

RESEARCH ARTICLE

Assessment of two-pool multiplex long-amplicon nanopore sequencing of SARS-CoV-2

Hongjie Liu¹ | Jinhui Li² | Yanfeng Lin^{2,3} | Xiaochen Bo¹ | Hongbin Song² |
Kuibiao Li⁴ | Peng Li² | Ming Ni¹ 

¹Department of Biotechnology, Beijing Institute of Radiation Medicine, Beijing, China

²Department of Bio-security, Chinese PLA Center for Disease Control and Prevention, Beijing, China

³Graduate School, Academy of Military Medical Sciences, Beijing, China

⁴Laboratory of Clinical Microbiology, Guangzhou Center for Disease Control and Prevention, Guangzhou, Guangdong, China

Correspondence

Peng Li, Department of Bio-security, Chinese PLA Center for Disease Control and Prevention, No. 20, Dongda Road, Beijing 100071, China.

Email: jiekenlee@126.com

Ming Ni, Department of Biotechnology, Beijing Institute of Radiation Medicine, No. 27, Taiping Road, Beijing 100850, China.

Email: niming@bmi.ac.cn

Funding information

Mega-projects of Science and Technology Research; Beijing Nova Program

Abstract

Genomic surveillance of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) plays an important role in COVID-19 pandemic control and elimination efforts, especially by elucidating its global transmission network and illustrating its viral evolution. The deployment of multiplex PCR assays that target SARS-CoV-2 followed by either massively parallel or nanopore sequencing is a widely-used strategy to obtain genome sequences from primary samples. However, multiplex PCR-based sequencing carries an inherent bias of sequencing depth among different amplicons, which may cause uneven coverage. Here we developed a two-pool, long-amplicon 36-plex PCR primer panel with ~1000-bp amplicon lengths for full-genome sequencing of SARS-CoV-2. We validated the panel by assessing nasopharyngeal swab samples with a <30 quantitative reverse transcription PCR cycle threshold value and found that ≥90% of viral genomes could be covered with high sequencing depths (≥20% mean depth). In comparison, the widely-used ARTIC panel yielded 79%–88% high-depth genome regions. We estimated that ~5 Mbp nanopore sequencing data may ensure a >95% viral genome coverage with a ≥10-fold depth and may generate reliable genomes at consensus sequence levels. Nanopore sequencing yielded false-positive variations with frequencies of supporting reads <0.8, and the sequencing errors mostly occurred on the 5' or 3' ends of reads. Thus, nanopore sequencing could not elucidate intra-host viral diversity.

KEYWORDS

genome, multiplex polymerase chain reaction, nanopore sequencing, SARS-CoV-2, viral

1 | INTRODUCTION

Since the first whole genome of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was reported in January 2020,¹ numerous genome sequences have been obtained and shared by researchers worldwide. As of January 15, 2021, over 400 000 SARS-CoV-2 genome sequences had been detected and collected in over 200 countries and had been deposited in public databases such as GISAID.² Consequently, the global spread and phylodynamics of SARS-CoV-2 have been under surveillance,^{3–7} and lineages have

been designated.^{8,9} Moreover, genome sequencing has greatly facilitated the tracking of SARS-CoV-2 evolution, especially the emergence of variants with enhanced infectivity and transmissibility, such as viruses that express D614G in the spike protein^{10,11} and recent variants of concern such as Alpha, Beta, Gamma, and Delta.¹²

Compared with methods that focus on specific regions of the viral genome, such as real-time quantitative reverse transcription PCR (qRT-PCR)¹³ and isothermal amplification,^{14–17} full-length sequencing of primary samples provides the most comprehensive genomic information.^{7,18} The deployment of virus-specific multiplex

PCR followed by sequencing is a widely-used strategy to obtain viral genomes. PCR panels for SARS-CoV-2 usually have short amplicons. For example, the ARTIC network (<https://artic.network/ncov-2019>) proposed a SARS-CoV-2 panel with ~400-bp amplicons. For amplicons that are shorter than the read lengths obtained by massively parallel sequencing (MPS) devices such as Illumina MiSeq, it is unnecessary to fragment the PCR products in MPS library preparation. However, because of the inevitability of biased PCR efficiency among different amplicons, the use of multiplex PCR pools is likely to generate uneven coverage. For full-genome sequencing of SARS-CoV-2, distributing primers in multiple pools^{19,20} or amplifying each amplicon separately²¹ might reduce bias, but increase labor and economic cost.

One approach to improve the coverage evenness of multiplex PCR assays is to reduce the number of primers per panel. In our previous study, to recover Ebola virus genomes from clinical samples, two panels with ~1000-bp and ~500-bp amplicon sizes were implemented.^{22,23} A long-amplicon panel is preferred as it may confer higher coverage and evenness, and if it fails in the evaluation of highly degraded samples, a short-amplicon panel may be used. Long-amplicon panels are well-suited to the Oxford Nanopore MinION apparatus, which can generate ultra-long reads. MinION generates long reads and can be implemented outside conventional laboratories.^{24,25} It provides an important supplement to MPS devices and has been used to sequence the SARS-CoV-2 genome.^{20,26}

In this study, we developed a new two-pool 36-plex panel for SARS-CoV-2 with ~1000-bp tiling amplicons that can generate high coverage evenness of SARS-CoV-2 genomes. Consequently, more samples may be sequenced in a MinION flowcell, thus greatly reducing sequencing costs.

2 | MATERIALS AND METHODS

2.1 | SARS-CoV-2 nucleic acid testing

qRT-PCR assays of oropharyngeal swab specimens were performed to confirm SARS-CoV-2 infection. Characteristics of samples included in this study are provided in Table S1. Total RNA was isolated from a viral transport medium containing oropharyngeal swab specimens (TIANamp Virus DNA/RNA Kit; TIANGEN). qRT-PCR was performed with the Novel Coronavirus (2019-nCoV) Nucleic Acid Detection Kit (BioGerm) according to the manufacturer's instructions on the CFX96 Real-Time System (BIO-RAD).

2.2 | Primer design and synthesis

PCR primers for SARS-CoV-2-specific multiplex amplification were designed by using the Primer Scheme v1.3.2 tool (<http://github.com/aresti/primalscheme>),²⁴ using the reference genome of SARS-CoV-2 Wuhan-hu-1 strain (GenBank accession MN908947.3). Amplicon sizes were set at 1000-bp and 2000-bp. A ~3000-bp amplicon primer panel was selected manually and re-paired from the ARTIC V3 primer

panel (<http://artic.network/ncov-2019>), in which primers with high-level bias were excluded. Primers were synthesized by Sangon Biotech. Primer pairs of the 1000-bp, 2000-bp, and 3000-bp panels are presented in Tables S2–S4.

2.3 | Multiplex amplification

Extracted total RNA was first reverse-transcribed to complementary DNA (cDNA) using a NEBNext Ultra II RNA First-Strand Synthesis Module (New England Biolabs). The generated first-strand cDNA was used as the template for SARS-CoV-2-specific amplification with different multiplex PCR primer panels and NEBNext High-Fidelity 2X PCR Master Mix (New England Biolabs) following the manufacturer's instructions. The thermal cycling regimen for the multiplex PCR primer panels followed the ARTIC protocol: 30 s at 98°C; 35 cycles of 15 s at 98°C and 5 min at 65°C; then held at 4°C. Primers were then pooled and cleaned with the GeneJET Gel Extraction Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. PCR-negative controls were used during amplification. The PCR products were analyzed with the Qsep 100 capillary electrophoresis (CE) system (BiOptic).

2.4 | Library preparation for MPS

PCR products obtained with the ARTIC primer panel had an amplicon size of ~400-bp, and were prepared for MPS libraries without fragmentation with NEBNext Ultra II FS DNA Library Prep Kit for Illumina (New England Biolabs). Products with amplicon sizes of ~1000-bp, 2000-bp, and 3000-bp were, respectively, first fragmented enzymatically for 20 min at 37°C and 30 min at 65°C, and then libraries with barcodes were prepared using NEBNext Ultra II FS DNA Library Prep Kit for Illumina (New England Biolabs). The libraries without fragmentation (insert sizes ~400-bp) were sequenced by using the MiSeq system (Illumina) to generate 2 × 300 bp paired-end reads. The libraries with fragmentation were sequenced by using the MiSeq system or NovaSeq. 6000 system (Illumina) to generate 2 × 150 bp paired-end reads. MPS using the NovaSeq. 6000 system was performed by Berry Genomics Co. Ltd.

2.5 | Library preparation for MinION nanopore sequencing

For nanopore sequencing, cDNA was treated with NEBNext Ultra II End repair/dA-tailing Module (New England Biolabs) and ligated with barcodes from a native barcoding kit (Oxford Nanopore Technologies) with NEB Blunt/TA Ligase Master Mix (New England Biolabs). The adapter from the ligation sequencing kit (Oxford Nanopore Technologies) was then ligated using the NEBNext Quick Ligation Module (New England Biolabs). Sequencing was then performed using the MinION Mk1B sequencer (Oxford Nanopore Technologies) with

either the regular R9.4.1 or a smaller Flongle flow cell (FLO-FLG001, R9.4.1) for 24 h. Guppy v3.2.1 (Oxford Nanopore Technologies) was employed for high-accuracy model base-calling and demultiplexing according to ARTIC network recommendations (<http://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html>).

2.6 | Bioinformatics analysis

The bioinformatics analysis of MPS data was similar to that used in our previous studies.^{22,27,28} Primer trimming of reads was performed with iVar v1.0.²⁹ Reads were then aligned to the reference genome (Wuhan-hu-1 strain, GenBank accession MN908947.3) by using BWA mem v0.7.17.³⁰ The alignments were then analyzed with SAMtools v1.9³¹ to obtain a sequencing depth file and “mpileup” formatted files. A previously developed homemade workflow named “iSNV-calling” (<http://github.com/generaliti/iSNV-calling>) was implemented to identify viral single nucleotide variations (SNVs) with requirements of \geq Q20 base quality, \geq 100-fold depth, and \geq 20% reads supporting each SNV. Based on our previous assessment,³² these bioinformatics workflows and filters can identify viral SNVs reliably during amplicon sequencing.

For MinION nanopore sequencing data, reads with the desired length (between 750 and 1100 bases for the 1K-panel amplification) were selected and trimmed with a start and end of 30 bases. We used both NGMLR v0.2.7³³ and Minimap2 v2.21³⁴ for alignment with the setting of “-x ont” and “-ax map-ont”, respectively. Depth profiles of sequencing reads and mpileup format alignments, based on NGMLR and Minimap2, respectively, were generated with SAMtools v1.9. SNVs were identified by using the homemade “iSNV-calling” workflow with requirements of \geq Q20 base quality, \geq 100-fold depth, and \geq 20% reads supporting the SNVs. For comparison, the bioinformatics workflow proposed by Bull et al.²¹ was also performed in parallel. Briefly, nanopore sequencing reads were aligned by using Minimap2 v2.21. VarScan2 v2.4.4³⁵ was employed for identification of SNV with requirements of \geq Q20 base quality, \geq 100-fold depth, and \geq 20% reads supporting the variant.

3 | RESULTS

3.1 | Determining the optimal amplicon length of multiplex PCR panels

We sought to improve the coverage evenness of the multiplex PCR panel targeting SARS-CoV-2 by increasing amplicon sizes. To find a favorable amplicon size, we designed and synthesized three two-pool panels consisting of \sim 1,000-bp, \sim 2,000-bp, and \sim 3,000-bp amplicons (Tables S2–S4). One nasopharyngeal swab sample with a C_t value of 25 was used for preliminary validation of the three panels. The size distribution of multiplex PCR products was analyzed with CE. The expected PCR lengths were obtained for the 1000-bp and 2000-bp panel, but not for the 3000-bp panel (Figure S1). PCR products were

then sequenced with MinION. Based on the alignment of sequencing reads with the reference genome (Wuhan-hu-1 strain), both the 1000-bp and 2000-bp amplicon panels generated full coverage of the viral genome. However, the 3000-bp panel failed to generate full coverage of the viral genome, which was consistent with the CE analysis and could be ascribed to RNA/cDNA degradation. The 2000-bp panel had a much larger coverage bias among amplicons than the 1000-bp panel (Figure S2). For the 2000-bp panel, 30.6% of sequencing data were assigned to one amplicon. Therefore, we determined that the \sim 1000-bp amplicons were preferable for the multiplex amplification panel of the SARS-CoV-2 genome. The 1000-bp panel contained 36 primer pairs in two pools (18 pairs each). Amplicon sizes ranged from 880-bp to 1027-bp with an average overlap of 112 bp. We thereafter referred to the panel with \sim 1000-bp amplicons as the 1K-panel.

3.2 | Coverage evenness of the 1K-panel

We next compared coverage evenness by using the 1K-panel and a widely used 98-plex primer panel provided by the ARTIC network (<http://artic.network/ncov-2019>, version V3). RNA was extracted from six nasopharyngeal swabs with varied SARS-CoV-2 titers (C_t values, 22.9–31.0). Aliquots were amplified with the 1K- and ARTIC panels, followed by sequencing using MiSeq. We defined viral genome regions with \geq 20% mean depth as high-coverage regions. The proportion of the high-coverage regions was used to quantify coverage evenness.

We found that the 1K-panel generated a more even sequence coverage than the ARTIC panel (Figure 1A). With the 1K-panel, the proportion of high-coverage regions averaged 93.0% ($SD = 1.7\%$) for the six samples, compared with 80.6% ($SD = 8.7\%$) for the ARTIC panel. Coverage evenness was dependent on sample viral titers. We found that the 1K-panel was less affected by low viral titers than the ARTIC panel (Figure 1B). As C_t values increased, the proportions of the high-coverage regions decreased slightly from 95.3% to 90.5% for the 1K-panel, compared with 88.1%–62.4% for the ARTIC panel. We also evaluated a higher cutoff (\geq 30% mean depth) to define high-coverage regions and found that the 1K-panel maintained its advantage (Figure 1B).

Next, we included another 49 nasopharyngeal swabs to further evaluate the performance of the 1K-panel to recover the viral genome. We observed a comparable efficiency of viral amplification and genome coverage evenness (Figure 1C). Among the 29 samples with C_t values $<$ 30, all except one had a viral genome recovery with \geq 90% high-coverage regions (mean 93.8%, $SD = 8.1\%$). Evenness decreased for samples with C_t values $>$ 30, but five of six samples with a 30–33 C_t value had $>$ 70% high-coverage regions.

3.3 | Assessment of long-amplicon nanopore sequencing

Recent studies have shown that nanopore sequencing can generate accurate consensus genomes of SARS-CoV-2, but are error-prone to

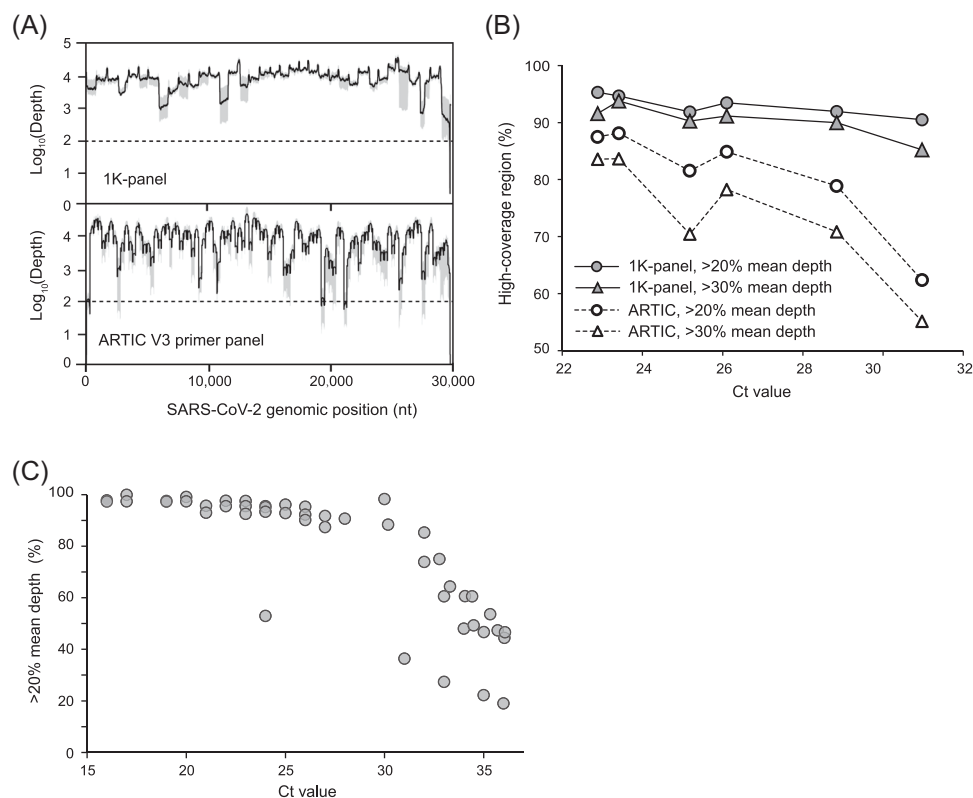


FIGURE 1 Coverage evenness of SARS-CoV-2 genome of multiplex PCR mix panels. (A) Coverage depth profile of six clinical samples for the 1K-panel (~1000-bp amplicons) and the ARTIC V3 primer panel (~400-bp amplicons). Depth is in standard logarithmic scale, with 100-fold denoted by the dashed line. Bold black lines denote median depth, and grey shadows indicate the minimum and maximum depths. Mean depth is defined as the average site depth on the targeted viral genome (excluding the 3' and 5' ends). (B) Two thresholds, >20% and >30% mean depth, were used to define regions of high coverage. Viral titers were quantified by the qRT-PCR C_t value of SARS-CoV-2 detection. (C) The correlation between coverage evenness and viral titers of samples for the 1K-panel. The proportions of genome regions with a >20% mean depth for 49 clinical samples. SARS-CoV-2, severe acute respiratory syndrome coronavirus 2

TABLE 1 Sequencing of samples with the amplification by using the 1K-panel

Sample	C_t value	MinION data (Mbp)	MinION coverage			MiSeq data (Mbp)	MiSeq coverage		
			Mean	≥100X	≥1000X		Mean	≥100X	≥1000X
S1	22.88	483.7	13 794	100	97.62	335.0	9733	100	97.55
S2	23.41	369.1	10 433	100	96.55	320.2	9088	100	99.39
S3	25.18	508.4	14 361	100	93.23	339.0	9852	99.92	95.66
S4	26.1	453.4	12 830	100	96.11	266.2	7604	99.93	96.24
S5	28.84	363.4	10 356	100	91.94	315.2	9054	99.94	93.04
S6	30.97	429.6	9718	97.77	91.75	336.3	7590	98.8	91.83
S7	31.58	423.3	11 983	97.74	89.37	263.8	7644	99.22	90.29
S8	33.15	435.5	12 234	95.57	81.65	307.9	8992	99.76	85.63

detect sub-clonal SNVs.^{20,21,36} Thus, we focused on assessing noise levels to identify SNVs, especially sub-clonal variations, via 1K-panel amplification followed by MinION sequencing. We sequenced the PCR products of the 1K-panel from eight samples by using both MinION (ligation-based kit and R9.4.1 flowcell) and MiSeq (enzymatic

fragmentation and pair-end 2 × 150 bp reads). High depth coverage, required to identify sub-clonal SNVs, was obtained by both sequencing devices (Table 1). We then implemented a homemade bioinformatics workflow based on piled MiSeq sequencing bases to identify SNVs. We included SNVs with a ≥0.2 frequency of

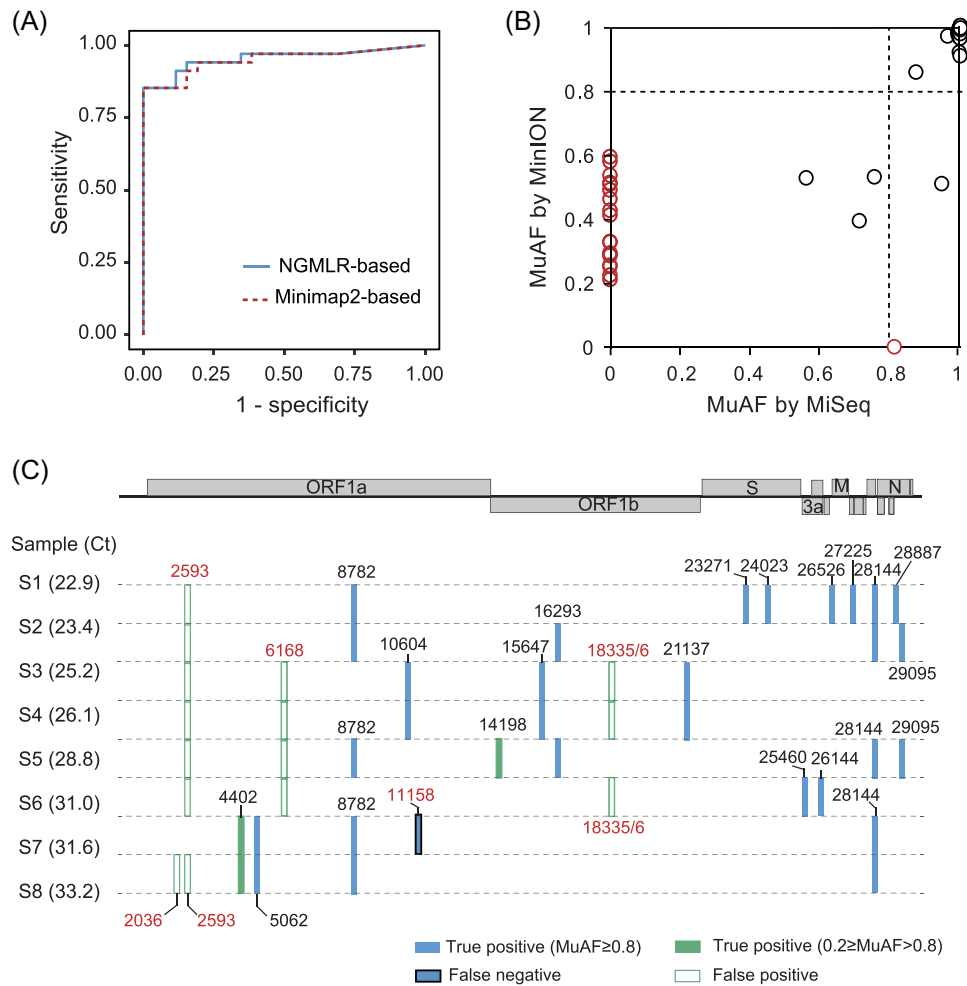


FIGURE 2 The identification of SARS-CoV-2 SNVs. (A) The ROC curves for SNV identification with bioinformatics workflows are respectively based on alignment tools NGMLR and Minimap2. Eight samples were included. Amplification was performed with the 1K-panel and MinION was used for sequencing. (B) SNVs were identified through the MiSeq- and MinION-based sequencing, with the MuAFs shown. SNVs that were not consistent between MinION and MiSeq sequencing are denoted in red. (C) The distribution of SNVs was identified by MinION sequencing. The true positive, false negative, and false positive SNVs were benchmarked by these based on MiSeq sequencing. MuAF, frequency of supporting reads for the mutated allele. Reference genome, Wuhan-hu-1 strain (GenBank accession MN908947.3). ROC, receiver operating characteristic; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; SNV, single nucleotide variation

supporting reads for the mutated allele (mutated allele frequency [MuAF]) for assessment (reference genome, Wuhan-hu-1). With this cutoff, the MiSeq-based identification of SNVs should be reliable, as previous studies have shown.^{32,37} Among the eight samples, a total of 34 SNVs were recognized by MiSeq. Three of them had a moderate MuAF within 0.2–0.8, and the remaining were regarded as SNVs at the consensus sequence level (MuAF \geq 0.8).

The SNVs detected by using MinION sequencing data had a high false-positive rate, and we observed that the artificial SNVs were dependent on the alignment tools we implemented. Based on the alignments of MinION reads respectively generated with NGMLR and Minimap2, we identified 18 artificial SNVs for each tool, but only ten of them were in common (Table S5). The difference of the artificial SNVs was irrelevant to the downstream workflows after alignment to detecting SNVs (the homemade workflow and VarScan2, Table S5). We obtained the receiver operating characteristic curves respectively

based on NGMLR and Minimap2 alignments (Figure 2A), and their area under curve values were 0.953 and 0.949, respectively, indicating their efficacy of SNV identifications were comparable. To address how these artificial SNVs occurred and differed based on the two alignment tools, we manually examined the alignment of reads harboring the errors. We found that all artificial SNVs, except one (G2036T in sample S8), were located on the 5' and 3' ends of MinION reads. Thus, these sub-clonal false-positive SNVs might be ascribed to noisy end regions of nanopore sequencing reads, and whether the ends of MinION-based reads were clipped in the alignment, somewhat depended on the software tools. We provide six examples to illustrate the occurrence of false-positive sub-clonal SNVs in Figure S3.

We illustrated the SNVs identified by MinION and NGMLR-based alignments of the eight samples in Figure 2B,C, benchmarked by the SNVs recognized by MiSeq. The 18 artificial SNVs had MuAFs ranged from 0.22 to 0.59 (mean: 0.39, median: 0.37), which were

indistinguishable for the three true positive SNVs with a moderate MuAF. This result suggests that the sub-clonal SNVs could not be identified reliably based on MinION sequencing. In contrast, all SNVs at the consensus sequence level (MuAF ≥ 0.8) could be reliably identified via MinION sequencing, except one (T11158C in sample S7) was filtered due to insufficient depth (51-fold).

3.4 | Genome recovery with low output of nanopore sequencing reads

As the 1K-panel generated an even sequencing coverage and the detection of sub-clonal SNVs was unnecessary, high MinION data amount per sample (300–500 Mbp) seemed excessive. Therefore, to test the limit of the 1K-panel to recover the SARS-CoV-2 genome, we re-sequenced the eight samples with an ultra-low throughput by using a MinION Flongle flowcell R9.4.1. The majority of viral genomes were recovered with sequencing reads of 1.21–6.14 Mbp data per sample (Figure 3). With cutoffs of ≥ 5 -fold sequencing depth and MuAF ≥ 0.8 , 29 (93.5%) of the 31 consensus SNVs were identified, and no false-positive SNVs were observed. The two SNVs (T11158C and T28144C) were not detected due to insufficient or absent coverage in sample S7, which had the least sequencing data (1.21 Mbp). Furthermore, we examined coverage evenness. In 7 of the 8 samples, $>90\%$ (90.5%–96.6%) of the viral genomes had been covered with a depth of $\geq 10\%$ mean depth, and for sample S8 the proportion was 81.6%. Thus, the coverage evenness by using the 1K-panel and MinION sequencing was not affected by extremely low data inputs.

4 | DISCUSSION

In countries where the outbreak appears to be leveling off, such as China, continual regional resurgences of COVID-19 have been observed.^{38–41} As the COVID-19 pandemic continues, full-genome-based surveillance of SARS-CoV-2 has become a routine approach to facilitate public health decision-making. Rapid recovery of viral genomes followed by phylogenetic analyses may determine viral lineages and elucidate introductions and transmission in local communities. Full genomes of newly emerged variants also provide data to further our understanding of SARS-CoV-2.

Multiplex PCR of SARS-CoV-2 followed by sequencing is the most widely-used strategy to obtain viral genomes from primary samples. The portable MinION sequencer provides field sequencing and could achieve near real-time genomic surveillance of SARS-CoV-2 during an outbreak. The high performance of the multiplex primer panel that targets SARS-CoV-2 is essential for the efficient recovery of viral genomes. A panel with a low amplification bias could provide high evenness of genome coverage. Low amplification bias reduces the risk of failure to identify lineage-determining variations, which is important to trace viral global phylogeny. Moreover, high genome coverage evenness dramatically decreases the per-sample sequencing cost. In our evaluation of the 1K-panel,

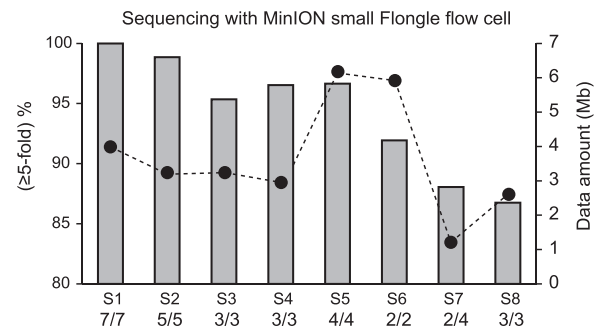


FIGURE 3 Nanopore sequencing of eight samples with a MinION Flongle flow cell. The bar plot denotes the proportions of viral genomes with a ≥ 5 -fold depth. The sequence data quantity per sample is shown by the dashed line and dot plot. The number of identified SNVs are shown below sample identifiers (i.e., 2/4 means two of four benchmarked SNVs were found). SNV, single nucleotide variation

~ 5 Mbp sequencing data generated accurate consensus genomes from samples with C_t values < 30 . Thus, we speculate that the sequencing of 96 or more barcoded samples per batch with one MinION flow cell is feasible, in which 20–30 Mbp data could be generated for each sample. This approach would be especially useful when large numbers of primary samples must be analyzed, such as for rapid lineage assignment of samples during epidemiologic and environmental surveillance.^{42,43}

To improve coverage evenness, we adopted ~ 1000 -bp amplicons, which are longer than those of several widely-used primer panels, such as ARTIC. Moore et al.¹⁹ recently proposed a multiplex primer panel with amplicons ranging from 956-bp to 1450-bp. However, their primer pairs were distributed in six multiplex pools, with an obvious bias of certain amplicons during amplification, which induced a more uneven coverage of the viral genome than the 1K-panel. Another major concern of the long-amplicon panel is its utility for evaluating clinical samples, as viral RNA is degraded easily. Although we did not observe a reduced efficiency of viral amplification among the clinical samples in this study, it should be noted that all samples in this study were tested within 2 weeks of collection. Compared with the short-amplicon panel, the long-amplicon panel would be more affected by RNA/cDNA degradation, especially when evaluating samples subjected to prolonged storage. Quality control for RNA integrity before amplification could be helpful, but requires additional equipment and procedures. We suggest that the long-amplicon panel could be used for the first round of viral amplification and that samples that fail to yield sufficient PCR products could be amplified further with a short-amplicon panel.

Consensus viral genomes can be accurately recovered by following 1K-panel amplification with MinION sequencing, but identifying viral sub-clonal variation is infeasible. Based on our assessment, a MuAF > 0.8 cutoff is more reliable to define consensus SNVs through MinION sequencing. However, false-positive SNVs with a moderate MuAF (0.2–0.6) were prevalent, and many were shared among samples, reflecting the characteristic systematic nature of nanopore sequencing errors.^{44,45}

The occurrence of artificial SNVs was also related to alignment software, and artificial SNVs were enriched on the 5' and 3' ends of nanopore sequencing reads. Thus, it provides a clue for further improvement of bioinformatic analyses for sub-clonal SNV identification. Currently, nanopore sequencing might not be applicable to studies that aim to analyze the intra-host diversity of viral genomes; MPS would be necessary.

In sum, we developed and validated a new two-pool, long-amplicon 36-plex PCR primer panel for the full-genome sequencing of SARS-CoV-2 from primary samples. The panel may generate a more even coverage of SARS-CoV-2 genomes than the ARTIC short-amplicon panel and subsequently may require less sequencing data. For the samples with a $C_t < 30$, ~5 Mbp data by MinION provided $\geq 95\%$ genome coverage with a ≥ 10 -fold depth. Meanwhile, our assessment showed that nanopore sequencing with MinION identified dominant viral populations (consensus viral genomes) reliably within samples, but was highly error-prone for the discovery of minor viral populations. SNVs with a < 0.8 MuAF should be regarded as unreliable.

ACKNOWLEDGEMENTS

We would like to thank Ms. Yan Zhang for revising this manuscript. Ming Ni and Peng Li were supported by the Beijing Nova Program (Z181100006218114 and Z181100006218110). This work was partially supported by grants from the Mega-projects of Science and Technology Research (No. 2018ZX10201001-003 and No. 2018ZX10305410-004).

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

AUTHOR CONTRIBUTIONS

Ming Ni and Peng Li designed this study. Hongbin Song supervised the study. Kuibiao Li, Yanfeng Lin and Peng Li collected the samples and clinical information. The primer design and multiplex viral amplification were performed by Jinhui Li and Yanfeng Lin. Hongjie Liu and Ming Ni analyzed the data, then created tables and figures. Ming Ni, Peng Li and Hongjie Liu drafted the manuscript. Xiaochen Bo revised this manuscript. All authors commented on previous versions of the manuscript and approved the final manuscript. All authors contributed to the revision of the manuscript and approved the submitted version.

ETHICS STATEMENT

The study was approved by the Ethics Committee of the Center for Disease Control and Prevention (CDC) of Guangzhou (GZCDC-ECHR-2020P0002), China. Written informed consent was obtained from patients regarding surveillance and data analysis. No identifiable participant data were included in this study. This study was supervised by both the Institutional Review Boards of the Guangzhou CDC and the Chinese PLA CDC, and was performed in Biosafety Level 2 with Enhanced Practices (BSL-2+) laboratories.

DATA AVAILABILITY STATEMENT

The codes for bioinformatics analyses are available at <https://github.com/Ming-Ni-Lab/1K-amplicon-ONT-sequencing-of-SARS-CoV-2>.

The consensus genomes of SARS-CoV-2 are available at the GISAID

database (<https://gisaid.org>), GenBank (<https://www.ncbi.nlm.nih.gov/genbank>) and the China National Microbiology Data Center (<http://www.nmdc.cn/coronavirus>) databases through the accession numbers provided in the Table S1.

ORCID

Ming Ni  <http://orcid.org/0000-0001-9465-2787>

REFERENCES

1. Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. *Nature*. 2020;579:265-269.
2. Shu YL, McCauley J. GISAID: global initiative on sharing all influenza data—from vision to reality. *Euro Surveill*. 2017;22(13):2-4.
3. Bedford T, Greninger AL, Roychoudhury P, et al. Cryptic transmission of SARS-CoV-2 in Washington state. *Science*. 2020;370(6516):571-575.
4. Eden JS, Rockett R, Carter I, et al. An emergent clade of SARS-CoV-2 linked to returned travellers from Iran. *Virus Evol*. 2020;6(1):veaa027.
5. Gámbaro F, Behillil S, Baidaliuk A, et al. Introductions and early spread of SARS-CoV-2 in France, 24 January to 23 March 2020. *Euro Surveill*. 2020;25(26):2001200.
6. Hadfield J, Megill C, Bell SM, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*. 2018;34(23):4121-4123.
7. Oude Munnink BB, Nieuwenhuijse DF, Stein M, et al. Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. *Nat Med*. 2020;26(9):1405-1410.
8. Rambaut A, Holmes EC, O'Toole Á, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol*. 2020;5(11):1403-1407.
9. Tang X, Wu C, Li X, et al. On the origin and continuing evolution of SARS-CoV-2. *Natl Sci Rev*. 2020;7(6):1012-1023.
10. Baric RS. Emergence of a Highly Fit SARS-CoV-2 Variant. *N Engl J Med*. 2020;383(27):2684-2686.
11. Korber B, Fischer WM, Gnanakaran S, et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 Virus. *Cell*. 2020;182(4):812-827.e819.
12. WHO. Tracking SARS-CoV-2 variants. 2021; <https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/>. Accessed June 25, 2021
13. Zhang K, Fan C, Cai D, et al. Contribution of TGF-beta-mediated NLRP3-HMGB1 activation to tubulointerstitial fibrosis in rat with angiotensin ii-induced chronic kidney disease. *Front Cell Dev Biol*. 2020;8:1.
14. Carter LJ, Garner LV, Smoot JW, et al. Assay techniques and test development for COVID-19 Diagnosis. *ACS Cent Sci*. 2020;6(5):591-605.
15. Kim Y, Yaseen AB, Kishi JY, et al. Single-strand RPA for rapid and sensitive detection of SARS-CoV-2 RNA. *MedRxiv*. 2020:20177006. <https://doi.org/10.1101/2020.08.17.20177006>
16. Xia S, Chen X. Single-copy sensitive, field-deployable, and simultaneous dual-gene detection of SARS-CoV-2 RNA via modified RT-RPA. *Cell Discov*. 2020;6:37.
17. Xue G, Li S, Zhang W, et al. Reverse-transcription recombinase-aided amplification assay for rapid detection of the 2019 novel coronavirus (SARS-CoV-2). *Anal Chem*. 2020;92(14):9699-9705.
18. Böhmer MM, Buchholz U, Corman VM, et al. Investigation of a COVID-19 outbreak in Germany resulting from a single travel-associated primary case: a case series. *Lancet Infect Dis*. 2020;20(8):920-928.
19. Moore SC, Penrice-Randal R, Alruwaili M, et al. Amplicon-based detection and sequencing of SARS-CoV-2 in nasopharyngeal swabs

- from patients with COVID-19 and identification of deletions in the viral genome that encode proteins involved in interferon antagonism. *Viruses*. 2020;12:10.
20. Paden CR, Tao Y, Queen K, et al. Rapid, sensitive, full-genome sequencing of severe acute respiratory syndrome coronavirus 2. *Emerg Infect Dis*. 2020;26(10):2401-2405.
 21. Bull RA, Adikari TN, Ferguson JM, et al. Analytical validity of nanopore sequencing for rapid SARS-CoV-2 genome analysis. *Nat Commun*. 2020;11(1):6272.
 22. Ni M, Chen C, Qian J, et al. Intra-host dynamics of Ebola virus during 2014. *Nat Microbiol*. 2016;1(11):16151.
 23. Tong YG, Shi WF, Liu D, et al. Genetic diversity and evolutionary dynamics of Ebola virus in Sierra Leone. *Nature*. 2015;524(7563):93-96.
 24. Quick J, Grubaugh ND, Pullan ST, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc*. 2017;12(6):1261-1276.
 25. Quick J, Loman NJ, Duraffour S, et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature*. 2016;530(7589):228-232.
 26. Lu J, du Plessis L, Liu Z, et al. Genomic epidemiology of SARS-CoV-2 in Guangdong Province, China. *Cell*. 2020;181(5):997-1003.
 27. Zhang Y, Yin Q, Ni M, et al. Dynamics of HIV-1 quasispecies diversity of participants on long-term antiretroviral therapy based on intrahost single-nucleotide variations. *Int J Infect Dis*. 2021;104:306-314.
 28. Chen C, Jiang D, Ni M, et al. Phylogenomic analysis unravels evolution of yellow fever virus within hosts. *PLoS Negl Trop Dis*. 2018;12(9):e0006738.
 29. Grubaugh ND, Gangavarapu K, Quick J, et al. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol*. 2019;20:20.
 30. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760.
 31. Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-2079.
 32. Ni M, Chen C, Liu D. An assessment of amplicon-sequencing based method for viral intrahost. *Analysis Virol Sin*. 2018;33(6):557-560.
 33. Sedlazeck FJ, Rescheneder P, Smolka M, et al. Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods*. 2018;15(6):461-468.
 34. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34(18):3094-3100.
 35. Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 2012;22(3):568-576.
 36. Lu J, du Plessis L, Liu Z, et al. Genomic epidemiology of SARS-CoV-2 in Guangdong Province, China. *Cell*. 2020;181(5):997-1003.e1009
 37. Schirmer M, Ijaz UZ, D'Amore R, Hall N, Sloan WT, Quince C. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res*. 2015;43(6):37.e37.
 38. Du P, Ding N, Li J, et al. Genomic surveillance of COVID-19 cases in Beijing. *Nat Commun*. 2020;11(1):5503.
 39. Jia HL, Li P, Liu HJ, et al. Genomic elucidation of a COVID-19 resurgence and local transmission of SARS-CoV-2 in Guangzhou, China. *J Clin Microbiol*. 2021;59:0007921.
 40. Lythgoe KA, Hall M, Ferretti L, et al. SARS-CoV-2 within-host diversity and transmission. *Science*. 2021;372(6539):eabg0821.
 41. Zhang Y, Pan Y, Zhao X, et al. Genomic characterization of SARS-CoV-2 identified in a reemerging COVID-19 outbreak in Beijing's Xinfadi market in 2020. *Biosaf Health*. 2020;2(4):202-205. <https://doi.org/10.1016/j.bsheal.2020.08.006>
 42. Xinghuo P, Lili R, Shuangsheng W, et al. Cold-chain food contamination as the possible origin of COVID-19 resurgence in Beijing. *Natl Sci Rev*. 2020. <https://doi.org/10.1093/nsr/nwaa264>
 43. Song ZG, Chen YM, Wu F, et al. Identifying the Risk of SARS-CoV-2 Infection and Environmental Monitoring in Airborne Infectious Isolation Rooms (AIIRs). *Virol Sin*. 2020;35:785-792.
 44. Kono N, Arakawa K. Nanopore sequencing: review of potential applications in functional genomics. *Dev Growth Differ*. 2019;61(5):316-326.
 45. Jain M, Olsen HE, Paten B, Akeson M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol*. 2016;17(1):239.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Liu H, Li J, Lin Y, et al. Assessment of two-pool multiplex long-amplicon nanopore sequencing of SARS-CoV-2. *J Med Virol*. 2022;94:327-334. <https://doi.org/10.1002/jmv.27336>