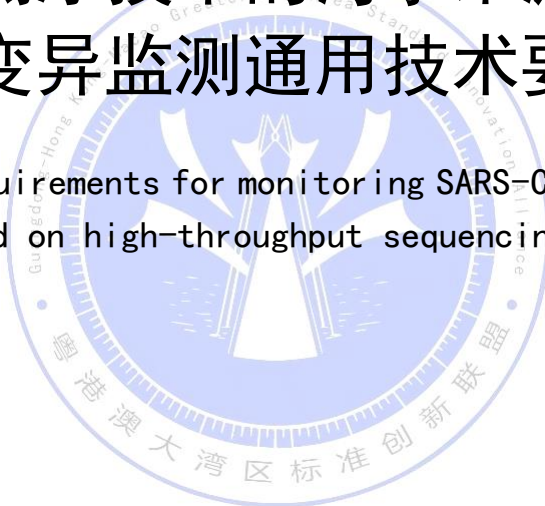


团 体 标 准

T/GBA 021—2024

基于高通量测序技术的污水来源新型冠状病毒 病毒变异监测通用技术要求

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



2024-12-31 发布

2025-01-10 实施

粤港澳大湾区标准创新联盟 发布

目 次

前言 II

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 缩略语 3

5 原理 3

6 试验步骤 3

 6.1 污水样本类型 3

 6.2 污水监测点位的布设 4

 6.3 污水样本的采集 4

 6.4 污水样本的富集浓缩 5

 6.5 污水样本的核酸提取及新冠病毒核酸检测 5

 6.6 测序文库制备 5

 6.7 测序 5

 6.8 数据分析 6

7 质量控制 6

 7.1 富集浓缩质量控制 6

 7.2 核酸提取质量控制 6

 7.3 PCR 检测质量控制 6

 7.4 文库构建质量控制 7

 7.5 测序数据质量控制 7

 7.6 新冠病毒变异株株系丰度分析质量控制 7

附录 A （规范性） 实验条件、实验室分区要求 8

附录 B （规范性） 试验试剂、材料、仪器和设备要求 9

附录 C （资料性） 污水中新冠变异株株系分布和丰度分析示例 10

参考文献 12

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

本文件由深圳市疾病预防控制中心提出。

本文件由粤港澳大湾区标准创新联盟归口。

本文件授权粤港澳大湾区标准创新联盟组织伙伴和所有成员单位使用，联盟组织伙伴需等同采用转化为自身团体标准，并在全国团体标准信息平台上公开标准基本信息。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件起草单位：深圳市疾病预防控制中心、香港大学、深圳华大智造科技股份有限公司、澳门科技大学、中国食品药品检定研究院、深圳大学、深圳万物传感科技有限公司、湖北国际旅行卫生保健中心（武汉海关口岸门诊部）、中国海关科学技术研究中心、深圳天祥质量技术服务有限公司。

本文件主要起草人：扈庆华、吕子全、付玉林、李迎慧、张彤、陈芳、王逸丛、黄锦伟、张晗、刘东来、杜琛、徐晓庆、龚睿、王艺凯、颜妙丽、苏凤侠、杨自飞、徐嘉铖、李兆新。

本文件为首次发布。



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

1 范围

本文件规定了基于高通量测序技术的污水来源新冠病毒变异监测通用技术要求的原理、试验步骤及质量控制要求等。

本文件适用于污水样本中新型冠状病毒的富集浓缩、核酸提取、高通量测序及数据分析过程，其中高通量测序适用于可逆末端终止测序法和联合探针锚定聚合测序法等技术为主要技术原理的测序平台。

本文件不适用于：

- 以桑格（Sanger）测序为主要技术原理的第一代基因测序领域；
- 以单分子实时荧光测序法、单分子纳米孔链测序法、单分子纳米孔标签测序法等技术为主要技术原理的具有连续测序特征的单分子基因测序领域；
- 宏基因组测序、探针捕获靶向测序。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB 19489 实验室生物安全通用要求

GB/T 30989—2014 高通量基因测序技术规程

WS/T 799—2022 污水中新型冠状病毒富集浓缩和核酸检测方法标准

YY/T 1723—2020 高通量基因测序仪

3 术语和定义

下列术语和定义适用于本文件。

3.1

污水 wastewater

排入污水管网系统的污水的总称，主要来源包括居民区、学校、医院、商业服务机构及各种公共设施等。

3.2

新型冠状病毒 severe acute respiratory syndrome coronavirus 2

属于 β 属冠状病毒，基因组为线性单股正链的RNA病毒，全长约30 Kb，有包膜，呈圆形或椭圆形颗粒状，直径为60nm~140nm。

[来源：WS/T 799—2022，术语和定义3.1]

3.3

富集浓缩 enrichment and concentration

通过特定的技术手段对污水中存在的目标病原微生物进行提取和浓缩的过程，目的是提高其在样品中的浓度，以便于后续的检测和分析。

3.4

循环阈值 cycle threshold

在实时荧光定量PCR中，反应体系中的荧光信号达到设定的阈值时所经历的循环数。

[来源：WS/T 799—2022，术语和定义3.2]

3.5

引物池 primer pool

将多个引物混合在一起，以实现同时扩增多个靶标区域。

3.6

多重 PCR multiplex PCR

在同一PCR反应体系中加上两对以上引物，同时扩增出多个核酸片段的PCR反应。

[来源：GB/T 19915.5—2005，术语和定义3.2]

3.7

多重PCR扩增子测序 multiplex PCR amplicon sequencing

通过多重PCR技术同时扩增多个靶标基因区域（称为扩增子，Amplicon），并结合高通量测序平台对这些扩增子进行序列测定和分析的一种技术。

3.8

基因测序 gene sequencing

对核酸分子不同碱基类型的测定，即测定组成核酸分子的腺嘌呤（A）、鸟嘌呤（G）、胞嘧啶（C）和胸腺嘧啶（T）或者尿嘧啶（U）等碱基的排列顺序。

[来源：YY/T 1723—2020，术语和定义3.1]

3.9

测序文库 sequencing library

作为测序模板的核酸片段，具有特定的大小范围，通常包含：接头和/或用于测序引物结合的特定序列、序列捕获和/或识别特定区域的标识。

[来源：ISO 20397—1:2022，术语和定义3.5，有修改]

3.10

测序读长 read length of gene sequencing

单次运行可读取的质量合格的序列片段长度，通常以碱基数表示。

[来源：YY/T 1723—2020，术语和定义3.4]

3.11

测序片段 reads

高通量测序平台产生的含有碱基序列和质量值的序列片段。单末端或双末端测序产生的一条序列信息，计为一条序列数。

[来源：GB/T 35890—2018，术语和定义3.2，有修改]

3.12

碱基识别质量 quality of base calling

评价碱基准确识别的概率，简称为Q，通常以数值定义，定义 $Q = -10 \log_{10}(P)$ ，其中P为碱基识别的错误概率。

注1：碱基识别质量值与碱基识别错误率呈负相关，二者遵循对数函数关系。

注2：碱基识别质量值越高，错误率越低。如Q20指碱基识别准确率为99%，或错误率为1%。Q30指碱基识别准确率为99.9%，或错误率为0.1%。

3.13

测序覆盖率 coverage rate of sequencing

待测样本核苷酸检测结果覆盖于参考序列上的比例（测序覆盖率=测序区域长度÷参考序列总长度）。

[来源：GB/T 30989—2014，术语和定义3.30]

3.14

测序深度 depth of sequencing

待测样本中某个核苷酸被检测到的次数。

[来源：GB/T 30989—2014，术语和定义3.31]

3.15

病毒变异株 viral variant

通常指病毒基因组发生突变，导致其在传播能力、致病性或免疫逃逸等方面与原始毒株存在显著差异的病毒群体。

注：新冠病毒变异株通常是由Pangolin工具根据基因组的突变情况进行分类和编号，将新冠病毒划分为不同的谱系，帮助追踪病毒的演变和传播情况。

3.16

单核苷酸位点突变 single nucleotide variant

指基因组中的不同位置发生的单核苷酸（A、T、C或G）替代，且该变异的频率不受群体内检测频率的限制，涵盖了所有单一差异。

3.17

比对质量 mapping quality

由比对软件输出的比对质量值，表示测序数据比对到参考基因组时，序列比对位置的置信度。

注：如比对质量 ≥ 20 指序列错误比对概率 $\leq 1\%$ 。

3.18

SNV 突变频率 SNV mutation frequency

待测样本中某个核苷酸位点突变在覆盖该位点所有测序数据中所占的比例。

3.19

株系丰度 abundance of variant

指在特定样本中，某个新冠病毒变异株在所有变异株中所占的相对比例。

4 缩略语

下列缩略语适用于本文件。

CT: 循环阈值 (Cycle threshold)

DNA: 脱氧核糖核酸 (Deoxyribonucleic acid)

dsDNA: 双链DNA (Double-stranded DNA)

Kb: 千碱基 (Kilobase pairs)

MAPQ: 比对质量 (Mapping quality)

Mb: 兆/百万条reads (Megabase)

PCR: 聚合酶链式反应 (Polymerase Chain Reaction)

RNA: 核糖核酸 (Ribonucleic acid)

RT-qPCR: 逆转录定量PCR (Reverse Transcription Quantitative Polymerase Chain Reaction)

SNV: 单核苷酸位点突变 (Single nucleotide variants)

SARS-CoV-2: 新型冠状病毒 (Severe acute respiratory syndrome coronavirus 2)

5 原理

基于高通量测序技术的污水来源新型冠状病毒变异监测采用多重PCR扩增子测序的方法对污水样本中的新型冠状病毒RNA进行测序，数据分析获取污水中新冠病毒变异株株系丰度，主要通过以下步骤实现：

- 样本采集与处理：布设污水监测点位、采集污水样本并进行病毒颗粒及核酸片段的富集浓度，以便后续的检测、测序；
- 样本的核酸提取及新冠病毒核酸检测：从浓缩后的样本中提取病毒 RNA，并进行定量检测；
- 文库构建与测序：将病毒 RNA 逆转录为 cDNA，构建多重 PCR 扩增子测序文库，并在高通量测序平台上完成测序；
- 数据分析：对测序数据进行质量控制、比对、变异株检测，获得污水中新冠病毒变异株株系丰度占比。

6 试验步骤

6.1 污水样本类型

污水样本来源包括但不限于城市污水处理厂（水质净化厂）、污水泵站、医疗机构、航空器、火车站等关键交通枢纽、畜禽养殖场以及农产品批发市场等场所。

样本的获取、前处理、保存、销毁、运输等操作应符合《关于印发新型冠状病毒肺炎防控方案（第九版）的通知》中的附件12《新冠病毒标本采集和检测技术指南》的要求。

试验操作中需要满足的实验条件、实验室分区要求应符合附录A的规定；试验操作中需要的试验试剂清单、试剂材料清单、仪器与设备清单应符合附录B的规定。

6.2 污水监测点位的布设

根据监测的具体目标布设污水监测点位，一般选择在城市污水处理厂的进水口和污水管网的关键节点设置监测点位。根据监测需求，可选择在医疗机构，特别是传染病医院（含带传染病房综合医院）布设监测点位，收集病区污水；或是在排水管道布局清晰的居民小区或城中村污水井设置监测点位。重点场所和社区采样点宜布设在监测区域下游距离最近的排水口或检查井。开展污水新冠病毒长期监测时，采样点宜保持相对稳定。在监测点位布设前，需明确监测点位的管网汇入及水流来源情况，并掌握污水收集管网覆盖区域的人口数。

6.3 污水样本的采集

6.3.1 采集器材

采样器材的材质应具有较好的化学稳定性，在样品采集过程及样品贮存期内不会与水样发生物理化学反应而引起水样组分浓度的变化。采样器具可选用聚乙烯、不锈钢、聚四氟乙烯等材质。样品容器可选用硬质玻璃、聚四氟乙烯塑料等材质，若重复使用，应事先消毒清洗，并能耐受60℃水浴且不会对后续检测造成影响。

6.3.2 采样方式

6.3.2.1 基本要求

水样采集方式可选择手工或自动化设备采样，进行长期、连续监测宜使用具有冷藏功能的污水自动采样器，保证水样的自动采集时间精准、定量准确与样品完整性。可根据监测目的选择采样方式，一般选择采集等时混合水样，即在某一时段内，在同一采样点位按等时间间隔采集等体积水样，形成混合水样。在疫情早期，可通过在社区布点进行3小时混合样采集，有利于实现早期预警和溯源；在疾病流行期间，可在社区、泵站与水质净化厂进行24小时混合样采集，用于评估疫情流行强度及变化趋势，监测病毒变异株的组成及丰度。

样本采集后应在现场使用75%酒精对采样瓶外表面进行消毒，然后将采样瓶装入密封采样袋中，对密封采样袋外表面再次进行消毒，并将样本放于0℃~4℃冷藏条件下，2小时内送至检测实验室。

6.3.2.2 3小时混合样采集

在用水高峰期（如上午8时—11时）进行3小时连续采样，每15分钟采集一次水样，共采集12次，获得混合样本。根据检测需要设定每次采集的水样体积，建议每次采集不少于250 mL，共采集水样不少于3 L，混匀后分装250 mL到样本容器中送检。

6.3.2.3 24小时混合样采集

进行全天24小时的连续采样，每1小时采集一次，共采集24次以制备混合样本。根据检测需要设定每次采集的水样体积，建议每次采集不少于125 mL，整个24小时采集期间，总采集水样量应不少于3 L。采集结束后，将所有样本混合均匀，并分装至250 mL的样本容器中，随后送至实验室进行检测。

6.3.3 采样频次

应根据新冠病毒感染的流行强度来调整采样频次。在疫情相对平稳期间，常规采样频次建议为每周1至2次。当疫情进入高峰期时，建议增加采样频次至每周至少3次，以便更准确的监测病毒的传播和变异情况。

6.3.4 污水样本的信息收集

采集包括监测点名称、监测点编号、详细地址和经纬度、覆盖人口数等信息。每次采样时记录水温、悬浮物、pH值、日均水流量、COD（化学需氧量，Chemical Oxygen Demand，COD）、氨氮等指标信息。

6.4 污水样本的富集浓缩

样本送达实验室后应保存于 0℃~4℃环境中，并在 24 小时内进行富集浓缩处理。富集前应将样本充分混匀后置于 60℃水浴 30 分钟进行病毒灭活处理。富集浓缩流程、试剂耗材和仪器设备等均可参照卫生行业标准《污水中新型冠状病毒富集浓缩和核酸检测方法标准》（WS/T 799—2022）进行，具体包括聚乙二醇沉淀法、铝盐混凝沉淀法和离心超滤法。若参照标准有更新，以更新的标准为准。

6.5 污水样本的核酸提取及新冠病毒核酸检测

6.5.1 污水样本的核酸提取

选择适用于污水等环境样本的病毒核酸提取试剂盒，按照试剂盒说明，在生物安全柜中取适量污水样本浓缩液进行病毒核酸提取。当浓缩液中存在大量杂质或浓缩液较黏稠时，可适当增加杂质漂洗次数，建议漂洗3次。核酸提取完成后应尽快进行RT-qPCR检测。剩余的浓缩液和核酸样本应于-80℃条件下冻存。其他操作参照所用核酸提取试剂盒说明。

6.5.2 新冠病毒核酸检测

采用反转录实时荧光定量 PCR 进行新冠病毒核酸检测。选择适用于污水样本新冠病毒核酸检测的 RT-qPCR 试剂盒。采用单/双靶标基因检测，针对开放读码框 lab（open reading frame lab, ORFlab）和/或核壳蛋白（nucleocapsid protein, N），必要时可根据检测需求和相关规定进行调整。按照 RT-qPCR 试剂盒说明书进行扩增反应体系配制并对提取得到的样本核酸进行扩增检测，每个样本应做 3 个平行。常规对 Orf 基因或者 N 基因平均 CT 值≤35 的样本进行高通量测序。具体样本类别选择标准见表 1。

表 1 进行高通测序的 RNA 样本类别选择标准

RNA样本类别	是否满足高通量检测需求
A类	CT值≤32，满足污水中病毒高通量测序需求
B类	32<CT值≤35，基本满足病毒高通量测序需求
C类	CT值>35，不满足测序，重要样本可尝试测序

6.6 测序文库制备

6.6.1 多重 PCR 靶向扩增

在生物安全柜中打开污水中提取的病毒核酸样本，使用逆转录试剂将 RNA 逆转录成 cDNA。后采用覆盖新冠病毒全基因组的引物池，配制多重 PCR 扩增体系，对病毒全基因组进行富集扩增。由于污水样本的特殊性（新冠病毒片段的高度碎片化、含有大量杂质和干扰物），新冠病毒全基因组的引物池需满足以下要求：

- PCR 扩增子片段的长度不宜过长，宜采用短扩增子（通常为 100~400bp）；
- 多重 PCR 引物池需要具有较高的特异性，减少非特异性扩增出现的概率；
- 根据新冠变异位点的变化及时更新多重 PCR 引物池。

6.6.2 多重 PCR 扩增产物测序文库制备

采用磁珠法或相关文库纯化试剂对扩增产物进行纯化处理。根据所使用的文库制备试剂盒说明书进行接头添加、文库扩增（可选）、文库产物纯化、定量及均一化混合。

6.7 测序

6.7.1 测序平台选择

选择符合本文件适用范围的高通量测序平台，如联合探针锚定聚合测序法、可逆末端终止测序法等为原理的测序平台。

6.7.2 测序

根据高通量测序仪配套测序试剂盒操作说明书进行测序。测序读长在 50–300bp 范围内。根据不同的文库制备方法明确单个样本测序数据量要求,根据测序平台通量选择合适的样本数量进行文库混合及测序,以保证测序覆盖度和深度。污水中一般包含多种新冠病毒变异株,为保证测序结果分析的需求,需适当提高测序数据量。根据不同的数据分析和时效性需求,选择不同的测序策略。

6.8 数据分析

6.8.1 数据质控

文库经过测序得到原始数据,应进行质量控制,去除含接头或低质量的序列,对于多重 PCR 法测序的数据应去除引物序列,再进行后续生物信息学分析。质控指标包括 Q20、Q30 等。

6.8.2 基因组比对

应以 GenBank 上的参考株 NC_045512.2 作为新型冠状病毒参考基因组。将质控后的测序数据与参考序列进行基因组比对,最终获得测序数据与参考序列的比对文件。

6.8.3 测序覆盖率和测序平均深度计算

根据式(1)和式(2)分别计算测序覆盖率值和测序平均深度值。根据实际需求计算不同测序深度的测序覆盖率值,如 10×覆盖度常用于评价测序数据的质量。

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

式中:
CVR——测序覆盖率;
CVB——覆盖到的基因组序列碱基数;
GB——基因组全部碱基数。

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

式中:
DP——测序平均深度;
MB——总比对碱基数;
CVB——覆盖到的基因组序列碱基数。

6.8.4 新冠病毒变异株株系丰度占比分析

污水样本中包含多种变异株,因此不适合直接进行有参组装后获得一致性序列的方式进行分析。在 6.8.2 获得比对文件基础上,通过软件或工具将比对文件中的 SNV 与公开可用的新冠病毒株系定义突变数据库,或自建数据库进行比对,获得 SNV 结果文件和 SNV 的测序深度文件;通过深度加权最小绝对偏差回归 (Depth-weighted Least Absolute Deviation Regression, DW-LAD) 方法解析混合样本短读长序列数据中的多个 SARS-CoV-2 株系丰度占比;最终通过分析软件或工具校正株系占比结果。有关此过程的具体操作,可参见附录 C 的示例流程。

7 质量控制

7.1 富集浓缩质量控制

应符合 WS/T 799—2022 的要求。

7.2 核酸提取质量控制

应符合 WS/T 799—2022 的要求。

7.3 PCR 检测质量控制

应符合 WS/T 799—2022 的要求。

7.4 文库构建质量控制

从以下几方面对文库进行质量控制：

- a) 清晰标记添加的标签接头，防止标签之间造成污染；
- b) 多重 PCR 扩增产物及测序文库浓度宜采用荧光定量方法检验，必要时可采用生物分析仪检测扩增产物及文库片段大小；
- c) 符合质量要求的文库进行均一化和混合，均一化浓度以测序平台要求为准。

7.5 测序数据质量控制

从以下几方面对测序数据进行质量控制：

- a) 根据测序和数据需求选择合适的测序策略，如SE100+10+10；
- b) 根据数据分析需求选择合适的数据量。若以10×测序深度的60%覆盖度为标准纳入监测数据，通常1Mb reads/样本的数据量即可满足基本的监测需求，5Mb reads/样本的数据量可获得更佳的测序结果；当污水样本中富集的新冠病毒浓度均较低时（CT值>32），5Mb reads/样本左右的数据量可满足基本的监测需求。综上，推荐每个污水样本测序数据量为5Mb reads；
- c) 测序数据质量控制包括但不限于去除重复序列、低质量序列、接头序列等。合格数据应满足以下要求：
 - 1) 测序读长在 50–300bp 范围内时，Q20 的碱基比例需≥90%，Q30 的碱基比例≥80%；
 - 2) 接头序列污染比例不超过 1%；
 - 3) 过滤低质量序列、接头序列后的剩余序列长度大于 35bp。

7.6 新冠病毒变异株株系丰度分析质量控制

从以下几方面对新冠病毒变异株株系丰度分析质量进行控制：

- a) 所使用的新冠病毒株系定义突变数据库，必须包含不同株系的特有位点或特有位点组合，以确定比对结果的准确性；
- b) 数据结果优化过程需参照以下规则：
 - 1) 根据分析需求筛选 10×覆盖度达到某个值以上的样本数据进行后续的分析 and 展示（20%、60%、90%或其他）；若为监测目的，使用 10×覆盖度≥60%为标准纳入测序数据；
 - 2) SNV 过滤阈值：MAPQ≥20、深度≥20X 且突变频率≥3%。当监测目的以早发现为主要目的，可将突变频率的筛选标准降低至 1%；
 - 3) 低丰度变异株的结果（丰度<1%）需去除；
 - 4) 株系结果校正：当输出的变异株间不存在株系判定的特有位点，即存在株系间支持的突变位点相同时，无法准确判定株系及其丰度，应将对应的株系以“|”间隔合并输出株系及其丰度和。如“BA5.2.1|BA5.2, 30%”。

附 录 A
(规范性)
实验条件、实验室分区要求

A.1 实验条件

实验室仪器和设备要求应符合GB 19489-2008的规定。实验室应符合《新型冠状病毒实验室生物安全指南》的要求,未经培养的感染性材料的操作在采用可靠的方法灭活前进行核酸提取及临床样本的灭活等操作,应在生物安全二级实验室(BSL-2)进行,同时采用生物安全三级实验室(BSL-3)的个人防护;感染性材料或活病毒在采用可靠方法灭活后进行的核酸检测操作应在BSL-2进行。新冠病毒RNA反转录后的相关操作可在生物安全一级实验室(BSL-1)进行。实验室应根据不同工作内容划分独立区域,并有明显标志,如试剂储存区和准备区、样本与文库制备区、扩增区、高通量测序区,各区域间要避免交叉污染。试验过程中产生的医疗废物需高压灭菌后再按照实验室规范丢弃。试验结束后需采用紫外照射、75%酒精或含氯消毒剂、去核酸酶等方式对试验环境进行消毒处理。

A.2 实验室分区

A.2.1 试剂储存区和准备区

用于分装、储存试剂、制备扩增反应混合液,以及储存和准备实验耗材。该区应配备冰箱或冰柜、离心机、试验台、涡旋振荡器、微量加样器等。为防止污染,该区宜保持正压状态。

A.2.2 样本与文库制备区

标本的分装以及核酸提取应在独立的生物安全二级(BSL-2)实验室进行,提取后的核酸可以转运至该区加至反应液中。该区应配备冰箱或冰柜、离心机、试验台、涡旋振荡器、微量加样器、磁力架等。为防止污染,该区宜保持正压状态。

A.2.3 核酸扩增区

进行核酸扩增反应和产物分析。该区应配备PCR扩增仪。为防止扩增产物污染环境,该区宜保持负压状态,压力低于样本与文库制备区。

A.2.4 高通量测序区

进行上机测序、数据分析、报告审核及发放。该区应配备高通量测序仪。该区宜保持负压状态。

附录 B

(规范性)

试验试剂、材料、仪器和设备要求

B.1 试验试剂清单

试验操作中需要的试验试剂清单如下：

- 病毒核酸提取试剂盒，用于释放和获得各类样本中的核酸并进行纯化；
- 新冠病毒核酸检测试剂盒（RT-qPCR 法），用于样本中新冠病毒的定量检测；
- 病毒 RNA 逆转录试剂，以 RNA 为模板合成 cDNA，应至少包含：逆转录酶、缓冲液；
- 新冠病毒全基因组多重 PCR 扩增试剂，用于富集新型冠状病毒全基因组，试剂盒内至少应包含：扩增引物、DNA 聚合酶及缓冲液；
- 测序文库构建试剂，用于核酸样品的测序文库构建，试剂盒内至少应包含：接头、标签、纯化磁珠、DNA 连接酶或 DNA 聚合酶、缓冲液；
- 荧光定量分析试剂盒，包括可用于定量 dsDNA 的荧光染料试剂；
- 测序试剂盒，用于对测序文库的上机测序，试剂盒内至少应包含：测序引物、高效测序反应酶及缓冲液；
- 无核酸酶水：分子生物学级。

注：其他试剂参照所使用的试剂盒。

B.2 试剂材料清单

试验操作中需要的试剂材料清单如下：

- 无菌带盖离心管：50 mL，1.5 mL 无核酸酶；
- 无菌移液管：50 mL，材质为聚丙烯；
- 无菌 PCR 管或 PCR 板：无核酸酶；
- 无菌移液器吸头：2 μ L、10 μ L、100 μ L、200 μ L、1 mL，带滤芯，无核酸酶；
- 适配荧光定量分析仪的 0.5 mL 薄壁透明检测管，无核酸酶。

注：其他材料参照所使用的试剂盒。

B.3 仪器与设备清单

试验操作中需要的仪器与设备清单如下：

- 冷冻离心机：4℃，50 mL 转子，可承受离心力 $\geq 5,000$ g；4℃，1.5 mL 转子，可承受离心力 $\geq 20,000$ g；
- 生物安全柜：II 级或 II 级以上；
- 高压灭菌器；
- 实时荧光定量 PCR 仪；
- 普通 PCR 仪；
- 荧光定量分析仪，可用于定量 dsDNA 的浓度；
- 磁力架；
- 掌上离心机；
- 涡旋混匀仪；
- 移液器：2 μ L、10 μ L、100 μ L、200 μ L、1 mL、50 mL。

注：其他仪器和设备参照所使用的试剂盒。

附 录 C

(资料性)

污水中新冠变异株株系分布和丰度分析示例

污水样本中混合了多种变异株，建议使用软件Freyja (Karthikeyan S, et al. Nature, 2022) 对污水中新冠病毒变异株不同株系丰度进行分析。Freyja软件的开发者使用matUtils包从UShER系统发育树 (Turakhia, Y. et al. Nat. Genet, 2021) 中获得了新冠病毒不同变异株的定义突变，形成包含不同变异株定义突变的数据库，“Barcode”矩阵文件——“usher_barcodes.csv”。

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

其中 $a_{i,j}$ 表示变异株 i 在 j 处的突变。Freyja将测序数据与新冠病毒参考序列比对后获得的比对文件作为输入文件，与新冠病毒株系定义突变数据库进行比对，估计SNV的突变频率和深度；通过深度加权最小绝对偏差回归分析混合样本短读长测序数据中的多个新冠病毒株系丰度分布结果（谱系反卷积，lineage deconvolution）。使用时需下载最新版本。

C.1 Freyja 软件下载安装

建议用户参照Github上Freyja软件官方提供的教程，借助Conda环境管理工具来完成Freyja软件的安装及配置工作。

C.1.1 Conda简介与安装

Conda是一个开源的软件包管理系统和环境管理系统，用于安装多个版本的软件包及其依赖关系，适用于Linux，OSX和Windows。实际使用时可应用Miniconda，其为conda轻量级的发行版。可通过miniconda官方网站或通过国内镜像源下载安装Miniconda最新版本，并根据自己需求添加channel（channel是软件包存在的位置）。

C.1.2 Freyja下载安装

建议用户访问Freyja软件的使用指南网址，在确认已成功安装Conda环境管理工具的基础上，严格按照官方流程完成Freyja软件的安装工作。

C.1.3 Freyja软件数据库更新

为了确保Freyja软件的正常使用，用户需定期下载其参考数据库文件“usher_barcodes.csv”。下载完成后，应将该文件重命名为一个包含下载日期的新文件名，随后用此新命名的文件替换本地Freyja软件调用的原参考数据库文件。

C.2 Freyja 软件使用流程

根据Freyja网站指引使用Freyja软件，输入测序数据与新冠病毒参考序列的比对文件，生成SNV表格文件和SNV的测序深度文件，获得SNV突变频率及深度，最终生成新冠病毒不同株系丰度文件demixed.tsv。

C.3 Freyja 结果进一步优化数据

Freyja输出结果中可能包含丰度极低，或非常见的新冠病毒株系丰度结果，或某些株系的定义突变（SNV）突变频率和测序深度较低，或者测序数据基因组覆盖度较低。为保证数据分析的准确性，需去除类似的低质量株系丰度分布结果，对结果进一步校正。该步骤可通过R语言、Python等来完成数据处理。数据校正需遵循标准正文部分7.6要求。



参 考 文 献

- [1] GB/T 19915.5—2005 猪链球菌2型多重PCR检测方法
- [2] GB/T 35537 高通量基因测序结果评价要求
- [3] GB/T 37872—2019 目标基因区域捕获质量评价通则
- [4] GB/T 40226—2021 环境微生物宏基因组检测高通量测序法
- [5] DB32/T 3762.11—2020 新型冠状病毒检测技术规范第11部分：全基因组高通量测序
- [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)

