doi: 10.13241/j.cnki.pmb.2021.13.001

·基础研究·

不同 DNA 聚合酶对运用 ARTIC 工作流的新型冠状病毒纳米孔测序的影响 *

李瑾慧[1] 李沛翰[1,2] 林彦锋[1,2] 王凯英[1,2] 李利忠[1] 邱少富[1] 李 鹏 宋宏彬[1] 贾雷立[1△]

（1 中国人民解放军疾病预防控制中心 北京 100071；2 军事科学院军事医学研究院 北京 100039）

摘要 目的：为了验证不同高保真 DNA 聚合酶是否会对运用 ARTIC 工作流进行新型冠状病毒纳米孔测序产生影响。**方法：**使用英国 Nanopore 公司 MinION 测序仪对 2 份已获得全基因组序列的新冠肺炎确诊病例核酸样本分别采用 KAPA HiFi HotStart ReadyMix，PrimeSTAR® GXL DNA Polymerase 和 NEBNext High-Fidelity 2X PCR Master Mix 进行 ARTIC 工作流的多重 PCR 扩增，对扩增产物进行测序，并对测序质量进行分析。**结果：**不同高保真 DNA 聚合酶在相同扩增条件下，扩增产物的质检结果和测序质量均不相同，NEBNext High-Fidelity 2X PCR Master Mix 在覆盖度和测序深度上明显好于另外两种酶。**结论：**NEBNext High-Fidelity 2X PCR Master Mix 在纳米孔新型冠状病毒 ARTIC 快速测序工作流中的应用效果较好。

关键词：新型冠状病毒；高保真 DNA 聚合酶；纳米孔测序；ARTIC 工作流

中图分类号：Q93-332；R-332；R373.1 **文献标识码：**A **文章编号：**1673-6273(2021)13-2401-05

The Influence of Different DNA Polymerases on the Novel Coronavirus Nanopore Sequencing with ARTIC Workflow*

LI Jin-hui[1], LI Pei-han[1,2], LIN Yan-feng[1,2], WANG Kai-ying[1,2], LI Li-zhong[1], QIU Shao-fu[1], LI Peng[1], SONG Hong-bin[1], JIA Lei-li[1△]

(1 Center for Disease Control and Prevention of PLA, Beijing, 100071, China;

2 Academy of Military Medical Sciences, Academy of Military Sciences, Beijing, 100039, China)

ABSTRACT Objective: To verify whether the selection of different high-fidelity DNA polymerases has an impact on the results of nanopore sequencing of SARS-CoV-2 with ARTIC-nCoV network workflow. **Methods:** Multiplex PCR amplification of ARTIC-nCoV network workflow were performed on nucleic acid samples of two confirmed COVID-19 patients with KAPA HiFi HotStart ReadyMix, PrimeSTAR® GXL DNA Polymerase or NEBNext High-Fidelity 2X PCR Master Mix. The amplicons were sequenced on Nanopore MinION platform, then were analyzed. **Results:** Under same conditions, different high-fidelity DNA polymerases lead to different results of quality inspection and sequencing quality. NEBNext High-Fidelity 2X PCR Master Mix is the best among the three enzymes in coverage and sequencing depth. **Conclusion:** NEBNext High-Fidelity 2X PCR Master Mix has better effect in nanopore sequencing with ARTIC-nCoV network workflow.

Key words: SARS-CoV-2; High-fidelity DNA polymerases; Nanopore sequencing; ARTIC-nCoV network workflow

Chinese Library Classification(CLC): Q93-332; R-332; R373.1 Document code: A

Article ID: 1673-6273(2021)13-2401-05

前言

自 2020 年 1 月以来，新型冠状病毒肺炎（COVID-19）疫情已迅速传播到世界各地，对国家安全、社会稳定和经济发展造成重大影响[1]。目前已经阐明该病系感染冠状病毒 SARS-CoV-2 所致。冠状病毒（CoVs）是迄今为止被鉴定出的最大的 RNA 病毒，冠状病毒科分为 4 个属，分别为 α，β，δ 和 γ 冠状病毒[2]。可以感染人类的七种冠状病毒（HCoV），SARS-CoV 和 SARS-CoV-2 及 MERS-CoV 同属于 β- 冠状病毒谱[3-6]。

疫情发生以来，高通量测序技术的应用在疫情预警防控过程中发挥了关键性作用。2020 年 2 月 5 日，首个基于 ARTIC 快速测序方案获得的新型冠状病毒基因组序列在 GISAID 数据库发布，项目组开发了一组针对新型冠状病毒的实验室和生物信息学工作流程，使用纳米孔测序仪，8 小时内即可完成病毒基因组的鉴定，国内及国际实验室正在使用该工作流进行针对该病毒的更深入的研究[7,8]。该工作流是建立在对病毒全基因组进行多重扩增的基础上，既往也有研究发现不同 DNA 聚合酶对模板的扩增具有偏好性，因此 DNA 聚合酶就成了影响扩

* 基金项目：国家科技重大专项项目（2018ZX10305410-004）

作者简介：李瑾慧（1988-），女，硕士，主要研究方向：生物安全，E-mail: jinhuili311@163.com

△ 通讯作者：贾雷立（1976-），男，硕士生导师，研究员，主要研究方向：生物安全，E-mail: jialeili@163.com，电话：010-66948269

（收稿日期：2020-12-31 接受日期：2021-01-27）

增准确性的关键因素之一[9,10]。ARTIC 工作流推荐使用的高保真 DNA 聚合酶是 NEB Q5 Hot Start HF Polymerase，本研究采用了操作更为简便的 NEBNext High-Fidelity 2× PCR Master Mix，以及高通量测序中经常使用的 KAPA HiFi HotStart ReadyMix 和 PrimeSTAR® GXL DNA Polymerase 分别对两例新冠肺炎确诊病例核酸样本进行多重 PCR 扩增，并运用 MinION 三代测序平台进行扩增子测序，进而分析了不同高保真 DNA 聚合酶对运用 ARTIC 工作流测序的影响，以期为新型冠状病毒全基因组测序中的应用提供数据支撑。

1 材料与方法

1.1 材料

1.1.1 样本来源 本实验室获得 2 份 COVID-19 确诊病例的核酸样本 40 μL，编号为 231 和 235，核酸检测 Ct 值分别为 27.39 和 29.62。

1.1.2 仪器与试剂 Qubit3.0 荧光计、Qubit dsDNA HS Assay Kit 购自美国 ThermoFisher 公司；NEBNext Ultra II RNA First Strand Synthesis Module、NEBNext High-Fidelity 2X PCR Master Mix 购自英国 NEB 公司；KAPA HiFi HotStart ReadyMix 购自德国 Roche 公司；PrimeSTAR® GXL DNA Polymerase 购自日本 Takara 公司；Agencourt AMPure XP beads 购自美国 Beckman 公司；Qsep100 全自动核酸蛋白分析系统购自中国台湾 BIOptic 公司；快速测序试剂盒（SQK-RBK004）、MinION 测序仪及 R9.4.1 Flowcell 芯片购自英国 Nanopore 公司。

1.2 方法

1.2.1 多重 PCR 扩增 使用 ARTIC Network 开发的工作流推荐的 98 对扩增引物，按照奇偶数将引物进行等体积混样为两个引物池，标记为 pool1(只包含奇数引物对)和 pool2(只包含偶数引物对)，将混好的引物池稀释到 10 μM 备用。核酸样本用 NEBNext Ultra II RNA First Strand Synthesis Module 反转录为一链 cDNA 并作为模板，使用三种不同 DNA 聚合酶进行多重 PCR 扩增，具体扩增体系如表 1 所示。扩增条件为：98℃预变性 30 s，98℃变性 15 s，65℃退火和延伸 5 min，4℃保存。

表 1 多重扩增体系

Table1 Multiple amplification system

DNA polymerases	Multiple amplification system（25 μL/pool）
KAPA HiFi HotStart ReadyMix	Ready mix: 12.5 μL primer pool1/pool2: 3.6 μL Nuclease-Free Water: 6.4 μL cDNA: 2.5 μL
NEBNext High-Fidelity 2× PCR Master Mix	Master mix: 12.5 μL primer pool1/pool2:3.6 μL Nuclease-Free Water: 6.4 μL cDNA: 2.5 μL
PrimeSTAR® GXL DNA Polymerase	GXL DNA Polymerase: 1 μL 5× GXL Buffer:5 μL dNTP: 2 μL primer pool1/pool2:3.6 μL Nuclease-Free Water: 10.9 μL cDNA: 2.5 μL

1.2.2 纯化与质检 将 pool1 和 pool2 扩增后产物混合后用 Agencourt AMPure XP beads 等体积纯化，回收后产物用 Qubit3.0 进行荧光定量，Qsep100 检测片段大小。

1.2.3 MinION 平台测序 使用快速测序试剂盒(Oxford Nanopore，货号 SQK-RAD004)进行文库构建，采用 MinION 测序平台，上机芯片为 R9.4.1 Flowcell 进行测序。

1.2.4 数据处理 MinION 下机数据后，使用软件 pycoQC v2.5.0.21 1 对数据进行质控，分别以 231（SARS-CoV-2/human/CHN/231/2020）和 235（SARS-CoV-2/human/CHN/235/2020）作为参考序列，使用 SAMTools v1.3.1 2 和 bwa v0.6.2 3 将原始 reads 和参考序列进行比对。并将比对结果使用 R 语言的软件包 ggplot2 v3.1.1 4 进行可视化。

2 结果

2.1 多重 PCR 产物定量质检结果

运用 Qubit3.0 对多重扩增后产物进行浓度定量结果如表 2 所示，两例样本中使用 NEBNext High-Fidelity 2× PCR Master Mix 进行多重扩增的产物浓度最高，分别为 61.2 ng/μL 和 21.6 ng/μL，将扩增产物稀释成 500 pg/μL 用 Qsep100 进行片段分析结果如图 1 所示，使用 NEBNext High-Fidelity 2X PCR Master Mix 的扩增产物相对杂片段更少，目的片段峰值更高。

2.2 测序质量分析

数据下机后，使用软件 pycoQC v2.5.0.21 1 对数据进行质控分析，分别以 231（SARS-CoV-2/human/CHN/231/2020）和 235（SARS-CoV-2/human/CHN/235/2020）作为参考序列进行比对，结果如表 3。231、235 样本用 NEBNext High-Fidelity 2X PCR Master Mix 进行多重扩增测序的病毒序列比例为 98.17% 和 96.48%，基因组覆盖度分别为 99.68%和 98.71，测序深度为 459.51 和 717.44，测序质量明显好于使用 KAPA HiFi HotStart ReadyMix 和 PrimeSTAR® GXL DNA Polymerase 进行多重扩增的产物结果。我们用 ggplot2 v3.1.1 4 将基因组覆盖度进行可视化比较，如图 2 所示，在基因组的各碱基位点上，用 NEBNext

High-Fidelity 2× PCR Master Mix 进行多重扩增的产物覆盖度均高于另外两个超保真 DNA 聚合酶，测序深度也较另外两组聚合酶高，可见 NEBNext High-Fidelity 2× PCR Master Mix 更适合 ARTIC 快速测序方案中多重扩增的引物对和扩增条件，测序质量较优。

表 2 三种高保真 DNA 聚合酶多重扩增产物定量结果

Table 2 Quantitative results of multiple amplification products of three high-fidelity DNA polymerases

Sample	Ct value	KAPA HiFi HotStart ReadyMix(ng/μL)	NEBNext High-Fidelity 2× PCR Master Mix(ng/μL)	PrimeSTAR® GXL DNA Polymerase(ng/μL)
231	27.39	6.34	61.2	33.8
235	29.62	3.02	21.6	13.9

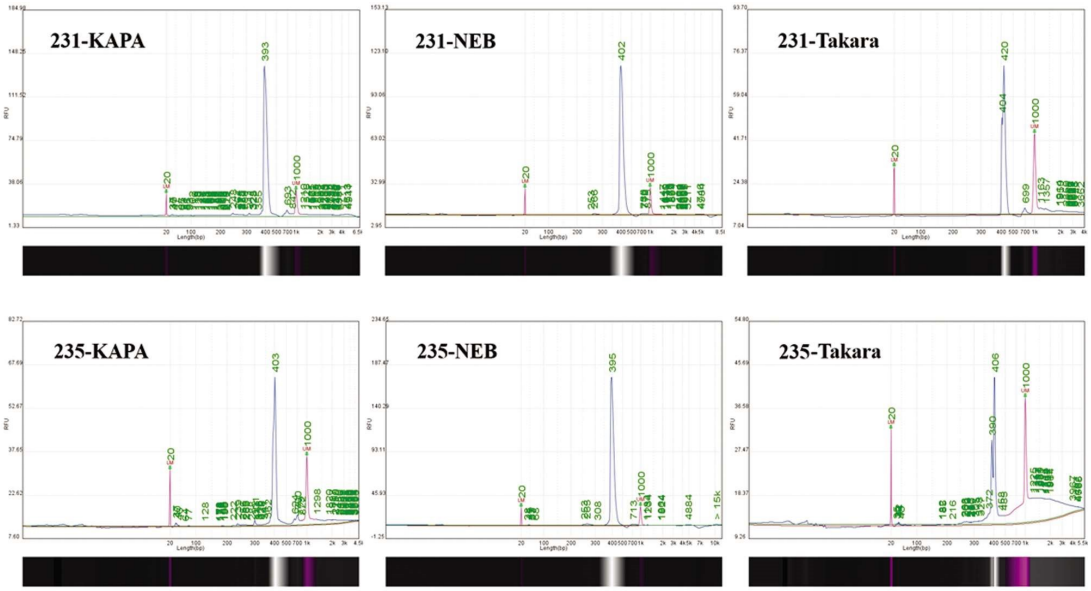

图 1 Qsep100 进行片段分析结果

Fig.1 Results of segment analysis performed by Qsep100

表 3 多重扩增产物测序结果比较

Table 3 Comparison of sequencing results of multiple amplification products

Samples	Total sequences	SARS-CoV-2 sequences	Host sequences	Sequences ratio (%)	Coverage(%)	Sequencing depth
231-KAPA	52957	48655	606	91.88	89.47	345.26
231-NEB	65477	64276	479	98.17	99.68	459.51
231-Takara	29913	27664	418	92.48	93.09	253.78
235-KAPA	10945	10050	294	91.82	73.63	78.42
235-NEB	104673	100990	850	96.48	98.71	717.44
235-Takara	8700	7861	207	90.36	71.69	58.56

3 讨论

高通量测序技术在此次疫情攻坚战中发挥了至关重要的作用[11-13]，但在病原检测应用上一直存在着技术瓶颈，宏基因组测序中病原含量极低，大量的宿主核酸浪费了大部分通量，测序结果高度依赖核酸的质量[14]。因此如何在短时间内快速获得病原全基因组序列信息以及时监控病原变异至关重要，前期实验中我们也尝试了样本过滤、离心等方法，效果甚微，ARTIC 工作流的出现，解决了 SARS-CoV-2 基因组测序的难题，通过多重扩增富集，增加病毒全基因序列，再通过扩增子测序获得病毒的基因组全序。该工作流自上线起，已在全世界广泛使用，截至 2020 年 8 月 30 日，ARTIC Network 网站已将该工作流更新至第三版[15,16]。

在建库过程中使用的三种超保真 DNA 聚合酶，均属于经过优化的工程酶，不同于来源于水生栖热菌的 Taq 酶与嗜热古细菌的 Pfu 酶，从使用范围来说均具有非常高的保真性能和校对功能，保真度大约是野生型 TaqDNA 聚合酶的 50 倍，大片断的扩增能达到 20kb 以上，并且对于高 GC 含量(GC＞70%)的模板也能达到很好的效果，这些聚合酶原则上都满足 ARTIC 工作流的实验要求。ARTIC 工作流推荐的多重 PCR 引物对具有高 GC 含量，大部分引物对的退火温度高于 65℃，因此，我们在保持扩增条件不变的情况下，进行了三种高保真 DNA 聚合

酶的尝试，结果发现，不同高保真 DNA 聚合酶的使用对测序结果产生一定的影响，虽然病毒序列比例均达到 90%以上，但在覆盖度和测序深度上差异较大，覆盖度可视化比较后可看出虽然扩增子之间丰度差异较大，但在相同扩增子上，NEBNext High-Fidelity 2X PCR Master Mix 的多重扩增效果明显高于其他两种酶。

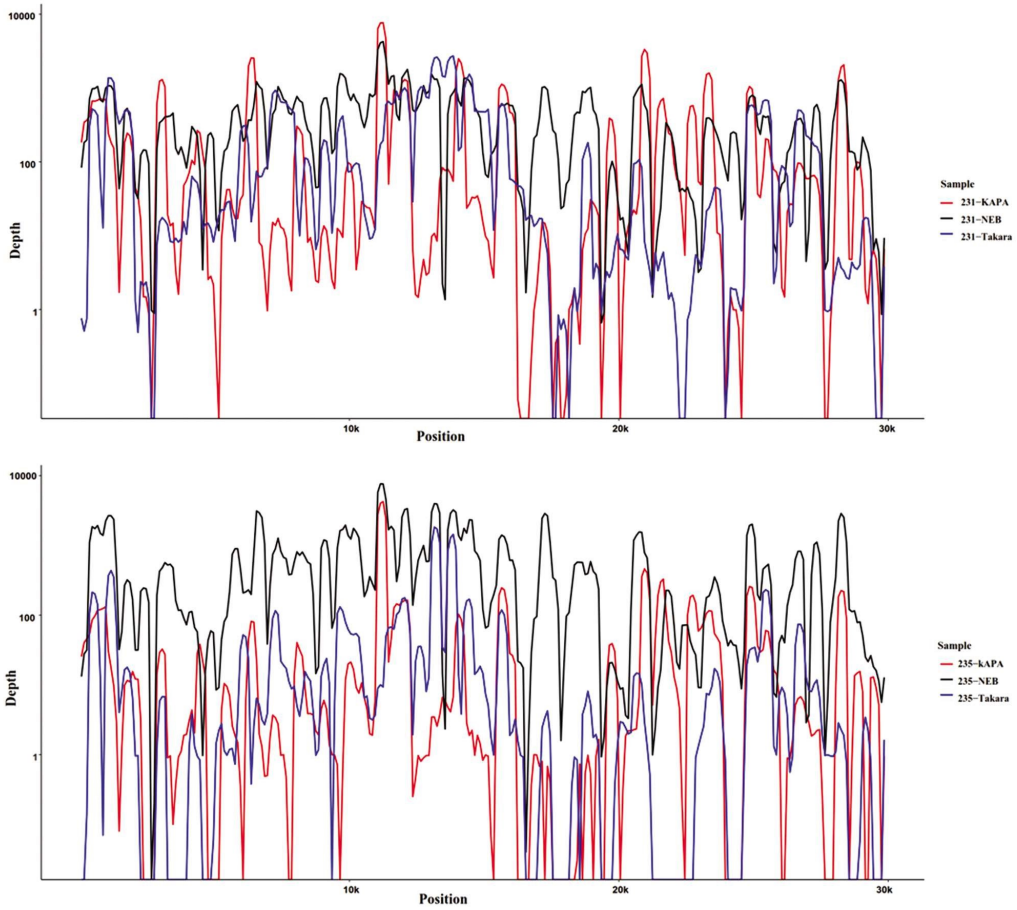

图 2 三种高保真 DNA 聚合酶多重扩增产物的测序深度分析

Fig.2 Sequencing depth analysis of multiple amplification products by three high-fidelity DNA polymerases

在本研究的扩增体系中，cDNA 占体系的 10%，ARTIC 工作流所使用引物池每条引物的浓度为 15 nM，每个引物池 49 对，原则上引物对数越多，扩增产物的产率就越低，特别是当引物对数大于 8 对时，非特异性的扩增产物就会增加；而在本实验的扩增体系中，由于引物对数过多，体系中引物浓度稍偏高就会引起错配从而导致非特异扩增反而降低了目的片段的产率[17]。从扩增产物的片段分析结果就能看出，在本研究中的多重扩增体系下，Q5 超保真聚合酶的扩增产物相对稳定。

由于多重 PCR 本质上是不同靶序列之间的竞争扩增，在 PCR 的早期阶段，随机效应会导致扩增效率的差异，特别是当模板 DNA 浓度很低的时候，优先扩增同样可以由靶序列本身具有的差异造成。因此，影响多重 PCR 的效率和特异性存在多种因素，成功的关键是反应中每个步骤、每种成分的系统优化[18-20]。既往已有报道使用 ARTIC 工作流进行新型冠状病毒基因组测序的研究，杭州市疾病预防控制中心运用该工作流完成了杭州市首例新型冠状病毒基因组组装，其研究选择了采用多重 PCR 扩增试剂盒（Qiagen Multiplex PCR Kit，德国 Qiagen 公司，货号：206143）进行多重 PCR 扩增，PCR 反应程序为三段式扩增，但也存在扩增子之间丰度差异大的问题，在 23277～23415 区域出现超过 1M 深度的高覆盖，在 6869～7041 区域和 29687～29835 区域的深度低于 100×，分析存在核酸 Ct 值较高的问题并需要增加测序深度[21]。在北京市 12 例境外输入新冠肺炎病例病毒基因组特征分析的研究中，使用了 Thermo Fisher 公司 Platinum Taq DNA polymerase High Fidelity 试剂盒进行多重 PCR 扩增并获得 12 条病毒全基因组序列，但文章没有就测序数据质量进行比较[22]。

由于研究样本数量的限制，本研究是在获得了全基因组序列的基础上进行实验条件的再摸索，但工作流在应用中仍有更多迭代和优化的空间。随着 ARTIC 工作流在新型冠状病毒基因组测序中的应用不断深入，国内外亦在成功应用该方法的基础上进行着进一步的优化和探索[23-26]，香港大学 Kelvin Kai-Wang To 团队利用纳米孔测序平台，结合并优化 ARTIC 和 Medaka 工作流程分析，发现全球首例 COVID-19 康复者二次感染病例，基因组学分析显示两次感染病毒不同[27]。当前，新冠肺炎疫情形势依然严峻，外防输入、内防反弹的压力巨大，因此建立并完善高通量测序技术在监控病毒变异中的应用十分重要[28]，虽然本研究存在着样本含量少的现实局限性，但对纳米孔测序结合 ARTIC 工作流在实际应用中提供了一定的数据支撑，也为防范病例发生以及监控新冠病毒变异提供了一定的参考。

参 考 文 献(References)

[1] WHO. WHO characterizes COVID-19 as a pandemic [EB/OL].https://www. who.int/emergencies/diseases/novel-coronavirus-2019/event-sas-they-happen

[2] Gorbalenya A E , Enjuanes L , Ziebuhr J , et al. Nidovirales: evolving the largest RNA virus genome[J]. Virus Research, 2006, 117(1):17-37

[3] Jie, Cui, Fang, et al. Origin and evolution of pathogenic coronaviruses [J]. Nature Reviews Microbiology, 2018, 17(3): 181-192

[4] Zumla A , Chan J F W , Azhar E I , et al. Coronaviruses - drug discovery and therapeutic options[J]. Nature Reviews Drug Discovery, 2016, 15(5): 327-347

[5] Lu R , Zhao X , Li J , et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding[J]. Lancet, 2020, 395(10224): 565-574

[6] 何文巧,陈清.SARS-CoV-2 溯源新进展[J/OL].南方医科大学学报: 1-6 [2020-12-28].https://kns-cnki-net.webvpn.ncepu.edu.cn/kcms/detail/44.1627.R.20201222.1755.002.html

[7] ARTIC NETWORK [EB/OL]. https://artic.network/ncov-2019

[8] QUICK J. nCoV-2019 sequencing protocol [EB/OL]. https: / /www. protocols.io /view /ncov-2019-sequencing-protocol-bbmuik6w

[9] 张敏,蔡俊鹏.不同 DNA 聚合酶对 PCR-DGGE 技术的影响[J].安徽农业科学, 2011, 39(14): 8231-8233

[10] 王彦杰,徐海燕,侯强川,等.DNA 聚合酶对 PacBio SMRT 测序结果影响的评价[J].中国微生态学杂志, 2018, 30(01): 93-99

[11] 赵康辰,朱小娟,葛以跃,等.江苏省 25 株新型冠状病毒全基因组分析[J].中华实验和临床病毒学杂志, 2020, 34(04): 352-356

[12] 郑燮,胡凤玉,孙静,等.第 1 株广州本地感染 COVID-19 病例的新型冠状病毒的分离鉴定[J].中国国境卫生检疫杂志, 2020, 43(05): 305-307+353

[13] 张炎华,陈炜,何文祥,等.福建省新型冠状病毒的分离和全基因组特征分析[J].中国人兽共患病学报, 2020, 36(05): 349-353

[14] 陶悦, 傅启华, 莫茜. 病原宏基因组测序在新型冠状病毒检测中的应用与挑战 [J/OL]. 中华检验医学杂志, 2020, (03): 217-218-219-220[2020-11-18]

[15] QUICK J.nCoV-2019 sequencing protocol v3 [EB/OL]. https://www.protocols.io/view/ncov-2019-sequencing-protocol-v3-locost-bh42j8ye.

[16] Tyson John R, James Phillip, Stoddart David, et al. Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore[J]. bioRxiv : the preprint server for biology, 2020[Epub ahead of print]

[17] J.萨姆布鲁克, M.R.格林. 分子克隆实验指南(第四版) [M]. 北京: 科学出版社, 2017

[18] Linz U, Delling U, RübsamenWaigmann H. Systematic studies on parameters influencing the performance of the polymerase chain reaction[J]. Journal of Clinical Chemistry & Clinical Biochemistry, 1990, 28(1): 5

[19] Kechin Andrey, Borobova Viktoria, Boyarskikh Ulyana, et al. NGS-PrimerPlex: High-throughput primer design for multiplex polymerase chain reactions[J]. PLoS computational biology, 2020, 16(12) [Epub ahead of print]

[20] 郑凯文,黄晓园,陈渡波,等.基于多重 PCR 和第二代高通量测序技术快速检测下呼吸道感染病原微生物方法的建立和应用[J].国际检验医学杂志, 2020, 41(17): 2066-2070

[21] 俞骅,王旭初,李钧,等.杭州市首例新型冠状病毒肺炎病例病毒全基因组序列测定及分析[J].中华预防医学杂志, 2020, (05): 486-490

[22] 李夫,潘阳,王雅杰,等.北京市 12 例境外输入新冠肺炎病例病毒基因组特征分析[J].国际病毒学杂志, 2020, 27(03): 177-181

[23] Zhang Chi, Xiu Le-shan, Li Ya-mei, et al. Multiplex PCR and Nanopore Sequencing of Genes Associated with Antimicrobial Resistance in Neisseria gonorrhoeae Directly from Clinical Samples [J]. Clinical chemistry, 2020[Epub ahead of print]

[24] Nikki E. Freed, Markéta Vlková, Muhammad B. Faisal, et al. Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore Rapid Barcoding [J]. Biology methods & protocols, 2020, 5(1)[Epub ahead of print]

[25] Jenjaroenpun Piroon, Wanchai Visanu, OnoMoore Kikumi D, et al. Two SARS-CoV-2 Genome Sequences of Isolates from Rural U.S. Patients Harboring the D614G Mutation, Obtained Using Nanopore Sequencing [J]. Microbiology resource announcements, 2020, 10(1) [Epub ahead of print]

[26] Li J, Wang H, Mao L, et al. Rapid genomic characterization of SARS-CoV-2 viruses from clinical specimens using nanopore sequencing[J]. Sci Rep, 2020, 10(1): 17492

[27] To Kelvin Kai Wang, Hung Ivan Fan Ngai, Ip Jonathan Daniel, et al. COVID-19 re-infection by a phylogenetically distinct SARS-coronavirus-2 strain confirmed by whole genome sequencing [J]. Clinical infectious diseases: an official publication of the Infectious Diseases Society of America, 2020[Epub ahead of print]

[28] Ming Wang, Aisi Fu, Ben Hu, et al. Nanopore Targeted Sequencing for the Accurate and Comprehensive Detection of SARS-CoV-2 and Other Respiratory Viruses [J]. Small, 2020, 16 (32)[Epub ahead of print]