

Two-Pool Multiplex Long-Amplicon Amplification and Full-Genome Sequencing of SARS-CoV-2

Hong-jie Liu

Beijing Institute of Radiation Medicine

Jin-Hui Li

Chinese PLA Center for Disease Control and Prevention

Yan-Feng Lin

Chinese Center for Disease Control and Prevention

Xiao-Chen Bo

Beijing Institute of Radiation Medicine

Hong-Bin Song

Chinese PLA Center for Disease Control and Prevention

Peng Li

Chinese PLA Center for Disease Control and Prevention

Ming Ni (✉ ni.ming@163.com)

Beijing Institute of Radiation Medicine <https://orcid.org/0000-0001-9465-2787>

Research

Keywords: SARS-CoV-2, Multiplex Polymerase Chain Reaction, Genome, Viral, Nanopore Sequencing

Posted Date: November 10th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-103455/v1>

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

[Read Full License](#)

Abstract

Background

Since the first public genome of SARS-CoV-2, over 170,000 genome sequences of the virus have been shared by researchers world-wide (till November 1st 2020). Multiplex PCR targeting SARS-CoV-2 followed by massively parallel sequencing (MPS) and/or nanopore sequencing is a widely used strategy to recover the genome from primary samples. However, the bias of amplification among different amplicons should not be ignored, which might lead to uneven sequencing coverage on the viral genome.

Methods

We aim to develop a novel multiplex PCR panel to achieve an improved coverage evenness of SARS-CoV-2. We adapt long amplicons (~1000-bp) for the panel and thus reduced the number of primer pairs. The panel was validated with clinical samples and sequenced via MPS sequencing systems and a portable nanopore sequencing device MinION. We evaluated the full-genome coverage evenness and its dependence on viral loads of the long amplicon panel; we then compared it with a 98-plex panel provided by the ARTIC network. The accuracy to identify viral genomic variations based on the panel and sequencing with MinION was assessed.

Results

We developed a two-pool 36-plex panel for full-genome sequencing of SARS-CoV-2, whose amplicon size ranged from 880 to 1027 bp. For samples with a $<30 C_t$ value, $>90\%$ viral genome could be recovered with a high sequencing depth (>0.2 mean depth) by using the long-amplicon panel ($n = 36$), compared with 79-88% highly covered genome region for the ARTIC panel ($n = 5$). The coverage evenness of the long-amplicon panel was also less affected by low viral titers and not dependent on sequencing data amount. With MinION sequencing, the consensus viral genomes could be reliably recovered. However, a high false positive rate was observed to identify sub-clonal genomic variations with a <0.6 frequency.

Conclusion

A novel multiplex PCR panel for full-genome sequencing of SARS-CoV-2 with improved coverage evenness and low requirement of data throughput was validated with clinical samples. Amplification of SARS-CoV-2 with the panel followed by MinION sequencing could generate reliable consensus genome sequences, but the detection of non-dominating viral populations within host is error-prone.

Background

Genomic surveillance plays an important role in the control and prevention of the COVID-19 pandemic induced by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). As a novel pathogen, the whole genome of SARS-CoV-2 was rapidly revealed in January 2020 [1]. Thereafter, numerous genome sequences of SARS-CoV-2 were obtained and shared by researchers worldwide. Till November 1st 2020,

over 170,000 genome sequences of SARS-CoV-2, which were detected from samples collected in over 200 countries, have been deposited in public databases such as the GISAID [2]. Consequently, global spreading and phylodynamics of SARS-CoV-2 were under surveillance [3-7], and lineages of the virus were designated [8,9].

As the pandemic goes on, recovering the full viral genome from primary samples is useful to trace the source and spread of the viral variants as well as to monitor viral genetic variations, which greatly facilitates informed public health decision-making [7,10]. Viral specific multiplex PCR followed by sequencing is a widely-used strategy to recover the viral genome. Due to the concern of viral RNA degradation, PCR panels for SARS-CoV-2 usually have amplicons with a size <600 bp, shorter than the maximum read length of Illumina MiSeq sequencing apparatus. However, as the bias of PCR efficiency among different amplicons is inevitable, highly multiplex PCR in a pool is more likely to generate uneven coverage. Distributing primers in more pools might reduce the bias [11, 12] but it increases the labor and economic cost of the experiment.

To improve the sequencing evenness of the multiplex PCR panel, one approach is to reduce the number of primers in a panel. In our previous study to recover the genome of Ebola virus from clinical samples, two panels with respectively ~1000-bp and ~500-bp amplicon sizes were both implemented [13]. The panel with longer amplicons is favorable as it could have higher coverage and evenness, and if it failed for the highly degraded samples, the short-amplicon panel would be used. Long-amplicon panel is well-adaptive with the Oxford Nanopore MinION apparatus, which could generate long read and can be implemented outside conventional laboratories [14,15]. MinION provides an important alternative to massive parallel sequencing (MPS) devices and has been used for the sequencing of the SARS-CoV-2 genome [12,16].

In this study, we developed a new two-pool 36-plex panel for SARS-CoV-2 with ~1000-bp amplicons, which could generate high coverage evenness of SARS-CoV-2 genomes. Consequently, more samples could be sequenced in a MinION flowcell and thus greatly reduce the cost.

Methods

Nucleic acid test of SARS-CoV-2

Real-time quantitative reverse transcription PCR (qRT-PCR) assay was performed to confirm the infection of SARS-CoV-2 based on oropharyngeal swabs specimen. Total RNA was isolated from viral transport median containing oropharyngeal swabs specimen through TIANamp Virus DNA/RNA Kit (TIANGEN, Beijing, China). The qRT-PCR was performed with Novel Coronavirus (2019-nCoV) Nucleic Acid Detection Kit (BioGerm, Shanghai, China) according to the manufacturer's instructions on the CFX96 Real-Time System (BIO-RAD, USA).

Primer design and synthesis

The PCR primers for multiplex SARS-CoV-2 specific amplification were designed by using the Primer Scheme v1.3.2 tool (<http://github.com/aresti/primalscheme>) [14], with the reference genome of SARS-CoV-2 Wuhan-hu-1 strain (GenBank accession MN908947.3). The sizes of amplicons were respectively set as 1000-bp (-a 1000) and 2000-bp (-a 2000). The primer panel with a ~3000-bp amplicon size was manually selected and re-paired from the ARTIC V3 primer panel (<http://artic.network/ncov-2019>), in which the primers with great bias were excluded. The primers were synthesized by Sangon Biotech (Shanghai, China). The primer pairs of the 1000-bp, 2000-bp and 3000-bp panels were provided in the Additional file.

Multiplex viral amplification

Extracted total RNA was first reverse-transcribed to cDNA by using NEBNext Ultra II RNA First Strand Synthesis Module (New England Biolabs, USA). 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 regimen of thermal cycling for the multiplex PCR primer panels followed the ARTIC protocol: 30 sec at 98 °C; 35 cycles of 15 sec at 98 °C and 5 min at 65 °C; then hold at 4 °C. Pooled and cleaned with GeneJET Gel Extraction Kit (Thermo Fisher Scientific, USA) according to the manufacturer's instructions. PCR negative controls were used during the amplification.

Library preparation for sequencing by using MPS

The PCR products with the ARTIC primer panel had an amplicon size of ~400-bp and were prepared for libraries of MiSeq sequencing without fragmentation. The products with the longer amplicon sizes were enzyme-based fragmented for 20 min at 37 °C and 30 min at 65 °C. The libraries with barcodes were prepared by using NEBNext Ultra II FS DNA Library Prep Kit for Illumina (New England Biolabs). The MiSeq system (Illumina, USA) and NovaSeq 6000 system (Illumina) were used to generate paired-end reads.

Library preparation for nanopore sequencing by using MinION

For nanopore sequencing, the cDNA was treated with NEBNext Ultra II End repair/dA-tailing Module (New England Biolabs) and ligated with barcodes from native barcoding kit (Oxford Nanopore Technologies, ONT) with NEB Blunt/TA Ligase Master Mix (New England Biolabs). The adapter from the ligation sequencing kit (Oxford Nanopore Technologies, UK) was next ligated by using NEBNext Quick Ligation Module (New England Biolabs). Sequencing was then performed using the Nanopore sequencer MinION with the regular R9.4.1 flow cell or a smaller Flongle flow cell (FLO-FLG001, R9.4.1), for 24 hours. Guppy v3.2.1 (Oxford Nanopore Technologies) was employed for high-accuracy model basecalling and demultiplexing according to the recommendation of the ARTIC network (<http://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html>).

Bioinformatics analysis

The bioinformatics analysis of the data generated by MiSeq was similar to our previous studies [17]. Primer trimming from reads was performed with iVar v1.0 [18]. Then reads were aligned to the reference genome by using BWA mem v0.7.17-r1188 [19]. The alignments were then analyzed with SAMtools v1.9 [20] obtain sequencing depth and 'mpileup' formatted files. A homemade workflow named 'iSNV-calling' developed previously (<http://github.com/generalitiy/iSNV-calling>) was implemented to identify viral SNVs with requirements of >Q20 base quality, >100-fold depth, and >0.2 MuAFs.

For the data of MinION sequencing, the reads with a desired length (between 750 to 1100 bases for the 1K-panel amplification) were selected and trimmed with start and end 30 bases. NGMLR v0.2.7 [21] was used for alignment with the default setting. Depth profiles were generated with SAMtools v1.9, and SNV identification was conducted by using the home-made 'iSNV-calling' workflow.

Results

We sought to improve the coverage evenness of multiplex PCR panel targeting SARS-CoV-2 by increasing the size of amplicons. To find a favorable size of amplicons, we designed and synthesized three two-pool panels respectively consisted of ~1,000-bp, and ~2,000-bp, and ~3,000-bp amplicon (Additional file). One nasopharyngeal swab sample was used for preliminary validation of the three panels. The multiplex PCR products were sequenced with MinION. The result shows that both the panels of 1000-bp and 2000-bp amplicons could obtain full viral genome, but the amplification of the 3000-bp panel failed due to that most RNA/cDNA degraded to short fragments. Then, the coverage profiles of the 1000-bp and 2000-bp panel were compared. The 2000-bp panel had a much larger bias of coverage among amplicons than the 1000-bp panel (Figure 1A). For the 2000-bp panel, 30.6% of sequencing data were assigned to one amplicon. Therefore, we selected the panel of ~1000-bp amplicons for the following analyses, referred to as 1K-panel. The 1K-panel contains 36 primer pairs in two pools (18 pairs each), and the sizes of the amplicons ranged from 880-bp to 1027-bp with an averagely 112-bp overlap.

We next compared the evenness of viral coverage 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), and aliquots were respectively amplified with the 1K-panel and the ARTIC panel, followed by sequencing with MiSeq. We defined the viral genome region with a >20% mean depth as the high-coverage region, and the proportion of the high-coverage regions was used to quantify coverage evenness.

We found that the 1K-panel generated a more even sequencing coverage than the ARTIC panel (Figure 1B). With the 1K-panel, the proportions of the high-coverage regions were averagely 93.0% (SD=1.7%) for the six samples; whereas it was 80.6% (SD=8.7%) for the ARTIC panel. The coverage evenness was dependent on the viral titer of samples, and we found that 1K-panel was less affected by low viral titers than the ARTIC panel (Figure 2A). As the C_t values increasing, the proportions of the high-coverage region slightly decreased from 95.3% to 90.5% for the 1K-panel, compared with 88.1% to 62.4% for the ARTIC

panel. We also evaluated a higher cutoff (>30% mean depth) to define the high-coverage region, and the 1K-panel maintained the advantage (Figure 2A).

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 2B). Among 29 of the 49 samples with a <30 C_t value, except one sample, all had a viral genome recovery with >90% high-coverage region (mean 93.8%, SD 8.1%). The evenness decreases for the samples with a >30 C_t value, but five of six samples with a 30-33 C_t value had >70% high-coverage regions.

The amplicon sizes of the 1K-panel exceeded the maximum read length of MPS, but are favorable for nanopore sequencing in which fragmentation of PCR products is not needed. Recent studies have shown that nanopore sequencing by using MinION sequencing could generate accurate consensus genome of SARS-CoV-2 (12, 16). Thus, we focused on assessing the level of intrinsic noise to identify viral single nucleotide variations (SNVs), especially the sub-clonal variations, via the 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 2×300 bp flowcell). A high depth coverage was obtained for both sequencing devices (Table 1) which is required to identify sub-clonal SNVs. Then we implemented a home-mode bioinformatics workflow based on piled sequencing bases (see Methods) to identify SNVs. We included the SNV with a mutated allele frequency (MuAF) larger than 0.2 for the assessments. With this cutoff of MuAF, the MiSeq-based identification of SNVs should be reliable as previous studies shown [17,22].

The SNVs respectively identified with MinION and MiSeq sequences are shown in Figure 3A. Based on the distributions of MuAFs of these SNVs, we found that, with a $\text{MuAF} \geq 0.8$ threshold to identify SNVs, 96.8% (30/31) SNVs based on MiSeq could be identified by using MinION. One SNV (G26526T, in sample S1) was missed due to a much reduced MuAF (MiSeq 0.95, MinION 0.51). No false positive SNV with a ≥ 0.8 MuAF was observed through MinION sequencing, which is consistent with previous studies [12,16]. However, MinION-based SNVs with a lower MuAF were highly error-prone. Benchmarked by the MiSeq-based SNVs with a 0.2-0.8 MuAF, the MinION-based SNVs had a recall rate of 100% (4 of 4) and a precision as low as 12.9% (4/31). We illustrated these artificial SNVs and the true positives along the viral genome (Figure 2B). Most artificial SNVs were shared among samples, which is consistent with the feature of systematic error of nanopore sequencing [23,24]. No obvious enrichment of artificial SNVs at certain viral genome region was observed. The result suggests that the sub-clonal SNVs could not be reliably identified based on MinION sequencing.

As the 1K-panel could generate an even sequencing coverage on the SARS-CoV-2 genome and no need to detect minor viral populations, a large amount of MinION sequencing data per sample (about 300 to 500 Mbp) seems over-abundant. Therefore, to test the limit of the 1K-panel to recover SARS-CoV-2 genome, we re-sequenced the eight samples with an ultra-low throughput by using MinION (Flongle R9.4.1). We obtained 1.21 to 6.14 Mbp data per sample, the majority of viral genomes were recovered with

sequencing reads (Figure 4). With a cutoff of ≥ 5 -fold sequencing depth, 29 (93.5%) of the 31 high-MuAF SNVs could be identified, and no false positive SNV was observed. The two SNVs (T11158C and T28144C) failed to be identified due to insufficient or no coverage in sample S7, which had the lowest 1.21 Mbp sequencing data. We furthermore examined the coverage evenness. In 7 of the 8 samples, >90% (90.5% - 96.6%) of the viral genome had high coverage, 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 the extremely low data input.

Discussion

As the COVID-19 pandemic goes on, the full-genome based surveillance of SARS-CoV-2 would become a routine approach for the control and prevention of the pandemic. Even in countries where the outbreak appears to be leveling off, such as China, continuously regional resurgence of COVID-19 has been seen [25]. Rapid recovery of viral genome and following phylogenetic analyses could determine the lineage of the virus and elucidate how the virus was spiked in and spread in local communities.

The multiplex PCR of SARS-CoV-2 followed by sequencing is the most widely-used strategy to recover the viral genome from primary samples. The portable sequencer MinION provides field sequencing and could achieve near real-time genomic surveillance of SARS-CoV-2 in an outbreak. The performance of the multiplex primer mix panel that targets SARS-CoV-2 is essential for the efficiency of recovering the viral genome. A panel with low bias in the amplification of amplicons could provide high coverage evenness on the viral genome. It reduces the possibility of failure to identify lineage-determining viral variations, which is important to trace the virus in the global phylogeny. Moreover, a high coverage evenness greatly decreases the cost per sample of sequencing. In our evaluation of the 1K-panel, approximate 5 Mbp sequencing data per sample could generate accurate consensus viral genome from samples with a <30 C_t value. Thus, we speculate that the sequencing of 96 or more barcoded samples in a batch with one MinION flow cell is feasible, in which about 20-30 Mbp data could be generated for each sample. It is especially useful when a large number of primary samples are needed to be analyzed, such as rapid lineage assignment of samples from screening testing or environmental contamination.

To improve the coverage evenness, we adopt ~ 1000 -bp amplicons for the panel, which are longer than those of some widely-used panels such as the ARTIC primer panel. Very recently, Moore et al. proposed a multiplex primer panel with amplicons ranged from 956-bp 1450-bp [11]. However, their primer pairs are distributed in six multiplex pools and the bias of certain amplicons during the amplification is obvious, which induces a more uneven coverage on the viral genome than the 1K-panel. Another major concern of long-amplicon panel is the applicability for clinical samples, as the RNA is easily degraded. Although we did not observe a reduced efficiency of the viral amplification among the clinical samples in this study, it should be noted that these samples were all tested within two weeks. Compared with short-amplicon panel, the long-amplicon panel would be more affected by the degradation of RNA/cDNA, especially for the samples stored for a long time. Quality control for RNA integrity before amplification could be helpful but it requires extra equipment and procedures. We suggest that the long-amplicon panel could be used

for the first round of viral amplification, and the samples that failed to obtain sufficient PCR products can be further amplified with another short-amplicon panel.

With the approach of 1K-panel amplification followed by MinION sequencing, consensus viral genome could be accurately recovered, but identifying viral sub-clonal variation is infeasible, possibly due to the systematic error of nanopore sequencing. Moreover, based on our assessment, a $\text{MuAF} > 0.8$ cutoff to define consensus SNVs through MinION sequencing is more reliable than the > 0.5 cutoff. In the study which analyzes the intra-host diversity of viral genomics, MPS is necessary. As the fragmentation of PCR products with the 1K-panel is needed to conduct MPS, one could choose the 1K-panel or other short-amplicon panels for viral amplification considering the trade-off between high coverage evenness and the extra experimental procedures.

Conclusions

We developed and validated a new two-pool, long-amplicon multiplex PCR primer panel for full-genome sequencing of SARS-CoV-2 genome from primary samples. The panel could generate a more even coverage of SARS-CoV-2 than the ARTIC short-amplicon panel, and subsequently has a reduced requirement for sequencing data amount. For the samples with a $C_t < 30$ titer, ~5 Mbp data by MinION provided a $> 95\%$ viral genome coverage with a > 10 -fold depth. Meanwhile, our assessment shows that nanopore sequencing with MinION enables reliable identification of dominating viral population (consensus viral genome) within samples, but was highly error-prone to discover the minor viral population. The SNVs with a < 0.6 MuAF should be regarded as unreliable.

Abbreviations

COVID-19: Coronavirus Disease 2019; SARS-CoV-2: severe acute respiratory syndrome coronavirus 2; PCR: polymerase chain reaction; qRT-PCR: real-time quantitative reverse transcription PCR; MPS: massive parallel sequencing; bp: base pair; C_t : cycle threshold; SNV: single nucleotide variation; MuAF: mutated allele frequency.

Declarations

Ethics approval and consent to participate

This study was conducted in accordance with the declaration of Helsinki approved and supervised by the Ethics Committee of Beijing Institute of Radiation Medicine & Chinese PLA Center for Disease Control and Prevention. Informed consent was obtained from the participants. No identifiable participants' data were included in this study.

Consent for publication

Not applicable

Availability of data and materials

The codes for bioinformatics analyses are available at <https://github.com/lhj920321/Long-amplicon-nanopore-sequencing-of-SARS-CoV-2>. The consensus genomes of SARS-CoV-2 for the eight samples are available at the GISAID and NCBI database through the accession numbers: S1, EPI_ISL_430722; S2, EPI_ISL_430724; S3, EPI_ISL_430735; S4, EPI_ISL_430742; S5, EPI_ISL_430741; S6, EPI_ISL_430740; S7, MT135041; S8, MT135041.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by grants from the Mega-projects of Science and Technology Research (No. 2018ZX10201001-003 and No. 2018ZX10305410-004). M.N and P.L were supported by the Beijing Nova Program (Z181100006218114 and Z181100006218110).

Authors' contributions

MN and PL designed this study. HS supervised the study. The primer design and multiplex viral amplification were performed by JL and YL. HL and MN analyzed the data, then created tables and figures. MN, PL and HL drafted the manuscript. XB revised this manuscript. All authors commented on previous versions of the manuscript and approved the final manuscript. All authors contributed to revision of the manuscript and approved the submitted version.

Acknowledgements

We would like to thank Ms. Yan Zhang for revising this manuscript.

Footnotes

Not applicable.

References

1. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. *Nature*. 2020;579(7798):265-9.
2. Shu YL, McCauley J. GISAID: Global initiative on sharing all influenza data - from vision to reality. *Eurosurveillance*. 2017;22(13):2-4.
3. Bedford T, Greninger AL, Roychoudhury P, Starita LM, Famulare M, Huang ML, et al. Cryptic transmission of SARS-CoV-2 in Washington state. *Science*. 2020;370(6516):571-5.

4. Eden JS, Rockett R, Carter I, Rahman H, de Ligt J, Hadfield J, et al. An emergent clade of SARS-CoV-2 linked to returned travellers from Iran. *Virus Evol.* 2020;6(1): veaa027.
5. Gambaro F, Behillil S, Baidaliuk A, Donati F, Albert M, Alexandru A, et al. Introductions and early spread of SARS-CoV-2 in France, 24 January to 23 March 2020. *Eurosurveillance.* 2020;25(26):8-14.
6. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics.* 2018;34(23):4121-3.
7. Oude Munnink BB, Nieuwenhuijse DF, Stein M, O'Toole A, Haverkate M, Mollers 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):1405-10.
8. Rambaut A, Holmes EC, O'Toole A, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol.* 2020;5(11):1403-7.
9. Tang XL, Wu CC, Li X, Song YH, Yao XM, Wu XK, et al. On the origin and continuing evolution of SARS-CoV-2. *Natl Sci Rev.* 2020;7(6):1012-23.
10. Bohmer MM, Buchholz U, Corman VM, Hoch M, Katz K, Marosevic DV, 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):920-8.
11. Moore SC, Penrice-Randal R, Alruwaili M, Randle N, Armstrong S, Hartley C, 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).
12. Paden CR, Tao Y, Queen K, Zhang J, Li Y, Uehara A, et al. Rapid, Sensitive, Full-Genome Sequencing of Severe Acute Respiratory Syndrome Coronavirus 2. *Emerg Infect Dis.* 2020;26(10):2401-5.
13. Ni M, Chen C, Qian J, Xiao HX, Shi WF, Luo Y, et al. Intra-host dynamics of Ebola virus during 2014. *Nat Microbiol.* 2016;1(11):16151.
14. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, 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):1261-76.
15. Quick J, Loman NJ, Duraffour S, Simpson JT, Ettore S, Cowley L, et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature.* 2016;530(7589):228-32.
16. Lu J, du Plessis L, Liu Z, Hill V, Kang M, Lin HF, et al. Genomic Epidemiology of SARS-CoV-2 in Guangdong Province, China. *Cell.* 2020;181(5):997-1003.
17. Ni M, Chen C, Liu D. An Assessment of Amplicon-Sequencing Based Method for Viral Intra-host Analysis. *Virol Sin.* 2018;33(6):557-60.
18. Grubaugh ND, Gangavarapu K, Quick J, Matteson NL, De Jesus JG, Main BJ, et al. An amplicon-based sequencing framework for accurately measuring intra-host virus diversity using PrimalSeq and iVar. *Genome Biol.* 2019; 20(1):8.

19. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-60.
20. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-9.
21. Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, et al. Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods*. 2018;15(6):461-8.
22. 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):e37.
23. Kono N, Arakawa K. Nanopore sequencing: Review of potential applications in functional genomics. *Dev Growth Differ*. 2019;61(5):316-26.
24. Jain M, Olsen HE, Paten B, Akeson M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol*. 2016;17:239.
25. Xinghuo P, Shuangsheng W, Wentai M, Jian Y, Lin D, Jie L, 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](https://doi.org/10.1093/nsr/nwaa264).

Tables

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)	MinION coverage		
			Mean	>100X	>1000X		Mean	>100X	>1000X
S1	22.88	483.7	13,794	100	97.62	335.0	9,733	100	97.55
S2	23.41	369.1	10,433	100	96.55	320.2	9,088	100	99.39
S3	25.18	508.4	14,361	100	93.23	339.0	9,852	99.92	95.66
S4	26.1	453.4	12,830	100	96.11	266.2	7,604	99.93	96.24
S5	28.84	363.4	10,356	100	91.94	315.2	9,054	99.94	93.04
S6	30.97	429.6	9,718	97.77	91.75	336.3	7,590	98.8	91.83
S7	31.58	423.3	11,983	97.74	89.37	263.8	7,644	99.22	90.29
S8	33.15	435.5	12,234	95.57	81.65	307.9	8,992	99.76	85.63

Figures

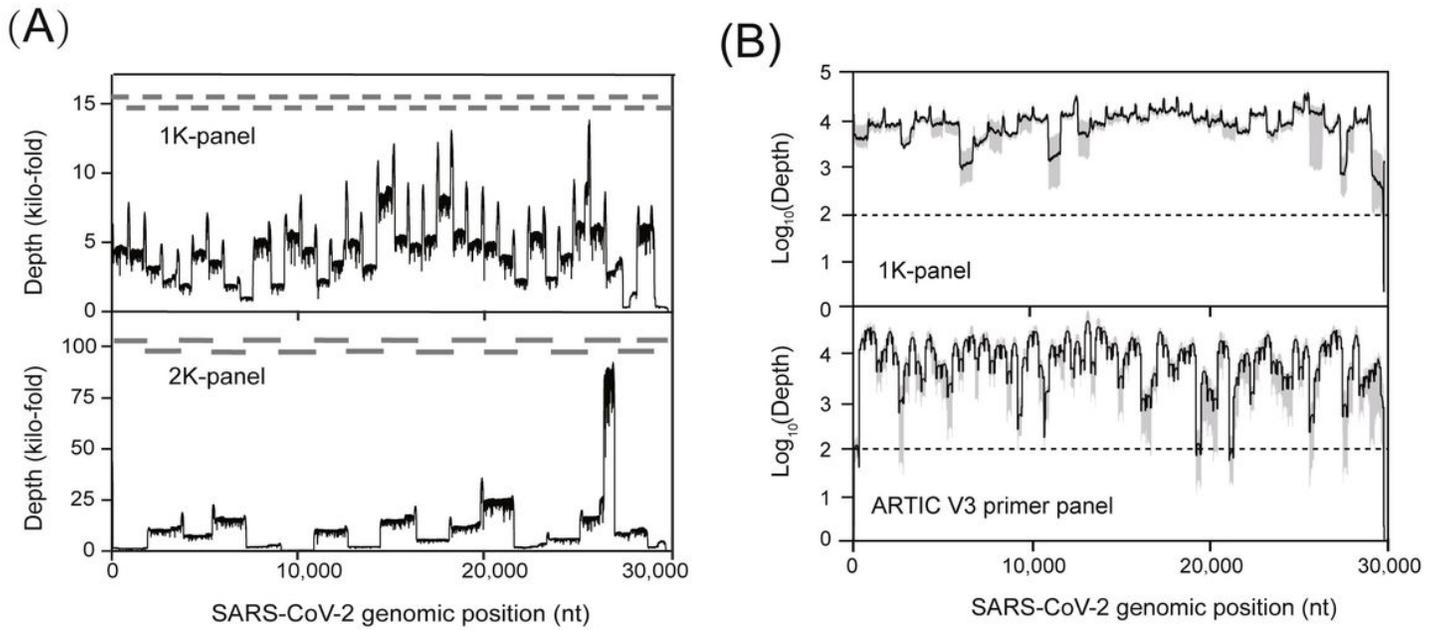


Figure 1

Coverage evenness of SARS-CoV-2 genome of multiplex PCR mix panels. (A) The coverage depth profiles of one clinical sample for the panels with ~1K-bp amplicons (the upper, 1K-panel) and ~2K-bp amplicons (the lower, 2K-panel). The sample had a viral titer of 25 Ct value of 25 for qRT-PCR. Genomic positions of amplicons are indicated by grey solid lines above the depth profile. The y-axis is in absolute value. (B) The coverage depth profile of six clinical samples for the 1K-panel and the ARTIC V3 primer panel. Depth is in standard logarithmic scale, with 100-fold denoted by the dashed line. Bold black line denotes median depth, and grey shadow indicates the minimum and maximum of depth. Mean depth is defined as the average site depth on the targeted viral genome (excluding the 3' and 5' end).

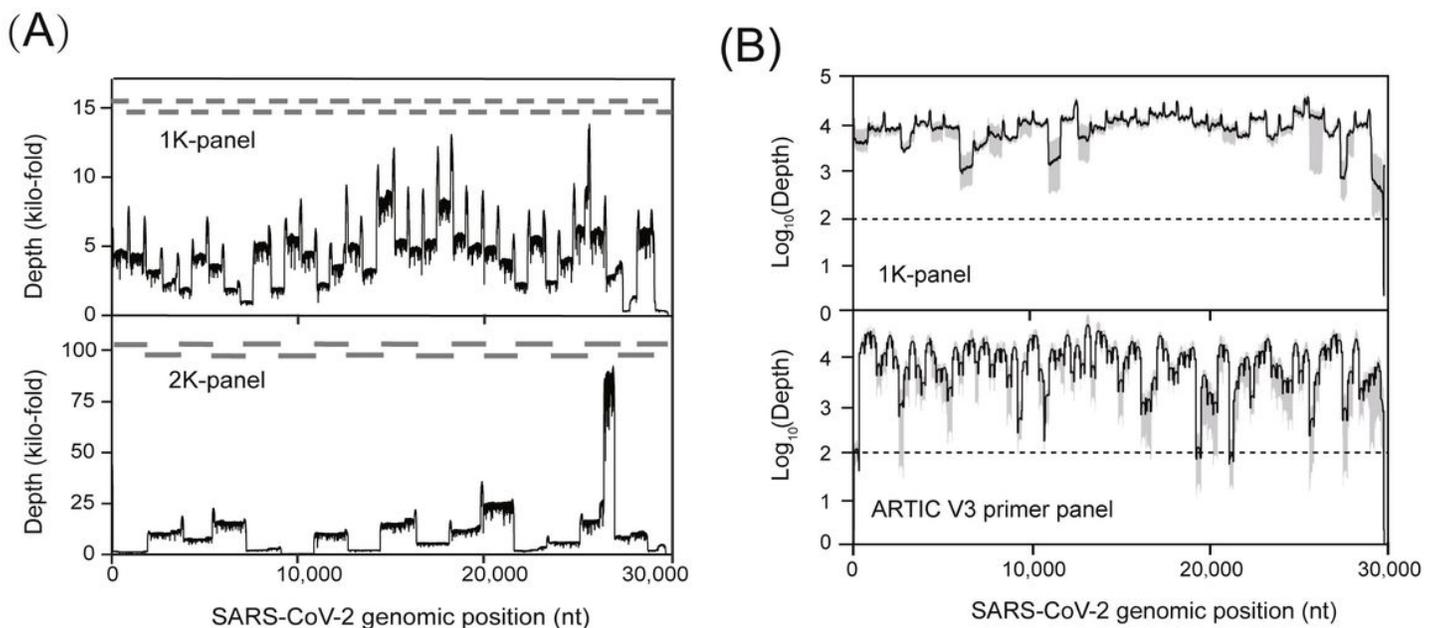


Figure 1

Coverage evenness of SARS-CoV-2 genome of multiplex PCR mix panels. (A) The coverage depth profiles of one clinical sample for the panels with ~1K-bp amplicons (the upper, 1K-panel) and ~2K-bp amplicons (the lower, 2K-panel). The sample had a viral titer of 25 Ct value of 25 for qRT-PCR. Genomic positions of amplicons are indicated by grey solid lines above the depth profile. The y-axis is in absolute value. (B) The coverage depth profile of six clinical samples for the 1K-panel and the ARTIC V3 primer panel. Depth is in standard logarithmic scale, with 100-fold denoted by the dashed line. Bold black line denotes median depth, and grey shadow indicates the minimum and maximum of depth. Mean depth is defined as the average site depth on the targeted viral genome (excluding the 3' and 5' end).

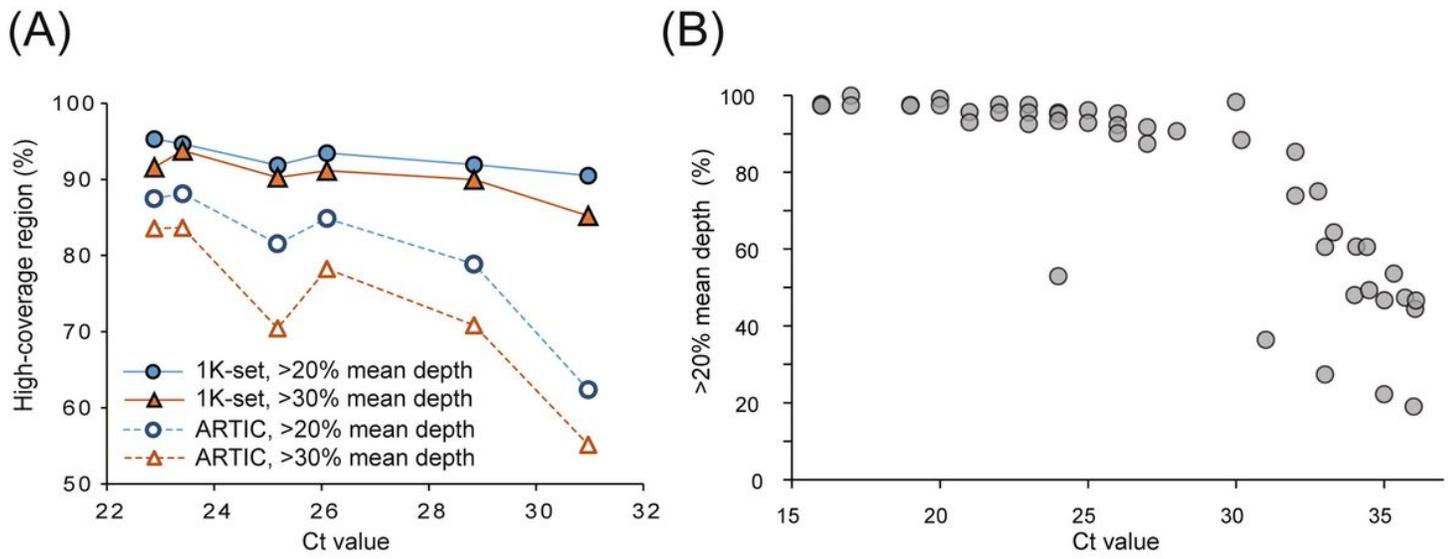


Figure 2

The correlation between coverage evenness and viral titers in samples. (A) Two thresholds, >20% and >30% mean depth, were used to define the region of high-coverage. Viral titers were quantified by the Ct value of qRT-PCR for SARS-CoV-2 detection. (B) The proportions of the genome region with a >20% mean depth for 49 clinical samples. Amplification was performed with the 1K-panel and PCR products were sequenced with Illumina NovaSeq 6000.

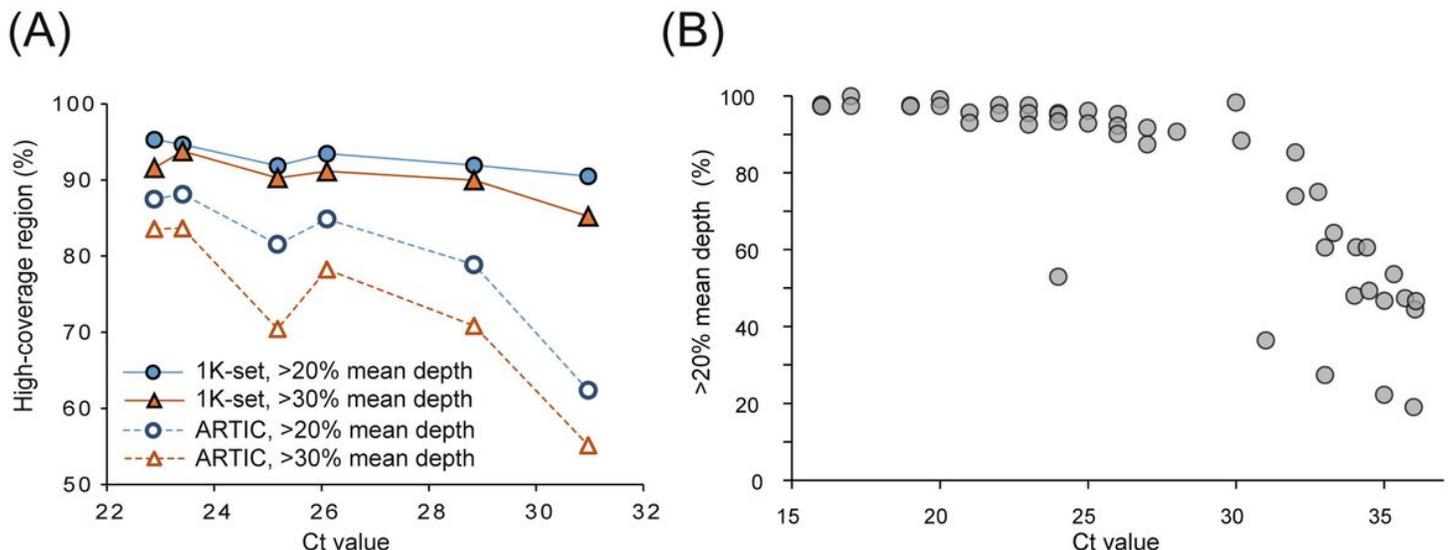


Figure 2

The correlation between coverage evenness and viral titers in samples. (A) Two thresholds, >20% and >30% mean depth, were used to define the region of high-coverage. Viral titers were quantified by the Ct value of qRT-PCR for SARS-CoV-2 detection. (B) The proportions of the genome region with a >20% mean depth for 49 clinical samples. Amplification was performed with the 1K-panel and PCR products were sequenced with Illumina NovaSeq 6000.

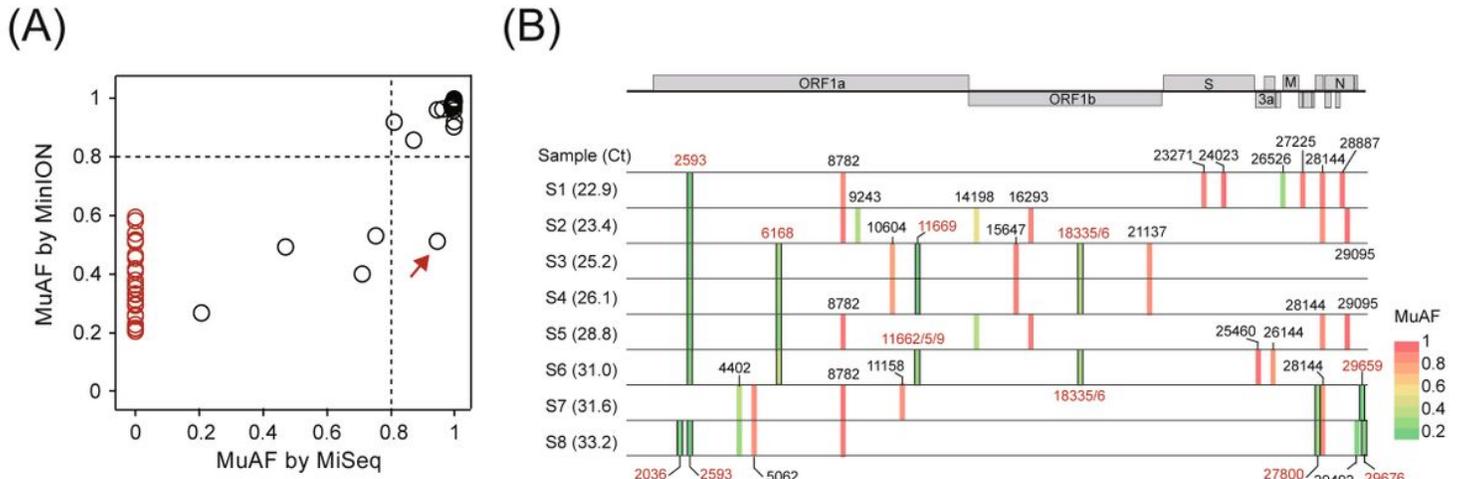


Figure 3

The identification of SNVs of SARS-CoV-2. (A) The SNVs were identified through the MiSeq-based and MinION-based sequencing, with the MuAFs shown. The SNVs that were not consistent between MinION and MiSeq sequencing are denoted in red. (B) The distribution of SNVs identified based on MinION sequencing. The MuAFs of SNVs are denoted by heat map. The false positive SNVs are indicated with positions in bars with a black border. Reference genome, the Wuhan-hu-1 strain (GenBank accession MN908947.3).

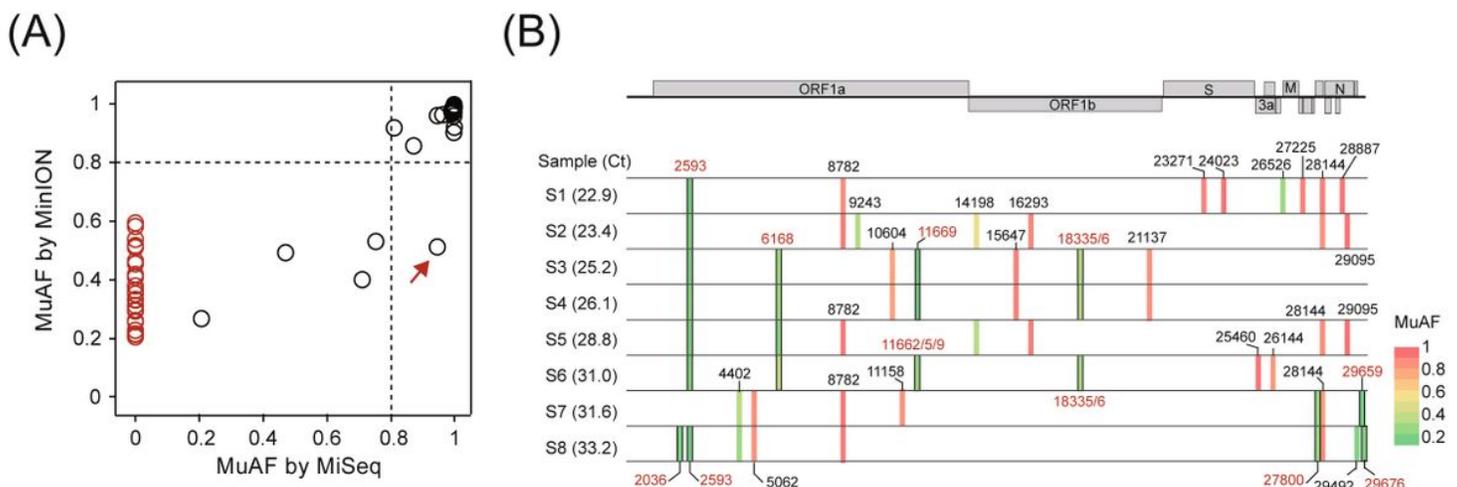


Figure 3

The identification of SNVs of SARS-CoV-2. (A) The SNVs were identified through the MiSeq-based and MinION-based sequencing, with the MuAFs shown. The SNVs that were not consistent between MinION and MiSeq sequencing are denoted in red. (B) The distribution of SNVs identified based on MinION sequencing. The MuAFs of SNVs are denoted by heat map. The false positive SNVs are indicated with positions in bars with a black border. Reference genome, the Wuhan-hu-1 strain (GenBank accession MN908947.3).

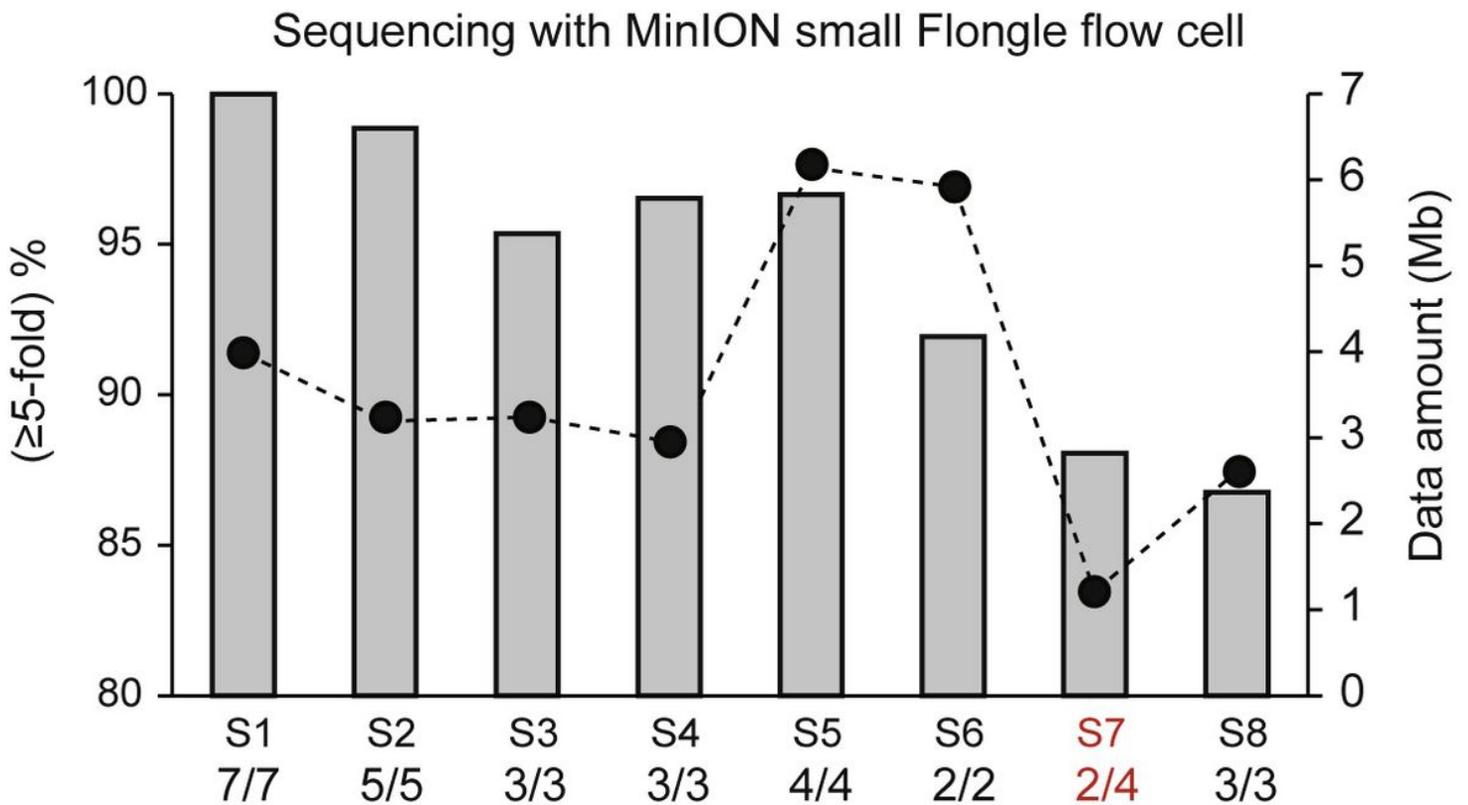


Figure 4

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

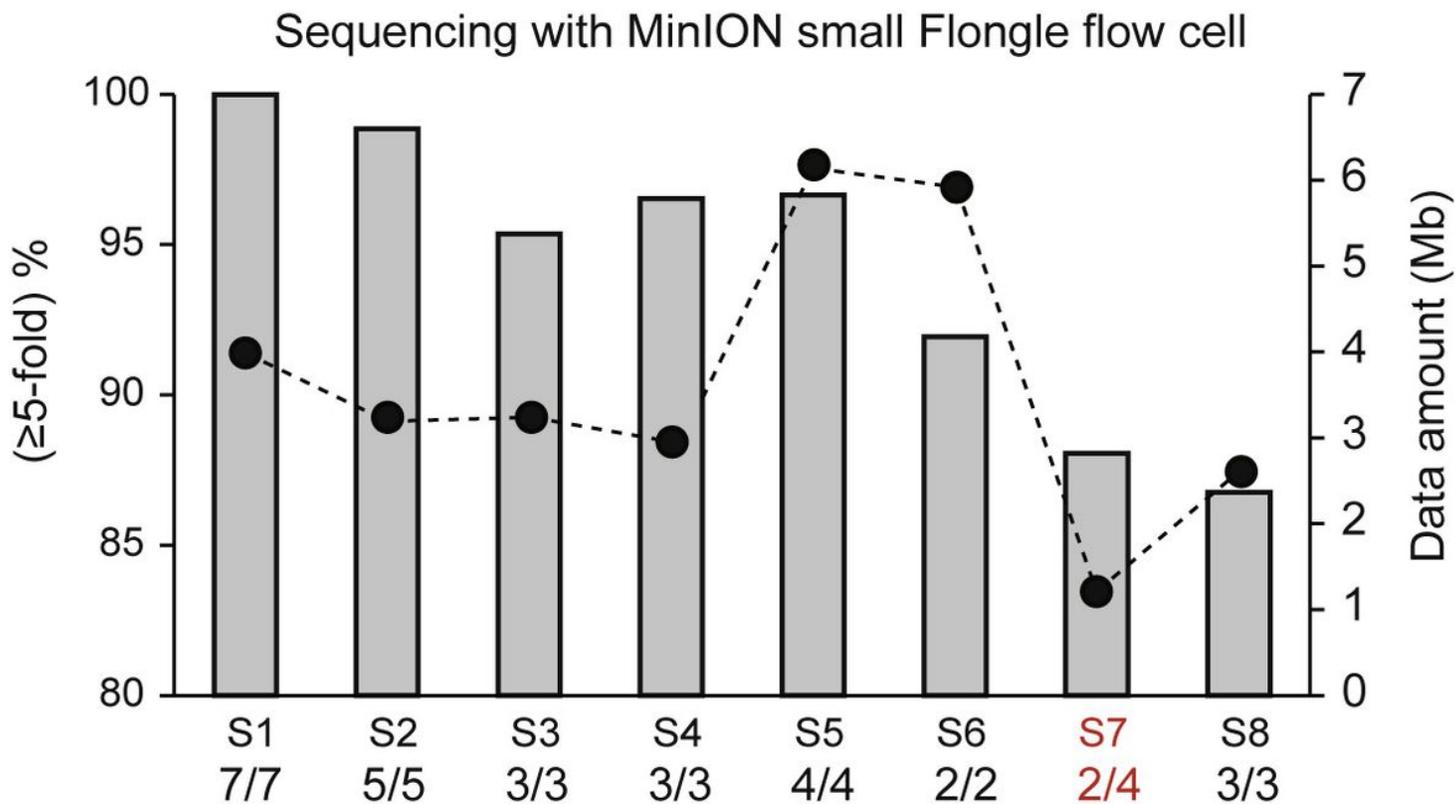


Figure 4

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

Supplementary Files

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

- [AdditionalFile.xlsx](#)
- [AdditionalFile.xlsx](#)