Titers in Wastewater Are Higher than Expected from Clinically Confirmed Cases.* mSystems, 2020. **5**(4).
 62. Artem Nemudryi, A.N., Tanner Wiegand, Kevin Surya, Murat Buyukyoruk, Calvin Cicha, Karl K. Vanderwood, Royce Wilkinson, Blake Wiedenheft, *Temporal Detection and Phylogenetic Assessment of SARS-CoV-2 in Municipal Wastewater.* Cell Rep Med, 2020. **1**(6): p. 100098.
 63. WHO, *WHO Coronavirus Disease (COVID-19) Technical Guidance: Laboratory Testing for 2019-nCoV in Humans.* 2020.
 64. Corman, V.M., et al., *Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR.* Euro Surveill, 2020. **25**: p. 2000045.
 65. Chan, J.F., et al., *Improved Molecular Diagnosis of COVID-19 by the Novel, Highly Sensitive and Specific COVID-19-RdRp/HeI Real-Time Reverse Transcription-PCR Assay Validated In Vitro and with Clinical Specimens.* J Clin Microbiol, 2020. **58**(5).
 66. CDC, U., *2019-Novel Coronavirus (2019-nCoV) Real-time rRT-PCR Panel: Primers and Probes.* 2020.
 67. Haugland, R.A., et al., *Evaluation of genetic markers from the 16S rRNA gene V2 region for use in quantitative detection of selected Bacteroidales species and human fecal waste by qPCR.* Syst Appl Microbiol, 2010. **33**(6): p. 348-57.
 68. Shanks, O.C., et al., *Quantitative PCR for genetic markers of human fecal pollution.* Appl Environ Microbiol, 2009. **75**(17): p. 5507-13.
 69. Reischer, G.H., et al., *A quantitative real-time PCR assay for the highly sensitive and specific detection of human faecal influence in spring water from a large alpine catchment area.* Lett Appl Microbiol, 2007. **44**(4): p. 351-6.
 70. Yampara-Iquise, H., et al., *Use of a Bacteroides thetaiotaomicron-specific alpha-1-6, mannanase quantitative PCR to detect human faecal pollution in water.* J Appl Microbiol, 2008. **105**(5): p. 1686-93.

Table 1. Critical stages involved in the WBE of SARS-CoV-2 and key points to be considered. Some of these important factors had ever been summarized in a previous review [51].

Critical stages	Key points to be considered for qPCR relative & absolute quantification
Virus shedding/Biomarker emission	<ul style="list-style-type: none"> ➤ Individual differences in shedding amount of virus/biomarker ➤ Fluctuating amount of virus shedding throughout the infection period ➤ Personal excreta amount ➤ Personal wastewater generation rate (fluctuating with the seasons) **
Hydraulic transport of virus/biomarker in the sewer system	<ul style="list-style-type: none"> ➤ Virus/Biomarker decay kinetics ➤ Temperature ➤ Content of wastewater ➤ Hydraulic retention time (HRT) ➤ Dilution effect of rainfall events**
Sampling	<ul style="list-style-type: none"> ➤ Metadata of sampling (Catchment area/population, flow, weather conditions) ➤ Sampling strategy (Specific manholes or/and the influent of WWTP) ➤ Sampling method (Autosampler for 24 h composite sample, volume of sample) ➤ Sample transportation (4 °C) and storage temperature (-80 °C)
Concentration (Virus/Biomarker recovery)	<ul style="list-style-type: none"> ➤ Timely treatment of samples (within 24 h) ➤ Thermal sample pre-treatment (disinfection for biosafety) ➤ Volume of sample for virus concentration ## ➤ Concentration method ➤ Surrogate control and viral recovery yield ➤ Adsorbed virions on the solid fraction
RNA/DNA extraction	<ul style="list-style-type: none"> ➤ RNA/DNA extraction kit## ➤ Final eluted RNA/DNA volume## ➤ Internal standard (reference biomarker)* ➤ Final RNA/DNA concentration## ➤ Inhibiting substances
RT-qPCR & sequencing	<ul style="list-style-type: none"> ➤ Type of qPCR (nonspecific fluorescent dyes or specific probe) ➤ qPCR instrument** ➤ Target and reference genes* ➤ Reaction mixture and thermal cycling parameters ➤ Data calibration and/or normalization ➤ Standard substance for absolute quantification** ➤ Sequencing the PCR products for polymorphism analysis and virus typing
Reporting the results	<ul style="list-style-type: none"> ➤ Type of reported values (relative abundance of viral genes, C_T values or gene copies/mL wastewater) ➤ Estimating the community epidemic burden (infection rate) ➤ Uncertainty of quantification ➤ Validation and efficiency of the method ➤ Estimating the daily distributed lag of the detection [14]

*: To realize the relative quantification of the target virus (gene), a reference biomarker (gene) whose content is closely related to population size need to be chosen.

** : The important factors that may **only** affect the absolute quantification, such as:

- the seasonal variation of personal sewage generation rate,
- the dilution effect of rainfall events on the combination sewage system,
- the standard substance (virus or gene) for constructing the calibration curve in absolute quantification.

##: The points are critical for absolute quantification but not so critical for relative quantification, such as:

- the volume of sample for virus concentration (need to be recorded for absolute quantification),
- RNA/DNA extraction kit,
- final eluted RNA/DNA volume at the step of nucleic acid extraction,
- final RNA/DNA concentration in the extracts.

Table 2. Reported detections of SARS-CoV-2 RNA in wastewater in different countries.

Country	City/location	Sample type	Virus concentration method	Target gene	Threshold cycle number (C_T) or detection rate (n)	Virus concentration (gene copies/L)	Cal. PSIP ^a (infection rate)	Reference
Australia	Southeast Queensland	Raw wastewater	A: direct RNA extraction from electronegative membranes; B: centrifugation (4,750 g × 30 mins) + ultrafiltration (Centricon® Plus-70 centrifugal filter, 10 kDa).	N	37.5-39 (C_T)	1.9×10 - 1.2×10^2	552-3571 (0.028–0.181%, 171-1,090 infected cases in the catchment)	[18]
China	First Affiliated Hospital of Zhejiang University	Inlets of preprocessing disinfection pool	NA	NA	29.37- 32.42 (C_T) 100% (n=3)	NA	NA	[45] ^b
		Outlet of preprocessing disinfection pool			33.55 (C_T) 100% (n=1)			
		Final outlet of sewage disinfection pool			0% (n=1)			
	Wuchang Cabin Hospital	Influent and effluent samples of septic tanks	Centrifugation (3,000 rpm) + PEG precipitation (PEG 6000 10%, NaCl 0.3 M) + centrifugation (10,000 g × 30 min)	ORF-1 N	NA	$(0.5-18.7) \times 10^3$	NA	[13]
Czech Republic	33 WWTPs	Untreated wastewater (24-h composite)	Direct flocculation (10 mL of beef extract solution in glycine buffer, 0.05 M +1 or 0.1 M HCl to pH 3.5-3.0)	NA	34-40 (C_T) 11.6% positive (n=112)	NA	391-4167 (24-256 per 100,000)	[52]
Ecuador	Quito	Quito's river water (polluted by wastewater)	Skimmed milk flocculation + centrifugation (8,000 g × 40 min)	N1	NA	2.84×10^5 - 3.19×10^6	NA (358-2077 cumulative cases in the influence area)	[53]
				N2	NA	2.07×10^5 - 2.23×10^6		
France	Parisian area	Raw wastewater from three WWTP	Ultracentrifugation (200,000 g × 1 h) **Virus in both liquid and solid phase should have been recovered.	E	NA	5×10^4 - 3×10^6	131,868 (91 confirmed cases in >12,000,000 inhabitants)	[44] ^c
Germany	North-Rhine Westphalia	Influent wastewater after the sand trap	Centrifugation (4,700 g × 30 min) + centrifugal ultrafiltration (Amicon® Ultra-15 Centrifugal Filter Unit, Sigma) **The authors reminding that quantifications must not neglect the solid-phase reservoir.	RdRP E M N	33-37 (RdRP, C_T) 32-35 (M, C_T)	3.0×10^3 - 2.0×10^4 2.7×10^3 - 3.7×10^4	454-1389 (72-220 cumulative cases per 100,000, and 30-174 acute cases per 100,000)	[17]
		Treated sewage at selected locations						
India	Ahmedabad	WWTP influent raw wastewater	Centrifugation (4,500 g × 30 min) + filtration (0.22 µm) + PEG precipitation (PEG 9000, 80 g/L and NaCl, 17.5 g/L) + Centrifugation (13,000 g × 90 min)	ORF1ab	32.65- 35.52 (C_T)	$\sim 3.5 \times 10^2$	522-1135.6 (4912-10674 in Ahmedabad)	[16] ^b
				N	34.18- 35.39 (C_T)			
				S	34.83- 39.56 (C_T)			
		Final effluents after UASB and aeration pond		ORF1ab, N, S	Negative	Negative		
	Jaipur	Raw wastewater form 6 WWTP and 2 hospitals	A: Filtration (0.45 µm) + PEG precipitation (0.9 g NaCl and 4 g PEG in 50 mL) + centrifugation (7,000 rpm × 30 min); B: Centrifugation (1 mL sample, 7,000 rpm ×	RdRP	36.04 (C_T)	NA	NA	[54]
				E	33.16 (C_T)	NA		
				N	33.06 (C_T)	NA		

			15 min) + RNA extraction from supernatant					
Israel	WWTP in Israel, Tel Aviv metropolis	Raw sewage	Centrifugation + PEG or alum (20 mg/L) precipitation + centrifugation (14,000 g × 30 min, 4°C) + filtration (0.22 µm) + ultrafiltration (Ultra-15 centrifugal tubes, 30 kDa)	E	32.76 -38.5 (C_T)	NA	NA (1235-2052 cases in different locations)	[55]
Italy	Milan and Rome	Raw influent sewage	Centrifugation + PEG/dextran precipitation **Concentrate was combined with the pellet from the initial centrifugation.	RdRP	50% positive (6/12)	NA	108,275 (29 cases in Milan)	[19] ^c
	Milano Metropolitan Area	Raw and treated wastewater, river water	Sequential filtration (0.7 µm + 0.22 µm) filtrate for RNA extraction	ORF1ab N E	100% positive for raw wastewater from all three WWTP 100% negative for treated wastewater Positive in one river water sample	NA	10,000 (up to 0.02%)	[15] ^b
	Milan Turin Bologna	Influent wastewater **Samples collected between 9th Oct. 2019 and 28th Feb. 2020.	Centrifugation (1,200 g × 30 min) + PEG precipitation (dextran ~1%, PEG 6000, 9.66%, NaCl 0.2 N)	RdRP ORF1ab	37.5% positive (15/40)	Up to 5.6×10 ⁴	NA	[56]
Japan	Yamanashi Prefecture	Influent wastewater	Electronegative membrane-vortex (EMV)	N	Negative	NA	22727 (4.4 cumulative cases per 100,000 inhabitants)	[23]
			Membrane adsorption-direct RNA extraction	N1	Negative	NA		
		Secondary-treated wastewater	Electronegative membrane-vortex (EMV)	N2	20% Positive (n=5)	2.4×10 ³		
			Membrane adsorption-direct RNA extraction		Negative	NA		
	Ishikawa & Toyama	Influent wastewater of five WWTP	Centrifugation (3,000 g × 5 min) + PEG precipitation (PEG 10%, NaCl 1 M) + Centrifugation (1,0000 g×30 min)	N2 & N3	46.7% positive (21/45, n=45)	1.2×10 ⁴ -3.5×10 ⁴	5000-333,333 (0.3-20 in 100,000)	[57]
Netherlands	Amsterdam, The Hague, Utrecht, Apeldoorn, Amersfoort, Schiphol, Tilburg	Raw wastewater	Centrifugation (4654 g × 30 min) + centrifugal ultrafiltration (Centricon® Plus-70, 100 kDa)	N1	NA	2.6 × 10 ³ - 7.9 × 10 ⁵	~ 100,000 (around or even below 1.0 case in 100,000 people can be detected)	[22]
				N2	NA	1.2 × 10 ⁴ - 2.2 × 10 ⁶		
				N3	NA	6.6 × 10 ³ - 1.8 × 10 ⁶		
				E	29.9- 39.2 (C_T)	NA		
Slovenia	Ljubljana (Department of Infectious Diseases, University Medical Center Ljubljana catchment area)	Untreated wastewater (pumping station)	Filtration (0.7µm) + ultrafiltration (30 kDa and 10 kDa) **The better results of 10-kDa filter are listed.	RdRP	66.7% positive (n=15) 29.65-36.81 (C_T)	~10 ⁷	Up to 4 hospitalized patients	[58]
				E	33.61-38.12 (C_T)	~10 ⁵		
Spain	Main cities of the Region of Murcia	Influent wastewater	Aluminum hydroxide adsorption-precipitation	N1	78.5% (n=42) 34-37.05 (C_T)	5.1 log10	714-11,806 (8.47-140 diagnosed cases per 100,000 inhabitants)	[20] ^b
				N2	73.8% (n=42) 35-38.12 (C_T)	5.5 log10		
				N3	85.7% (n=42) 35-37.29 (C_T)	5.5 log10		
		Secondary and tertiary effluent wastewater		N1, N2, N3	2/18 positive (secondary), 0/12 positive (tertiary)	NA		

	Ourense	Wastewater (24-h composite)	Centrifugation (4,600 g × 45 min) + ultrafiltration (Amicon 15 ml 10 kDa centrifugal devices)	RdRP	34.7- 37.57 (C_T)	7.5×10 ³ -1.5 × 10 ⁴	237-308 (324.4-422.4 cases per 100,000 inhabitants)	[43]
				N	36.39 -39.60 (C_T)			
				E	33.61- 34.87 (C_T)			
		Sludge	Glycin buffer (0.05 M glycine, 3% beef extract) detach virus bound + centrifugation (8,000 g × 30 min) + filtration (0.45 μm) + PEG precipitation + centrifugation (13,000 × 45 min)	RdRP	34.77-38.08 (C_T)	7.5×10 ³ -4.0 × 10 ⁴		
				N	35.78-39.87 (C_T)			
				E	33.41-36.13 (C_T)			
Turkey	Istanbul	Primary sludge	Shaking for detaching virus + centrifugation (7,471 g × 45 min) + filtration (0.45 μm & 0.2 μm) + PEG precipitation (PEG 8000 10%, NaCl, 0.3 M) + centrifugation (7,471 g × 120 min)	RdRP	34.71-35.86 (C_T)	1.25×10 ⁴ - 2.33×10 ⁴	NA	[59]
		Waste activated sludge			33.52- 35.67 (C_T)	1.17×10 ⁴ - 4.02×10 ⁴		
	Istanbul	Wastewater at influent of WWTP or manhole	A: Centrifugation (3,200 g × 45 min) + ultrafiltration (Amicon® Ultra-15, 10 kDa); B: Centrifugation (3,000 g × 45 min) + filtration (0.45 μm) + PEG precipitation	RdRP	34.667-39.54 (C_T)	2.89×10 ³ -9.33×10 ⁴	NA (284-6045 cases around the district)	[60]
	USA	Massachusetts	Raw sewage (24-h composite)	I: Filtration (0.2 μm) + PEG precipitation (PEG 8000 10%, NaCl 0.3 M) + centrifugation (12,000 g × 2h)	N1	33.87- 37.86 (C_T)	5.7 × 10 ⁴ -3.03 × 10 ⁵	~3846 (0.026% confirmed rate in the state of Massachusetts)
N2					38.39 (C_T)	2.1 × 10⁴-5.06 × 10⁵ (PMMoV normalization)		
N3					37.12- 38.14 (C_T)			
Bozeman, Montana		Raw influent wastewater (composite samples)	Sequential filtration (20 μm, 5 μm & 0.45μm) + Corning Spin-X ultrafiltration (100 kDa)	N1	NA	43.5-528.9	4152 (up to 12 positive daily tests within 49,831 inhabitants)	[62]
				N2	NA	22.2-1710.8		
New Haven, Connecticut		Primary sludge	Direct RNA extraction	N1	~38.75 (C_T)	1.7 × 10 ⁶ -4.6 × 10 ⁸	50-28571 (7-3,978 cases per 200,000)	[14] ^d
				N2				
Southern Louisiana		Raw wastewater	A: Centrifugation (3,000 g × 30 min) + ultrafiltration (Centricon® Plus-70 centrifugal filter, 100 kDa, 1,500 g × 15 min)	N1	NA	7.5 x 10 ³	NA (308-6173 cumulative confirmed COVID-19 cases in related area)	[21]
				N2	NA	3.1 × 10 ³ -4.3 × 10 ³		
			B: Adsorption–elution by electronegative Membrane (25 mM MgCl ₂ as salt bridge)	N1	NA	ND		
	N2			NA	ND			

^a: Calculated detection sensitivity (PSIP) for each study is shown in bold texts.

^b: Those studies detected the SARS-CoV-2 RNA in the influent of the WWTP but not in the effluent, which indicated the effective removal of virus *via* sewage treatment.

^c: Studies recovering virions from both liquid and solid phases of sewage samples gave high detection sensitivities (high PSIP values).

^d: Directly extracting RNA from **primary sludge** gave the highest detected concentration of virus RNA up to 4.6×10^8 copies/mL.

^e: **PMMoV normalization**: This study used **pepper mild mottle virus** (PMMoV) as an internal reference for quantification.

UASB: Upflow Anaerobic Sludge Blanket

NA: Not available.

ND: Not detected.

Table 3. Recommended primer/probe sets with high specificity and inclusivity according the sequence alignment analysis.

Institutes	Target	Name	Sequence (5'-3')	Tm (°C)	Product Size (bp)	Reference
France Pasteur	<u>RdRP-IP4</u>	RdRp-IP4-KY_F*	GGT <u>ARC</u> TGG TAT GAT TTC G	58	107	[63]
		RdRp-IP4_P	FAM-TCA TAC AAA CCA CGC CAG G-BHQ1			
		RdRp-IP4_R	CTG GTC AAG GTT AAT ATA GG			
Germany Charité	<u>RdRP</u>	RdRp_SARSr-KY-F*	GTG A <u>AA</u> TGG TCA TGT GTG GCG G	58	100	[64]
		RdRp_SARSr-P2	FAM-CAG GTG GAA CCT CAT CAG GAG ATG C-BHQ1			
		RdRp_SARSr-KY-R*	CA <u>A</u> ATG TTA AA <u>A</u> ACA CTA TTA GCA TA			
HKU SKL of EID	<u>S</u>	Chan-S_F	CCT ACT AAA TTA AAT GAT CTC TGC TTT ACT	55	158	[65]
		Chan-S_P	FAM-CGC TCC AGG GCA AAC TGG AAA G-BHQ1			
		Chan-S-KY_R*	CAA GCT ATA AC <u>R</u> CAG CCT GTA			
US CDC	N	2019-nCoV_N2-F	TTA CAA ACA TTG GCC GCA AA	55	67	[66]
		2019-nCoV_N2-P	FAM-ACA ATT TGC CCC CAG CGC TTC AG-BHQ1			
		2019-nCoV_N2-R	GCG CGA CAT TCC GAA GAA			

*The sequences of the primers marked in bold have been amended at the underlined sites according to the sequence alignment and/or Primer-BLAST analysis, and the primers have been renamed accordingly.

Table 4. Human-associate biomarkers in MST assays as promising candidates of the references for the relative quantification of SARS-CoV-2.

Organism	Assays	Primers and Probes	Tm (°C)	Product Size (bp)	Sensitivity (n)	Specificity (n)	References
<i>Bacteroidales</i>	HF183	HF183-1: ATCATGAGTTCACATGTCCG BFDRRev: CGTAGGAGTTTGGACCGTGT BFDFAM: 6-FAM-CTGAGAGGAAGGTCCCCCACATTGGA-TAMRA	60	167	100% (16)	90.9% (174)	[48, 67]
	HumM2	Hum2F: CGTCAGGTTTGTTCGGTATTG Hum2R: TCATCACGTAACCTATTATATGCATTAGC HumP: FAM-TATCGAAAATCTCACGGATTAACCTTGTGTACGC-TAMRA	60	101	100% (16 feces, 20 SEW)	99.2% (285)	[68]
	HumM3	Hum3F: GTAATTCGCGTTCTTCCTCACAT Hum3R: GGAGGAAACAAGTATGAAGATAGAAGAATTAA Hum3P: FAM-AGGTCTGTCCTTCGAAATAGCGGT-TAMRA	60	83	100% (16 feces, 20 SEW)	97.2% (285)	[68]
	BacH	BacH-f: CTTGGCCAGCCTTCTGAAAG BacH-r: CCCCATCGTCTACCGAAAATAC BacH-pC: FAM-TCATGATCCCATCCTG-NFQ-MGB BacH-pT: FAM-TCATGATGCCATCTTG-NFQ-MGB	61	93	95% feces (21), 100% wastewater (20)	99% (302)	[69]
	BtH	BtH-F: CATCGTTCGTCAGCAGTAACA	60	63	100% (10 feces, 20 SEW)	100% (160)	[70]
	(<i>B. thetaiotomicron</i> <i>α-mannanase</i>)	BtH-R: CCAAGAAAAAGGGACAGTGG BtH-P: FAM-ACCTGCTG-NFQ					
	CPQ_056	056F1: CAGAAGTACAACTCCTAAAAAACGTAGAG 056R1: GATGACCAATAAACAAGCCATTAGC 056P1: FAM-AATAACGATTTACGTGATGTAAC-MGB	60	126	100% (9 SEW)	98.6% (222)	[33]
	CPQ_064	064F1: TGTATAGATGCTGCTGCAACTGTACTC 064R1: CGTTGTTTTCATCTTTATCTTGTCCAT 064P1: FAM-CTGAAATTGTTTCATAAGCAA-MGB	60	148	100% (9 SEW)	98.6% (222)	[33]
	Mitochondria	mtCytb					
	(Cytochrome b)	Hum_mtCytb-F: AGTCCCACCCTCACACGATTCTTT Hum_mtCytb-R: AGTAAGCCGAGGGCGTCTTTGATT Hum_mtCytb-P: FAM-ACCCTTCATTATTGCAGCCCTAGCAGCACT- TAMRA	60	185	100% (2 feces, 10 SEW)	100% (18)	[47]

Figure 1. The concept of wastewater-based epidemiology (WBE). The figure is amended from reference [28]. The detection targets of WBE include drugs, population markers, PCPP, health-related biomarkers and industrial chemicals etc. This work focuses on the quantitative pathogen surveillance and the estimation of community infection rate.

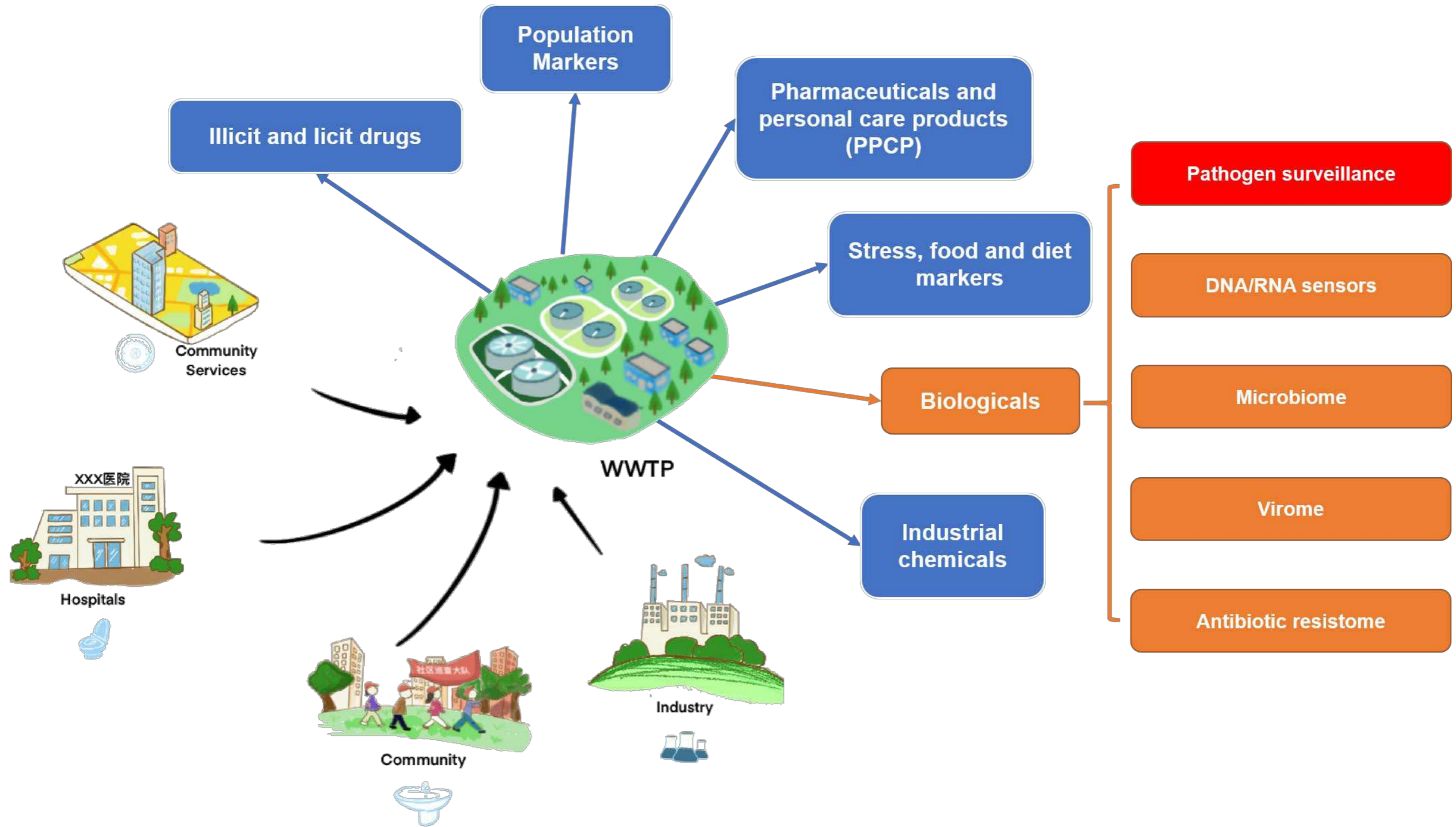


Figure 3. Factors affecting the detectability of SARS-CoV-2 in sewage. The detection sensitivity (PSIP) is affected by personal virus shedding rate (S_r) and sewage generation rate (V_s) (A). The detection sensitivity (PSIP) is also affected by factors that influence the decay kinetics of virus in sewage (B). The decay rate constant (k_T) at different temperatures (T) are calculated according to the reported experimental data (C). The calculated values of T_{50} and T_{90} are compared with experimental ones (D). The experimental data of the decay kinetics of SARS-CoV-2 RNA are available in reference [39].

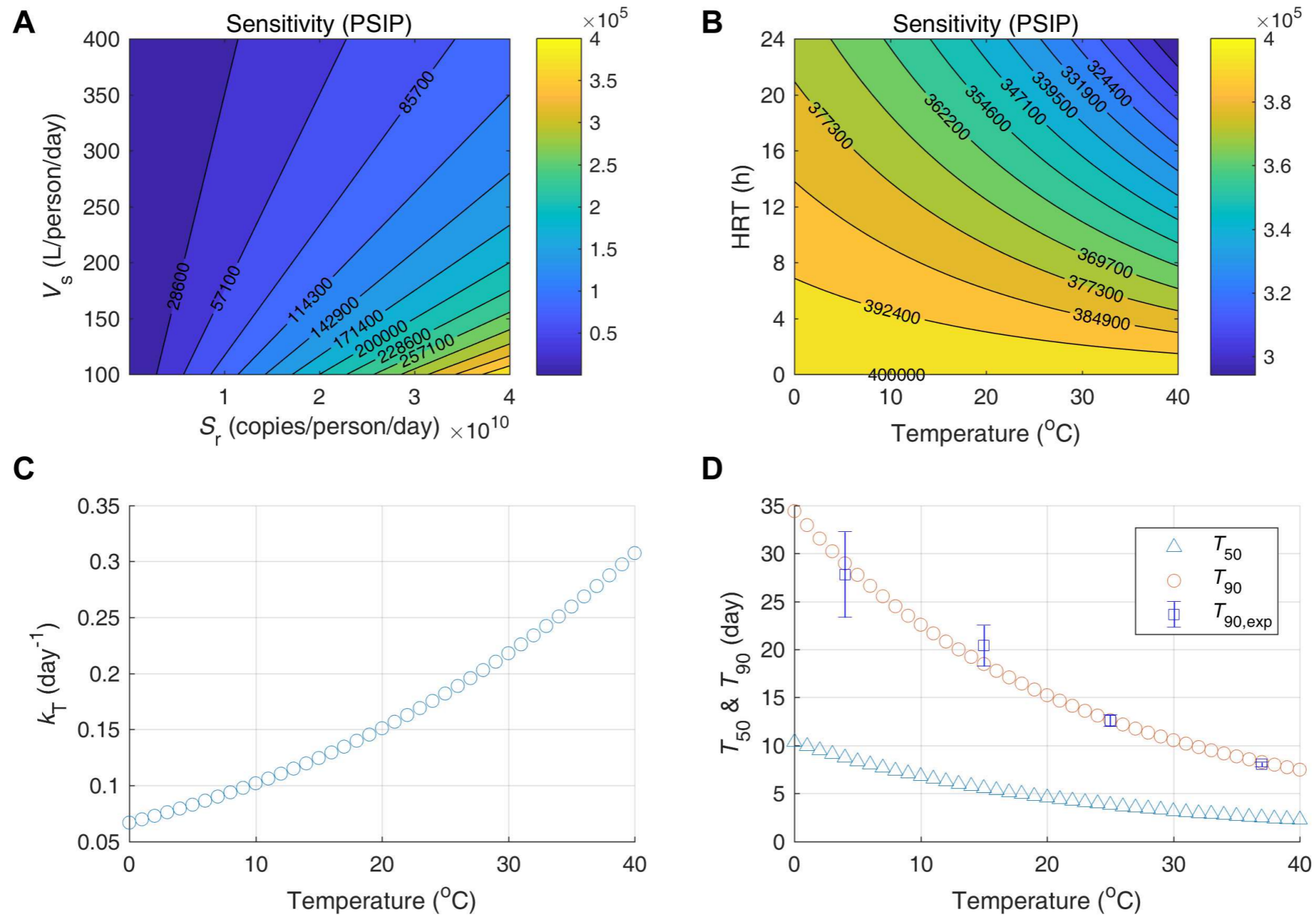
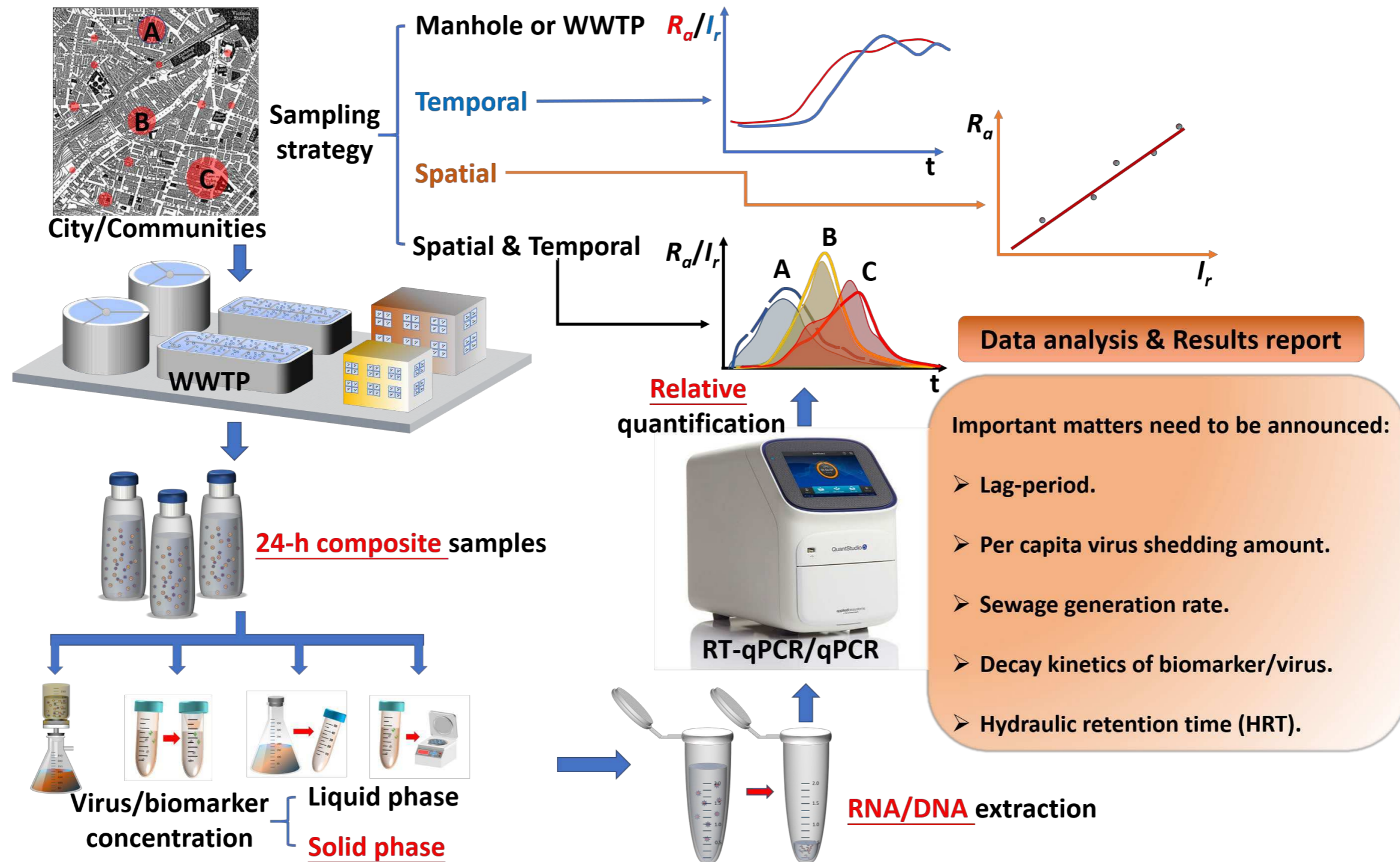


Figure 4. Illustration of the technical framework of the WBE for SARS-CoV-2 based on relative quantification *via* qPCR. Different strategies can be adopted at the sampling step: temporal, spatial and combination of spatial and temporal. The 24-hour flow-weighted-composite samples are recommended for testing. It is recommended to recover virus from both solid and liquid phase and to extract both RNA and DNA from concentrated samples. When estimating the community infection rate, the relative abundance of virus RNA (against the reference gene) in the sewage is adopted.



Key points in the figure:

- Spatial or temporal sampling (**24-h** flow-weighted **composite** samples),
- Concentration of virus from both liquid and **solid** phases of the sewage samples,
- Simultaneous **DNA/RNA** extraction,
- RT-qPCR/qPCR **relative quantification**,
- **Important matters** need to be announced when reporting the results: lag-period between WBE and clinical data, personal virus shedding amount, sewage generation rate, decay kinetics of virus (RNA) and biomarkers (reference genes) in sewage, and HRT.

Figures

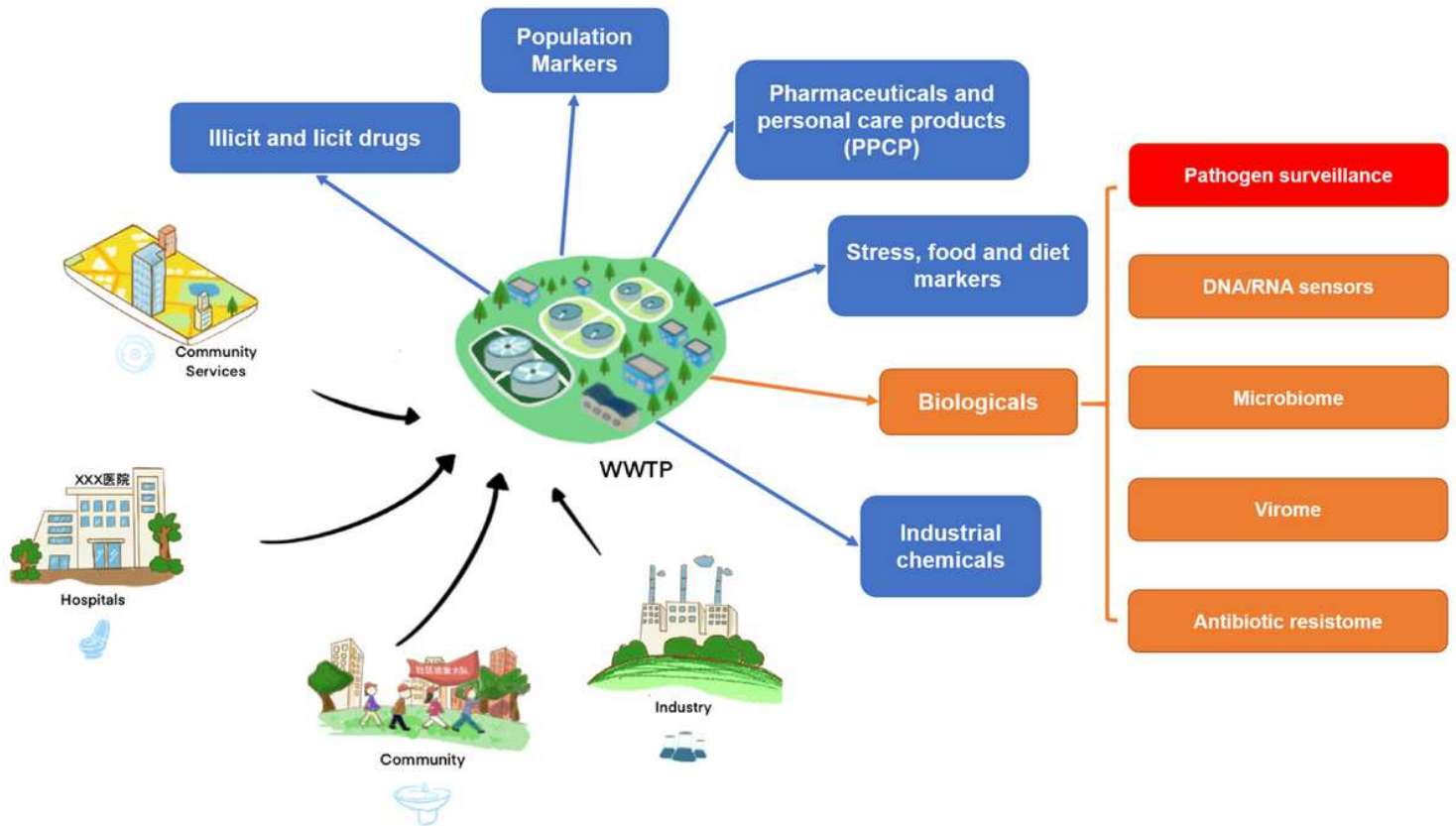


Figure 1

The concept of wastewater-based epidemiology (WBE). The figure is amended from reference [28]. The detection targets of WBE include drugs, population markers, PPCP, health-related biomarkers and industrial chemicals etc. This work focuses on the quantitative pathogen surveillance and the estimation of community infection rate.

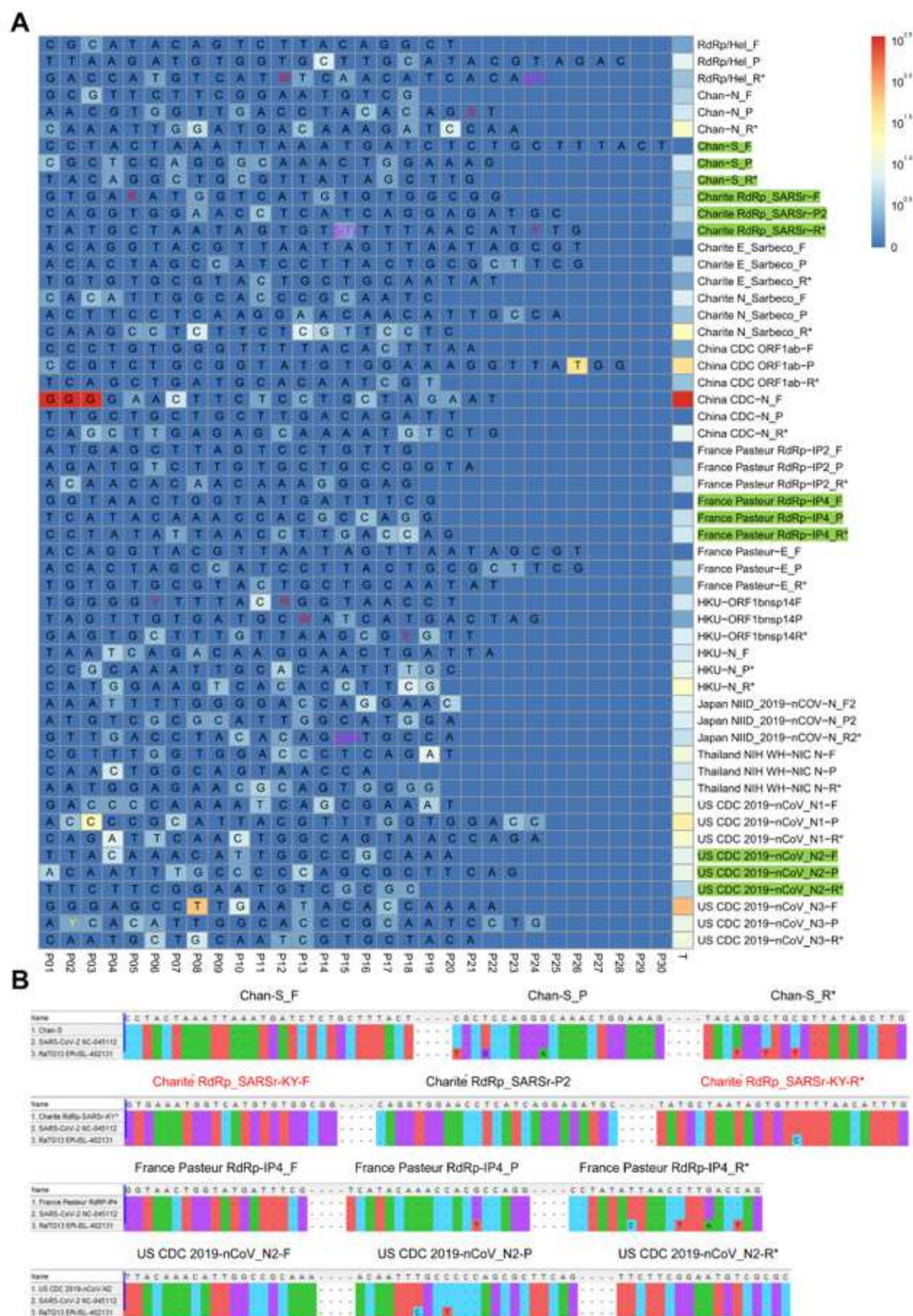


Figure 2

Heatmap indicating the mutation rate of SARS-CoV-2 genome at the probe/primer targeting regions (A) and the specificity (exclusivity) evaluation of recommended primer/probe sets by matching their sequences with the genome of the bat coronavirus RaTG13 (B). The color of each square in the heatmap (A) represents the number of sequences (among the total 2257 sequences) that mutated at the site, which is displayed in an exponential gradient. The sequence of each primer/probe is exhibited

accordingly. Magenta texts indicate mismatched sites of the primer/probe to the PCR template and the expected matched base is listed in the bracket. The primers/probes indicated with asterisk in both panels (A) and (B) are exhibited the reverse complementary sequences. Most of them are reverse primers. Those primer/probe sets recommended in this work are marked in green in panel (A), among which the sequences of Charité RdRP-SARSr primer/probe set has been amended (indicated with red texts in panel (B)).

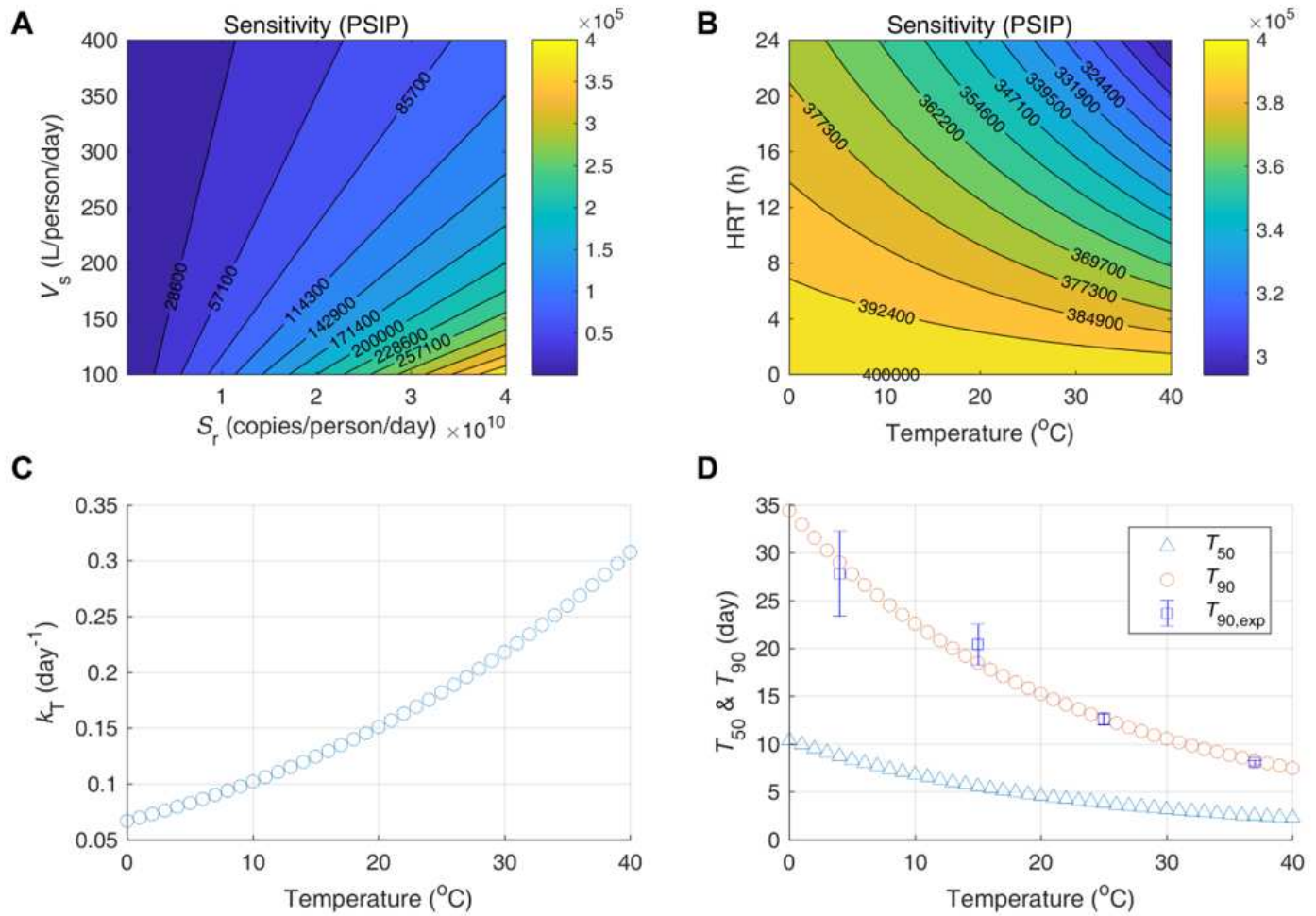


Figure 3

Factors affecting the detectability of SARS-CoV-2 in sewage. The detection sensitivity (PSIP) is affected by personal virus shedding rate (S_r) and sewage generation rate (V_s) (A). The detection sensitivity (PSIP) is also affected by factors that influence the decay kinetics of virus in sewage (B). The decay rate constant (k_T) at different temperatures (T) are calculated according to the reported experimental data (C). The calculated values of T_{50} and T_{90} are compared with experimental ones (D). The experimental data of the decay kinetics of SARS-CoV-2 RNA are available in reference [39].

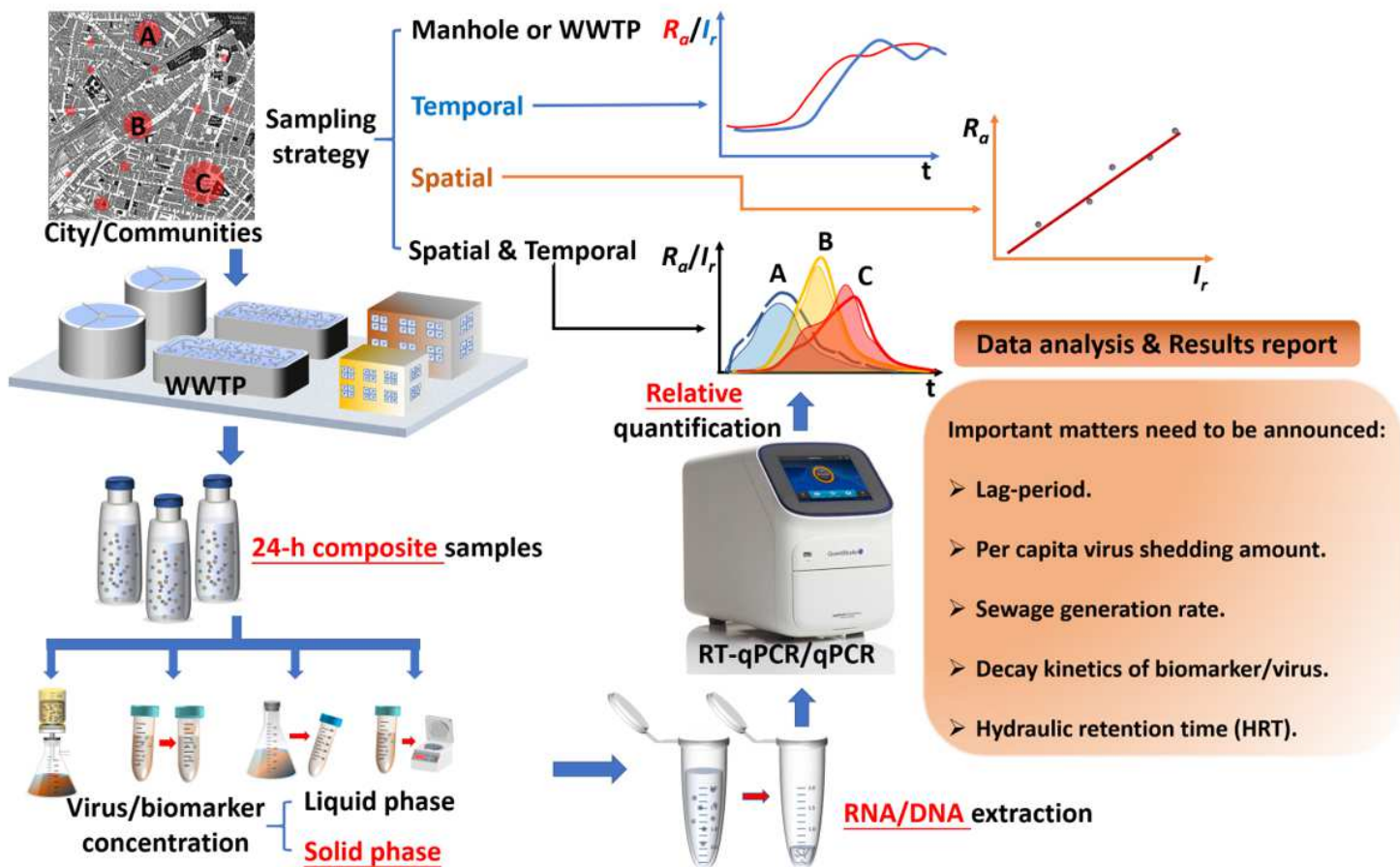


Figure 4

Illustration of the technical framework of the WBE for SARS-CoV-2 based on relative quantification via qPCR. Different strategies can be adopted at the sampling step: temporal, spatial and combination of spatial and temporal. The 24-hour flow-weighted-composite samples are recommended for testing. It is recommended to recover virus from both solid and liquid phase and to extract both RNA and DNA from concentrated samples. When estimating the community infection rate, the relative abundance of virus RNA (against the reference gene) in the sewage is adopted.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [WBEforSARSCoV2Supportinginformation.docx](#)
- [SupplementaryDatasets.xlsx](#)