ICS 07.100
CCS A 40

Social Organization Standard

T/GBA 021—2024

General technical requirements for monitoring SARS-CoV-2 variants from wastewater based on high-throughput sequencing technology

基于高通量测序技术的污水来源新型冠状病毒变异监测通用技术要求

(English Translation)

Issue date:2024-12-31 Implementation date:2025-01-10

Issued by Guangdong-Hong Kong-Macao Greater Bay Area Standards Innovation Alliance

Contents

Foreword

This document is drafted in accordance with the rules set forth in the GB/T 1.1-2020 *Directives for Standardization -- Part 1: Rules for the Structure and Drafting of Standardizing Documents*.

This document was proposed by the Shenzhen Center for Disease Control and Prevention.

This document was prepared by the Guangdong-Hong Kong-Macao Greater Bay Area Standards Innovation Alliance.

This document is authorized for use by organizational partners and all members of the Guangdong-Hong Kong-Macao Greater Bay Area Standards Innovation Alliance, who are required to adopt the same transformation into their own group standards, and publicize the standard basic information on the National Group Standards Information Platform.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. The issuing body of this document shall not be held responsible for identifying any or all such patent rights.

This is the first release.

General technical requirements for monitoring SARS-CoV-2 variants from wastewater based on high-throughput sequencing technology

1 Scope

This document specifies the general technical requirements for monitoring SARS-CoV-2 variants in wastewater samples based on high-throughput sequencing technology, including the principles, experimental procedures, and quality control requirements.

This document is applicable to the processes of sample concentration, nucleic acid extraction, high-throughput sequencing, and data analysis for monitoring SARS-CoV-2 variants in wastewater samples. And high-throughput sequencing is applicable to sequencing platforms based on technologies such as reversible terminator sequencing and combinatorial probe-anchor synthesis technology.

This document does not apply to:
———First-generation gene sequencing methods based on Sanger sequencing as the primary technology;
———Single-molecule gene sequencing technologies characterized by continuous sequencing, such as single-molecule real-time fluorescence sequencing, single-molecule nanopore strand sequencing, and single-molecule nanopore tag sequencing;
———Metagenomic sequencing and probe-based target sequencing.

2 Normative references

The contents of the following documents constitute essential provisions in this document. For dated references, only the version specified applies. For undated references, the latest version, including all amendments, is applicable.

GB 19489 General requirements for biosafety in laboratories
GB/T 30989—2014 Technical regulation of high-throughput gene sequencing
WS/T 799—2022 Method for enrichment and nucleic acid detection of SARS-CoV-2 in sewage
YY/T 1723—2020 High-throughput gene sequencer

3 Terms and definitions

The following terms and definitions shall apply to this document.

3.1

wastewater

The general term of wastewater discharged into the wastewater pipe network system is mainly derived from residential areas, schools, hospitals, commercial service agencies and various public facilities.

3.2

Severe Acute Respiratory Syndrome Coronavirus 2

SARS-CoV-2 belongs to the β genus coronavirus, with a genome consisting of linear single-stranded positive-sense RNA approximately 30 Kb in length. It is an enveloped virus, characterized by spherical or oval-shaped particles with diameters ranging from 60 nm to 140 nm.
[Source: WS/T 799--2022, Terms and Definitions 3.1]

3.3

enrichment and concentration

Refers to the process of extracting and concentrating target pathogenic microorganisms from wastewater using specific technical methods, with the aim of increasing their concentration in the sample to facilitate subsequent detection and analysis.

3.4

threshold cycle

Refers to the number of amplification cycles required for the fluorescence signal to reach the predefined threshold in a real-time fluorescence quantitative PCR reaction system.
[Source: WS/T 799--2022, Terms and Definitions 3.2]

3.5

the primer pool

Refers to a mixture of multiple primers used to achieve simultaneous amplification of multiple target regions.

3.6

multiple PCR

Refers to a PCR reaction in which multiple nucleic acid fragments are simultaneously amplified in the same reaction system.
[Source: GB/T 19915.5-2005, Term and Definitions 3.2]

3.7

multiplex PCR amplicon sequencing

Refers to a technology that uses multiplex PCR to simultaneously amplify multiple target gene regions (amplicons), in conjunction with high-throughput sequencing platforms for sequencing and analysis of the amplicons.

3.8

gene sequencing

Refers to the determination of different types of nucleotide bases in nucleic acid molecules, specifically the sequencing of the arrangement of adenine (A), guanine (G), cytosine (C), and thymine (T) or uracil (U) in nucleic acids.
[Source: YY/T 1723-2020, Terms and Definitions 3.1, modified]

3.9

sequencing library

Refers to a collection of nucleic acid fragments used as templates for sequencing. These fragments have a specific size range and typically include: adapters and/or specific sequences for sequencing primer binding, sequence capture, and/or identifiers for recognizing specific regions.
[Source: ISO 20397-1:2022, Terms and Definitions 3.5, modified]

3.10

read length of gene sequencing

Refers to the length of high-quality sequence fragments that can be read in a single run, typically expressed in the number of bases.
[Source: YY/T 1723-2020, Term and Definitions 3.4]

3.11

reads

Refers to sequence fragments containing base sequences and quality scores generated by high-throughput sequencing platforms. A single sequence fragment generated from either single-end or paired-end sequencing is counted as one read.
[Source: GB/T 35890-2018, Terms and Definitions 3.2, modified]

3.12

quality of base calling

Refers to the probability of accurately identifying a base, abbreviated as Q. It is typically expressed numerically as: $Q = -10\log_{10}(P)$, where P is the error probability of base calling.

Note 1: The base calling quality value is inversely correlated with the base calling error rate, following a logarithmic function.

Note 2: Higher base calling quality values indicate lower error rates. For instance, Q20 refers to a base calling accuracy of 99% (error rate of 1%), while Q30 refers to a base calling accuracy of 99.9% (error rate of 0.1%).

3.13

coverage rate of sequencing

Refers to the proportion of nucleotides sequence detected in the sample covered on the reference sequence, calculated as:
Sequencing Coverage Rate= [Length of Sequenced Region ÷ Total Length of Reference Sequence]
[Source: GB/T 30989-2014, Terms and Definitions 3.30]

3.14

depth of sequencing

Refers to the number of times a specific nucleotide is detected in the test sample.
[Source: GB/T 30989-2014, Terms and Definitions 3.31]

3.15

viral variants

Generally refer to groups of viruses that have mutations in their genomes, resulting in significant differences from the original strain in terms of transmissibility, pathogenicity, or immune evasion.

Note: SARS-CoV-2 variants are typically classified and assigned lineage numbers by the Pangolin tool based on genomic mutations, which classifies SARS-CoV-2 into different lineages to help track the evolution and transmission of the virus.

3.16

single nucleotide variants

Refers to single nucleotide (A, T, C, or G) substitution that occurs at different locations in the genome, and the frequency of this variant is not limited by the frequency of detection within the population and covers all single differences.

3.17

mapping quality

Refers to the values output by the alignment software that indicate the confidence in the position of the sequence alignment when the sequencing data is aligned to the reference genome.
Note: As mapping quality≥20 indicates that the probability of sequence misalignment is ≤ 1%.

3.18

SNV mutation frequency

Refers to the proportion of a nucleotide mutation at a specific site in the tested sample relative to all sequencing data covering that site.

3.19

abundance of variants

Refers to the relative abundance of a SARS-CoV-2 variant among all variants in a given sample.

4 Abbreviation

The following abbreviations apply to this document.
Cycle threshold (CT)
Deoxyribonucleic acid(DNA)
Double-stranded DNA(dsDNA)
Kilobase pairs(Kb)
Mapping quality(MAPQ)
Megabyte reads(Mb)
Polymerase chain reaction(PCR)
Ribonucleic acid(RNA)
Reverse Transcription Quantitative PCR(RT-qPCR)
Single nucleotide variants(SNV)
Severe acute respiratory syndrome coronavirus 2(SARS-CoV-2)

5 Principle

Monitoring SARS-CoV-2 variants in wastewater based on high-throughput sequencing technology employs a multiplex PCR amplicon sequencing method to sequence the enriched RNA of the SARS-CoV-2 extracted from wastewater samples. Data analysis is conducted to obtain the abundance proportions of various variants of the SARS-CoV-2 in the wastewater. This process is primarily achieved through the following steps:

a) Sample collection and processing: setting up wastewater monitoring sites, collecting wastewater samples, and concentrating and enrich viral particles and nucleic acid fragments to facilitate subsequent detection and sequencing;

b) Nucleic acid extraction and detection: extracting viral RNA from the concentrated samples and performing quantitative detection;

c) Library preparation and sequencing: reverse transcribe viral RNA into cDNA, constructing a multiplex PCR amplicon sequencing library, and perform sequencing on a high-throughput sequencing platform;

d) Data analysis: performing quality control, alignment and conducting variant detection of sequencing data to obtain the relative abundance of SARS-CoV-2 variants in wastewater.

6 Test procedure

6.1 Wastewater sample types

The sources of wastewater samples include but are not limited to urban wastewater treatment plants (water purification plants), wastewater pumping stations, medical institutions, aircraft, railway stations and other key transportation hubs, livestock and poultry farms, agricultural products wholesale markets and other places.

Sample pre-processing, preservation, disposal, and transportation procedures should comply with the requirements outlined in Annex 12, "Technical Guidelines for SARS-CoV-2 Specimen Collection and Testing," of the "Notice on Issuing the COVID-19 Prevention and Control Protocol (9th Edition)".

The experimental conditions and laboratory zoning requirements that need to be met during the experimental operations shall comply with the provisions specified in Appendix A. The necessary reagents, materials, instruments, and equipment requirements for experimental operations shall comply with the provisions specified in Appendix B.

6.2 Sampling Site Selection for Wastewater surveillance

Based on the specific objectives of the surveillance, wastewater monitoring sites should be strategically placed. Typically, these sites are established at the inlets of municipal wastewater treatment plants and at key junctions within the sewage network. Depending on the surveillance requirements, monitoring sites may also be set up in medical institutions, particularly in hospitals for infectious diseases (including general hospitals with dedicated infectious disease wards), to collect sewage from these areas. Additionally, monitoring sites can be installed in residential communities or urban villages where the drainage system is clearly defined, by placing them at sewage manholes. Key locations and community sampling points should be positioned at the nearest outfalls or inspection wells downstream of the monitored area. When conducting long-term surveillance for SARS-CoV-2 in wastewater, it is advisable to maintain a relative stability in the sampling sites. Prior to the establishment of monitoring sites, it is necessary to clarify the inflow and water source of the sewage network at the monitoring location, as well as to understand the population size within the area served by the sewage collection system.

6.3 Collection of wastewater samples

6.3.1 Collection equipment

Sampling equipment must have good chemical stability to avoid reactions with water samples that could alter component concentrations. Suitable materials include polyethylene, stainless steel, and polytetrafluoroethylen (PTFE). Sample containers can be rigid glass or PTFE plastic.

Reused containers should be disinfected, cleaned, and able to withstand a 60° C water bath without affecting testing.

6.3.2 Sampling method

6.3.2.1 Basic requirements

Water samples can be collected manually or with automated equipment. For long-term monitoring, wastewater autosamplers with refrigeration are recommended for precise timing and sample integrity. The sampling method should match monitoring objectives, with time-composite samples preferred. In early epidemic stages, establish community sampling points and collect 3-hour composite samples for early warning. During peak transmission, deploy sampling points in communities, pumping stations, and wastewater treatment plants, collecting 24-hour composite samples to enable assessment of infection trends and monitor the composition and abundance of viral variants.

After sample collection, the outer surface of the sample container should be disinfected with 75% alcohol on site and put into a sealed sample bag. The outer surface of the sealed sample bag should be disinfected again, and the sample should be stored at 0℃-4℃ and sent to the testing laboratory within 2 hours.

6.3.2.2 3-hour mixed sample collection

During the peak water consumption (e.g., 8 am-11 am), mixed samples were usually collected every 15 minutes. The volume of water samples should be set according to the monitoring requirement. It is recommended to collect no less than 250 mL each time, and no less than 3 L of water samples should be collected in total. After mixing, 250 mL is dispensed into the sample container for examination.

6.3.2.3 24-hour mixed sample collection

Wastewater was usually collected continuously from 0:00 to 24:00 every day at each sampling site, with no less than 125 mL collected every hour, resulting in no less than 3 L of wastewater collected per site per day. After mixing, 250 mL is dispensed into the sample container for examination.

6.3.3 Sampling frequency

Adjust the sampling frequency based on the intensity of SARS-CoV-2 transmission: the conventional sampling frequency should be at least 1-2 times/week, and recommended ≥3 times a week during peak period of the epidemic.

6.3.4 Sampling information collection

The collected information includes sampling point names, numbers, addresses, latitude and longitude, population data, and other details. Indicators recorded at each site include water temperature, suspended solids, pH, daily average water flow, chemical oxygen demand (COD), and ammonia nitrogen.

6.4 Enrichment and concentration of the wastewater samples

Samples should be stored at 0℃-4℃ after delivery and enriched within 24 hours. Before enrichment, mix samples thoroughly and incubate in a 60℃ water bath for 30 minutes to inactivate viruses. Follow the health industry standard “Method for Enrichment and Nucleic Acid Detection of SARS-CoV-2 in Sewage” (WS/T 799-2022) for enrichment and concentration processes, including polyethylene glycol precipitation, aluminum salt coagulation, and centrifugal ultrafiltration. The latest reference standard will prevail if updated.

6.5 Nucleic acid extraction and SARS-CoV-2 nucleic acid detection of wastewater samples

6.5.1 Nucleic acid extraction of wastewater samples

Select a virus nucleic acid extraction kit that is suitable for environmental samples, such as wastewater. From the biosafety cabinet, take the appropriate concentration of wastewater samples for virus nucleic acid extraction, following the kit's instructions. If the concentrated solution contains a significant amount of impurities or is particularly viscous, you may increase the number of rinses; it is recommended to rinse three times. RT-qPCR should be conducted as soon as possible after nucleic acid extraction. The remaining concentrate and nucleic acid samples should be stored at -80° C. Follow the instructions provided with the nucleic acid extraction kit for any additional procedures.

6.5.2 Detection of SARS-CoV-2 Nucleic Acid

SARS-CoV-2 nucleic acid detection was conducted using reverse transcription real-time PCR (RT-qPCR). Select an appropriate RT-qPCR kit for detecting SARS-CoV-2 nucleic acid in wastewater samples. Single or double target gene testing should be employed for the open reading frame 1ab (ORF1ab) and/or the nucleocapsid protein (N), with adjustments made as necessary according to testing requirements and relevant regulations. The amplification reaction system was prepared and executed in accordance with the instructions provided with the RT-qPCR kit. Each sample should be processed in parallel. High-throughput sequencing is routinely performed for samples containing the ORF1ab gene or the N gene, particularly those with an average CT value of 35. Specific criteria for selecting sample categories are outlined in Table 1.

Table 1 Standards for Selecting RNA Sample Types Suitable for High-Throughput Sequencing

RNA sample types	Whether the requirements are met
A class	CT value ≤32, meet the high-throughput sequencing requirements
B class	32<CT value≤35, basically meet the requirements
C class	CT value>35, not meet the requirements, important samples can be tried to sequence

6.6 Sequencing library preparation

6.6.1 Targeted amplification by multiplex PCR

The extracted viral nucleic acid samples from wastewater are processed within a biosafety cabinet. RNA is reverse-transcribed into cDNA using reverse transcription reagents. A primer pool designed to cover the entire SARS-CoV-2 genome is then employed to prepare a multiplex PCR system for genome enrichment and amplification. Due to the unique characteristics of wastewater samples (high fragmentation of viral genomes and the presence of numerous impurities or interfering substances), the primer pool for amplifying SARS-CoV-2 genome must meet the following requirements:

a) Short amplicons (typically 100 - 400bp) are recommended. The length of amplicons should not be excessively long;

b) The multiplex PCR primer pool must exhibit high specificity;

c) The multiplex PCR primer pool should be promptly updated according to the mutation of the SARS-CoV-2 genome.

6.6.2 Sequencing library preparation of the multiplex PCR amplification products

Purification of amplification products is performed using the magnetic bead method or other library purification reagents. Adaptor additions, library amplification (optional), library purification, quantification and homogenization mixing are performed according to the instructions of the library preparation kit used.

6.7 Sequencing

6.7.1 Selection of the sequencing platform

Select high-throughput sequencing platforms that meet the applicable scope of this document, such as reversible terminator sequencing and combinatorial probe anchored polymerization sequencing.

6.7.2 Sequencing

Sequencing was conducted following the guidelines provided by the high-throughput sequencing kit. The length of the sequencing reads should range of 50－300bp. The volume of sequencing data for individual samples should be determined based on the various library preparation methods employed. The appropriate number of samples should be selected for library pooling and sequencing in accordance with the throughput of the sequencing platform, ensuring adequate coverage and sequencing depth. Wastewater samples generally contain a variety of SARS-CoV-2 variants, which necessitates an increase in the volume of sequencing data to meet the analytical requirements. Different sequencing strategies should be chosen based on specific data analysis needs and timeliness considerations.

6.8 Data analysis

6.8.1 Data quality control

After library sequencing, raw sequencing data were obtained. Quality control is required to remove sequences containing adapters or low quality, and primer sequences should be removed for data sequenced by multiplex PCR amplicon method before subsequent bioinformatics analysis. Quality control indicators include Q20, Q30, etc.

6.8.2 Genome alignment

The reference strain NC_045512.2 on GenBank should be used as the SARS-CoV-2 reference genome. Alignment files were obtained by genomic comparison of sequencing data with reference sequences after quality control.

6.8.3 Calculation of sequencing coverage and average depth

The sequencing coverage and average sequencing depth were calculated according to equation (1) and (2). Sequencing coverage values for different sequencing depths are calculated according to actual needs, e.g. 10× coverage is commonly used to evaluate the quality of sequencing data.

$$CVR = \frac{CVB}{GB} \times 100\% \quad (1)$$

Eq:

CVR——Sequencing coverage;

CVB——Number of genomic sequence bases covered;

GB——Total bases number in the genome.

$$DP = \frac{MB}{CVB} \quad (2)$$

Eq:

DP——Average sequencing depth;

MB——Number of total aligned bases;
CVB——Number of genomic sequence bases covered .

6.8.4 Analysis of the lineage abundance ratio of SARS-CoV-2 variants

The wastewater samples contain multiple variants, rendering them unsuitable for analysis through direct consensus sequence generation via reference-based assembly. Based on the alignment file obtained in Section 6.8.2, use software or tools to mapping the SNVs in the alignment file against publicly available SARS-CoV-2 lineage-defining mutations library or self-constructed library to generate the SNV results file and SNV sequencing depth file. Apply the Depth-weighted Least Absolute Deviation Regression (DW-LAD) method to resolve the abundance ratios of multiple SARS-CoV-2 lineages in the short-reads data of the mixed sample. Finally, lineage abundance results are corrected using specific tools. For practical application, please refer to Annex C for an example process.

7 Quality control

7.1 Quality control for enrichment and concentration

Reference to WS/T 799-2022, “Method for enrichment and nucleic acid detection of SARS-CoV-2 in sewage”.

7.2 Quality control for nucleic acid extraction

Reference to WS/T 799-2022, “Method for enrichment and nucleic acid detection of SARS-CoV-2 in sewage”.

7.3 Quality control for PCR detection

Reference to WS/T 799-2022, “Method for enrichment and nucleic acid detection of SARS-CoV-2 in sewage”.

7.4 Quality control for the library preparation

Perform quality control for the sequencing library from the following aspects:

a) Indexes and adapters should be clearly added to prevent cross-contamination between indexes;
b) The concentration of multiplex PCR amplification products and sequencing libraries should be verified using fluorescence quantification methods. If necessary, a bioanalyzer should be used to determine the size of amplified products and library fragments;
c) Libraries that meet quality requirements are normalized and pooled, with the normalized concentration determined by the sequencing platform requirements.

7.5 Quality control of the sequencing data

Perform quality control on sequencing data from the following aspects:

a) Select an appropriate sequencing strategy based on sequencing and data requirements, such as SE100+10+10;
b) Choose an appropriate data volume based on analytical needs. For monitoring data based on 10× sequencing depth with 60% coverage, a data volume of 1Mb reads per sample can meet basic monitoring requirements, while 5Mb reads per sample can provide better sequencing results. When the concentration of enriched SARS-CoV-2 in wastewater samples is relatively low (CT value > 32), approximately 5Mb reads per sample can meet basic monitoring needs. Therefore, it is recommended that the sequencing data volume for each wastewater sample be 5Mb reads;

c) Sequencing data quality control includes, but is not limited to, the removal of duplicate sequences, low-quality sequences, and adapter sequences. Qualified data should meet the following requirements:

1) For sequencing read lengths ranging from 50-300 bp, the proportion of bases with Q20 quality should be ≥ 90%, and the proportion of bases with Q30 quality should be ≥ 80%;
2) The proportion of adapter sequence contamination should not exceed 1%;
3) The remaining sequence length after cutting low-quality and adapter sequences should be greater than 35bp.

7.6 Quality control of lineage abundance analysis of SARS-CoV-2 variants

Perform quality control on the abundance analysis of SARS-CoV-2 variants from the following aspects:

a) The SARS-CoV-2 lineage-defining mutations library used must include specific mutations or specific mutations combinations for different lineages to ensure the accuracy of the alignment results;

b) The data optimization process should follow the rules below:

1) Sequencing data for subsequent analysis and presentation should be selected based on analysis requirements, with a 10× coverage threshold (e.g., ≥ 20%, ≥ 60%, ≥ 90%, or other values). For monitoring purposes, include sequencing data with 10× coverage ≥ 60%;
2) SNV filtering thresholds: MAPQ ≥ 20, depth ≥ 20X, and mutation frequency ≥ 3%. For monitoring purposes focusing on early detection, the mutation frequency threshold can be lowered to ≥1%;
3) Remove results for low-abundance variants (relative abundance <1%);
4) Correction of lineage abundance results: If the identified variants lack unique mutations for lineage definition or exhibit overlapping mutation sites between lineages, the lineage and its abundance cannot be determined with precision. In such cases, the corresponding lineages and their abundances should be merged in the output, separated by a "|". For example: "BA5.2.1|BA5.2, 30%".

Annex A
(Normative)
Experimental conditions and laboratory zoning requirements

A.1 Experiment condition

Laboratory instruments and equipment must comply with the requirements specified in GB 19489-2008. The laboratory should meet the requirements outlined in the "Laboratory Biosafety Guidelines for Novel Coronavirus". Operations involving unprocessed infectious materials, including nucleic acid extraction and inactivation of clinical samples before reliable inactivation methods are applied, should be conducted in a Biosafety Level 2 (BSL-2) laboratory. Personal protective equipment designed for Biosafety Level 3 (BSL-3) must be used during these operations. Nucleic acid detection on infectious materials or live viruses should be carried out in a BSL-2 laboratory after they have been inactivated using reliable methods. Operations related to after reverse transcription of SARS-CoV-2 RNA can be carried out in a Biosafety Level 1 (BSL-1) laboratory. The laboratory should be divided into distinct areas based on the type of operation being performed, with clear signage identifying the areas, such as reagent storage and preparation areas, sample and library preparation areas, amplification areas, and high-throughput sequencing areas. Cross-contamination between these areas should be avoided. Medical waste generated during the experiment must be autoclaved before disposal according to laboratory standards. After the experiment, the laboratory environment should be disinfected using ultraviolet light, 75% alcohol, chlorine-based disinfectants, or nuclease-free agents.

A.2 Laboratory zoing

A.2.1 Reagent storage and preparation areas

This area is used for the aliquoting and storage of reagents, preparation of amplification reaction mixtures, and storage and preparation of experimental consumables. The area should be equipped with refrigerators or freezers, centrifuges, laboratory benches, vortex mixers, micropipettes, etc. To prevent contamination, this area should maintain positive pressure.

A.2.2 Samples and the library preparation areas

Sample aliquoting and nucleic acid extraction should be carried out in a dedicated BSL-2 laboratory. Extracted nucleic acids could be transported to this area for addition to reaction mixtures. This area should be equipped with refrigerators or freezers, centrifuges, laboratory benches, vortex mixers, micropipettes, magnetic stands, etc. To prevent contamination, this area should maintain positive pressure.

A.2.3 Amplification areas

This area is used to perform nucleic acid amplification reactions and analyze amplification products. The area should be equipped with PCR amplification instruments. To prevent contamination of the environment by amplification products, this area should maintain a negative pressure that is lower than the pressure in the sample and library preparation areas.

A.2.4 High-throughput sequencing areas

This area is designated for sequencing, data analysis, report review, and issuance. The area should be equipped with high-throughput sequencing instruments. The area should maintain negative pressure.

Annex B
(Normative)
Requirements for reagents, materials, instruments and equipment

B.1 Reagents list

The list of reagents required for the experiment is as follows:

a) Viral nucleic acid extraction kit – used for releasing and obtaining nucleic acids from various samples and purifying it;
b) SARS-CoV-2 nucleic acid detection kit (RT-qPCR method) – used for quantitative detection of SARS-CoV-2 in samples;
c) Viral RNA reverse transcription kit – used to synthesize cDNA from RNA templates, should include at least: reverse transcriptase and buffer;
d) SARS-CoV-2 whole-genome multiplex PCR amplification kit – used to enrich the whole genome of SARS-CoV-2, should include at least: amplification primers, DNA polymerase, and buffer;
e) Sequencing library construction kit – used for preparing sequencing libraries from nucleic acid samples, should include at least: adapters, tags, purification beads, DNA ligase or DNA polymerase, and buffer;
f) Fluorescence quantitative analysis kit – includes fluorescent dye reagents for quantifying dsDNA;
g) Sequencing kit – used for sequencing libraries, should include at least: sequencing primers, high-efficiency sequencing enzymes, and buffer;
h) Nuclease-free water – molecular-biological grade.

Note: Other reagents should follow the instructions of the reagents used in the kits.

B.2 Reagents and materials list

The list of reagents and materials required for the experimental procedures is as follows:

a) Sterile screw-cap centrifuge tubes: 50 mL, 1.5 mL nuclease free;
b) Sterile pipette: 50 mL, made of polypropylene;
c) Sterile PCR tubes or PCR plates: nuclease-free;
d) Sterile pipette tips: 2 μL, 10 μL, 100 μL, 200 μL, 1 mL, with filter, nuclease-free;
e) 0.5 mL thin-walled transparent detection tubes compatible with fluorescence quantitative analyzers, nuclease-free.

Note: Other materials should follow the instructions of kits used.

B.3 Instruments and Equipment List

The list of instruments and equipment required for the experiment is as follows:

a) Refrigerated centrifuge: 4° C, 50 mL rotor, capable of withstanding centrifugal forces ≥5,000 g; 4° C, 1.5 mL rotor, capable of withstanding centrifugal forces ≥20,000 g;
b) Biological safety cabinet: Class II or higher;
c) Autoclave;
d) Real-time fluorescence quantitative PCR instrument;
e) Conventional PCR instrument;
f) Fluorescence quantitative analyzer – used for quantifying dsDNA concentration;
g) Magnetic stand;

h) Mini-centrifuge;
i) Vortex mixer;
j) Pipettes: 2 μL, 10 μL, 100 μL, 200 μL, 1 mL, 50 mL.

Note: Other instruments and equipment should follow the instructions of the kits used.

Annex C
(Informative)
Example of lineage relative abundance analysis procedure of SARS-CoV-2 variants in wastewater

The wastewater samples contain a mix of various SARS-CoV-2 variants, and it is recommended to utilize the software Freyja (Karthikeyan S, et al., Nature, 2022) for analyzing the abundance of different lineages within these samples. The developers of the Freyja software used the matUtils package to extract lineage-defining mutations for various SARS-CoV-2 variants from the UShER global phylogenetic tree (Turakhia, Y. et al. Nat. Genet., 2021), forming a library of lineage-defining mutations. This library is represented as a "Barcode" matrix file: "usher_barcodes.csv".

$$A = \begin{bmatrix} a_{1,1} & \cdots & a_{1,N} \\ \vdots & \ddots & \vdots \\ a_{M,N} & \cdots & a_{M,N} \end{bmatrix}$$

where $a_{i,j}$ denotes the i-th lineage at mutation j. Freyja uses the alignment file obtained mapping the sequencing data with the SARS-CoV-2 reference sequence as an input file. To encode each sample, Freyja stores the SNV frequencies for each of the lineage-defining mutations. As SNV frequencies at positions with greater sequencing depth more accurately estimate the true mutation frequency, Freyja recovers relative lineage abundance by solving a DW-LAD regression problem (lineage deconvolution). Need to download the latest version of Freyja when use.

C.1 Download and installation of Freyja

It is recommended that users refer to the official tutorial provided by the Freyja project on Github and use the Conda environment management tool to complete the installation and configuration of the Freyja software.

C.1.1 Introduction and installation of Conda

Conda is an open-source package and environment management system designed for installing multiple versions of packages and their dependencies on Linux, OSX, and Windows systems. Miniconda, a lightweight distribution of Conda, is often used in practice. The latest version can be downloaded and installed through the official website, or via domestic mirror sources. Users can also add channels, which are the locations where packages are stored, based on their specific needs.

C.1.2 Download and installation of Freyja

It is recommended that users visit the usage guide website of the Freyja software and, after confirming the successful installation of the Conda environment management tool, strictly follow the official process to complete the installation of the Freyja software.

C.1.3 Freyja database update

To ensure the proper use of the Freyja software, users are required to periodically download its reference database file named “usher_barcodes.csv”. After the download is complete, the file should be renamed to a new filename that includes the download date. Then, this newly named file should be used to replace the original reference database file called by the local Freyja software.

C.2 Workflow of Freyja

Follow the guidelines on the Freyja website to run the software. Input the alignment file, which mapping sequencing data to the SARS-CoV-2 reference genome, to generate a SNV table and a SNV depth file, which contain SNV frequencies and depths. Finally, recovers relative lineage abundance SARS-CoV-2 variants (demixed.tsv).

C.3 Further correction of the Freyja output

Freyja output results usually include very low abundance SARS-CoV-2 lineage, or uncommon SARS-CoV-2 lineage, or low mutation frequency and sequencing depth for lineage-defining mutations in some lineages, or low genome coverage of sequencing data. To ensure accuracy of the results, they must be subjected to further refinement, which includes the removal of low-quality lineages. This additional correction can be performed using programming languages such as R or Python, and it is essential that it adheres to the specifications outlined in section 7.6 of the standard's main text.

Bibliography

[1] GB/T 19915.5-2005 Protocol of multiplex PCR identification of Streptococcus suis type 2.

[2] GB/T 35537 Requirements of the high-throughput gene sequencing result evaluation

[3] GB/T 37872-2019 Guidelines for validation of next-generation target region sequencing

[4] GB/T 40226-2021 Detection of environmental microbial metagenome — high throughput sequencing

[5] DB32/T 3762.11-2020 Technical specifications for SARS-CoV-2 detection Part 11: Whole-genome high-throughput sequencing

[6] Xiao, M., et al. Multiple approaches for massively parallel sequencing of SARS-CoV-2 genomes directly from clinical samples. Genome Med 12(1): 57

[7] Karthikeyan S, et al. Wastewater sequencing reveals early cryptic SARS-CoV-2 variant transmission. Nature. 2022 Sep;609(7925):101-108

[8] Turakhia, Y. et al. Ultrafast sample placement on existing tRees (USHER) enables real-time phylogenetics for the SARS-CoV-2 pandemic. Nat. Genet.53, 809–816 (2021)