

# Rapid Genomic Characterization of SARS-CoV-2 Viruses From Clinical Specimens Using Nanopore Sequencing

Jun Li (✉ [junli.hz.cn@gmail.com](mailto:junli.hz.cn@gmail.com))

Microbiology Laboratory, Hangzhou Center for Disease Control and Prevention

Haoqiu Wang

Microbiology Laboratory, Hangzhou Center for Disease Control and Prevention

Lingfeng Mao

Hangzhou Baocheng Biotechnology Co., Ltd.

Hua Yu

Microbiology Laboratory, Hangzhou Center for Disease Control and Prevention

Xinfen Yu

Microbiology Laboratory, Hangzhou Center for Disease Control and Prevention

Zhou Sun

Department of Infectious Diseases, Hangzhou Center for Disease Control and Prevention

Xin Qian

Microbiology Laboratory, Hangzhou Center for Disease Control and Prevention

Shi Cheng

Microbiology Laboratory, Hangzhou Center for Disease Control and Prevention

Shuchang Chen

Microbiology Laboratory, Hangzhou Center for Disease Control and Prevention

Junfang Chen

Department of Infectious Diseases, Hangzhou Center for Disease Control and Prevention

Jingcao Pan

Microbiology Laboratory, Hangzhou Center for Disease Control and Prevention

Jueliang Shi

Hangzhou Baocheng Biotechnology Co., Ltd.

Xuchu Wang

Microbiology Laboratory, Hangzhou Center for Disease Control and Prevention

---

## Research Article

**Keywords:** SARS-CoV-2, Novel coronavirus, Genome characteristics, Nanopore sequencing, Variations

**Posted Date:** May 6th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-27215/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

**Version of Record:** A version of this preprint was published on October 15th, 2020. See the published version at <https://doi.org/10.1038/s41598-020-74656-y>.

## Abstract

The outbreak of the novel SARS-CoV-2 has swiftly spread worldwide. Rapid genome sequencing of the SARS-CoV-2 strains has become a helpful tool for better understanding of virus genomic characteristics and the origin. To obtain the virus whole-genome sequence directly from the clinical specimens, we performed the nanopore sequencing using a modified ARTIC protocol on portable nanopore sequencer, and validated the routine 8 hours workflow and 5 hours rapid pipeline. We had made some optimizations to improve the genome sequencing workflow. The sensitivity of the workflow was also tested by serially diluting RNA from clinical samples. The optimized pipeline was finally applied to obtain the whole genomes from 17 clinical specimens in Hangzhou from January 2020 to March 2020. In the obtained 17 complete genomes of SARS-CoV-2, 12 variations were found and analyzed. The genomic variations and phylogenetic analysis hinted the multiple sources and different transmission pattern during the COVID-19 epidemic in Hangzhou, China. In conclusion, the genomic characteristics and the origin of the virus could be quickly determined by nanopore sequencing following our workflows.

## Introduction

In December 2019, an outbreak of atypical pneumonia with unclear etiology began in Wuhan, a major transportation hub in the center of China <sup>1</sup>. A novel coronavirus similar to severe acute respiratory syndrome coronavirus (SARS-CoV) was then identified as the causative pathogen <sup>2</sup>, which was officially named as SARS-CoV-2 by the International Committee on Taxonomy of Viruses (ICTV). Before that, most of human infected coronaviruses (HCoV) can only cause mild upper respiratory infections (HCoV-229E, HCoV-NL63, HCoV-OC43 and HCoV-HKU1) <sup>3</sup>, but sometimes they cross species and cause fatal respiratory disease and outbreaks, as experienced in cases of SARS-CoV <sup>4</sup> or Middle East respiratory syndrome coronavirus (MERS-CoV) <sup>5</sup>. The seventh HCoV SARS-CoV-2 made a spillover event in late December 2019 and is swiftly spreading across the borders of city and province in the mainland of China, and soon became an emergency of major international concern. As of 24 April 2020, the cumulative number of confirmed human infections rapidly increased to 2,549,632 (175,825 deaths) reported by WHO (<https://who.sprinklr.com/>), almost 300 times more than the total number of SARS-CoV infected cases.

Hangzhou is a national tourism city with a registered population of 10.36 million, and located at the south wing of the Yangtze River Delta with a humid, subtropical climate facilitating the airborne survival and transmission of viruses associated with respiratory infections. The first case was found from a Wuhan returnee in January 19, 2020. As of March 2020, 186 infections were confirmed by viral RNA detection. As the virus genome could be sequenced rapidly on the portable MinION sequencer, the accurate genomic sequencing data generated could be used to trace back to the origin during the viruses spreading, which could bring the molecular epidemiology analysis close to the aim of front-line application.

Therefore, we conduct a modified ARTIC protocol for SARS-CoV-2 genome sequencing on the MinION platform. Two workflows were applied and validated by amplifying and sequencing the genome from SARS-CoV-2 infected clinical samples, and the 17 genomes characteristics of SARS-CoV-2 viruses in Hangzhou were analyzed to study the origin and transmission history of these viruses.

## Results

### 8h and 5h Workflow for SARS-CoV-2 Nanopore Sequencing

In order to acquire the whole genome sequence of SARS-CoV-2 more efficiently, 8 hours workflow was designed for its sequencing throughput and speed after loading library into the flow cell, and 5 hours workflow was designed for rapid library building need (Figure 1). These two work flows were tested on the sample of HZCDC0001 with Ct value 26.51/27.03 (Orf1ab/N), the first sample appeared in Hangzhou, Zhejiang province. In the 8h workflow, nanopore ligation sequencing kit was applied because this protocol can maximize the sequencing throughput and the length of reads, in figure 2c, the total bases and coverage of SARS-CoV-2 increased much faster than the 5h workflow, in only 10 minutes after loading library to the flow cell, the coverage reached to almost 100%, here we defined a base had a coverage with at least 15× depth. Meanwhile, 5h workflow took more than 1.5 hours to approach the 100% coverage. The 5h workflow advantaged at the rapid 15 minutes library preparation time, especially for the extreme condition, this workflow can hugely shorten the library preparation time compared to the 2 hours of ligation protocol (figure 1). However, as the nanopore rapid protocol would cleavage DNA to quickly add transposase adapters, sequencing throughput and speed behaved much poorly comparing with the 8h workflow (figure 2a-c). In both workflows, two regions (5231 - 5644 bp and 22798 - 23214 bp, primer pairs #18 and #76) appeared to be the short boards in genome mapping, which needs the further optimization.

Meanwhile, we had made some optimizations to improve the workflow. Since the step of RNA extraction is vital for the follow-up sequencing, we compared the magnetic beads extraction on NP968 (Tianlong, China) with column RNA extraction using RNeasy Mini Kit (QIAGEN, Germany) (Figure 2d). The latter approach seems to yield higher quality RNA, making the depth of coverage nearly doubled (Figure 2f). What's more, PCR procedure took more than 40% time of total workflow, so we tried to decrease the annealing and extension time from 5 minutes to 3 minutes and 1 minute, corresponding to the total time of PCR from about 3 hours to 1 hour. Even if the annealing time reduced to 1 minute, the whole genome sequence still could be obtained from the products of 1h PCR procedure (Figure 2e, g). On account of the different viral titers of SARS-CoV-2 in clinical samples, 3-min annealing time could be considered as the equilibrium point.

# Sensitivity Test and the Application in Clinical Samples

As clinical specimens may have extremely low viral titers, we tested the sensitivity of the routine 8 hours workflow by 10 times serially diluting RNA (Starting from the RNA samples with Ct value ~ 22) from two COVID-19 patients to determine whether we could amplify the whole genome of SARS-CoV-2 from samples that had a failure in qRT-PCR test. Approximately 85% reads were mapped to the reference genome (Genbank accession number MN908947.3) with an average depth of coverage greater than 250 × across > 97.56% of SARS-CoV-2 genome for both two samples, including the 100,000 × diluted sample that was undetectable by qRT-PCR on ABI 7500 (Figure 2 h-j). Average depth was not obviously decreased among the diluted samples, however the genome-wide depth fluctuated much significantly while the dilution rate was above 10,000×.

The routine workflow was also applied to obtain the genomes from 17 clinical samples with the average depth from 233.75 - 754.88 and the genome coverage from 98.08% - 100% (Table 1). The information of specimens including genome-wide depth by nanopore sequencing and Ct value by RT-PCR are listed in Table 1 and Figure 2k. Especially, some specimens with low Ct value, such as HZCDC0090 and HZCDC0091, was sequenced with the genome coverage from

99.15 to 100%, which provided evidence that some problematic clinical samples can be solved by nanopore sequencing. Besides, both the upper and lower respiratory tract specimens (HZCDC0048, HZCDC0048L, HZCDC0090, HZCDC0090L, HZCDC0091 and HZCDC0091L) from three COVID-19 patients were sequenced, and the genome mapping results shows that the virus genomes from different parts of respiratory tract are consistent (Table 1).

## Genomic Variations and Phylogenetic Analyses of SARS-CoV-2

The length of reference SARS-CoV-2 genome (MN908947.3) is 29,903 bp. However, a considerable part of the submitted SARS-CoV-2 genomes are uncompleted. So, the strategy of building phylogenetic tree based on SNPs was applied to investigate the traceability of interested samples. To avoid introducing the errors of nanopore sequencing, we first filtered the low quality reads and only homozygous SNP with high quality (phred value  $\geq 20$ ) and high site depth ( $\geq 50$ ) were considered for downstream analysis. Besides, we also perform the illumina sequencing in all samples including HZCDC6706 and HZCDC6789 imported from abroad, which provided the evidence that SNPs from our standard was 100% consistent to illumina data and SNPs can be called on only nanopore sequencing in SARS-CoV-2 genome (Table 2).

In all obtained 17 complete genomes of SARS-CoV-2, 12 substitutions distributed in three coding sequences (CDS) and 5'UTR were identified based on sequences alignment (Figure 3), including C241T in 5'UTR, 5 synonymous variations and 6 missense variations (Table 2). Nine variations were found in orf1ab gene, the longest gene in SARS-CoV-2 genome, and only one

variation was detected in S and ORF3a gene, respectively. The couple infected in Wuhan, HZCDC0012 and HZCDC0013 owned the same two variations in 11083 and 21282 site. Nine variations were detected in the samples HZCDC6789, which was nearest to now in time and imported from abroad, and few variations were detected in the early samples.

Based on marker variants from 507 full genomes shared (GISAID), three major subclades were named as S (Orf8: L84S), G (S: D614G) and V (Orf3: G251V). The phylogenetic relationship reveals that most of our genomes from early COVID-19 patients linked to Wuhan are in the main clade, except HZCDC0025 and HZCDC0167 belonging to subclade V (Figure 3). Two genomes from the international imported cases are in subclade G and V respectively (Figure 3), which hinted multiple sources of transmission from oversea. However, the clade classification will need to be adjusted when the marker is not stable occurring in the other newly sequenced genomes.

## Discussion

The outbreak of COVID-19 caused by SARS-CoV-2 has swiftly spread worldwide. However, probable origin of SARS-CoV-2 associated with the COVID-19 pandemic is still unclear<sup>6</sup>. Recently reports about COVID-19 cases with none or mild upper respiratory tract symptoms suggest the potential for asymptomatic or oligo-symptomatic transmission during the first week of symptoms<sup>7,8</sup>. Hence, there is an urgent need for rapid identification and traceability of pathogens for the diseases control and prevention. The deep understanding of the novel virus first comes from the analysis of the genome sequence. In this study, we demonstrate the utility of nanopore sequencing for SARS-CoV-2 genomes from clinical specimens based on modified ARTIC protocol. It allowed confirmation of SARS-CoV-2 infection at genomic-level within a few minutes by sequencing and simultaneously mapping the reads to reference genome and analyzing the output data in real-time.

Comparing with nasal/oropharyngeal swabs, the lower respiratory tract specimens from COVID-19 could be detected more readily<sup>9</sup>. Our data shows that the virus genomes from different parts of respiratory tract are consistent. However, the difference of viral loads in samples will affect the stability of the average depth and genome-wide coverage, and increase the difficulty of whole genome mapping, suggesting the importance of sample collection for the later genome sequencing.

To characterize the genomic variations, we found 12 different substitution sites distributed in four coding regions among 17 genomes of SARS-CoV-2, without any recombination event. Genomic evidence supported that most of the first early 15 infections were directly or indirectly linked to Wuhan, and

the genomes from two imported infections occur in March 2020 have specific variation comparing with domestic strains. The phylogenetic analysis indicated the potential for human to human transmissions as previously reported elsewhere<sup>10</sup>. For example, the identical substitutions (G11083T: Orf1ab Leu3606Phe and A21282G: Orf1ab Asp7006Gly) in the virus genomes from a couple hinted the family-cluster transmission history of SARS-CoV-2 from a Wuhan traveler to her husband.

In summary, we performed the SARS-CoV-2 genome sequencing on portable nanopore sequencer. Combined with 8 hours workflow, the genomic characteristics and the origin of the virus could be quickly determined. The rapid 5 hours workflow with 15 minutes fast library preparation could be applied on trace-back task out of lab, bringing the genome-level molecular epidemiology analysis to the front-line of the outbreaks. Therefore, based on prompt diagnosis and rapid whole-genome analysis, the swift and decisive response to the SARS-CoV-2 outbreak will be benefit to the diseases control and prevention.

## Methods

### Ethics statement

This study and all experimental protocols were approved by the Institutional Review Board of Hangzhou Center for Diseases Control and Prevention. We confirm that all methods were carried out in accordance with relevant guidelines and regulations. Signed informed consents were obtained from the patients or their spouses and personally identifiable information was anonymized.

### Viral Infections Diagnose

Upper and/or lower respiratory tract samples, including nasal, oropharyngeal swabs, sputum, tracheal aspirate samples and bronchoalveolar-lavage fluid, were collected from suspected cases with informed consent from patients or their spouses, and were sent to Hangzhou Center for Diseases control and prevention on ice for diagnosis within six hours. The viral RNA was extracted directly from 200  $\mu$ L supernatant of clinical sample using the RNeasy Mini Kit (QIAGEN, Germany) according to the manufacturer's instruction in a biosafety cabinet in Biosafety Level 2 Laboratory, and tested for the presence of SARS-CoV-2 using the diagnostic real-time reverse transcription polymerase chain reaction (qRT-PCR) on ABI7500 (ABI, USA) following the diagnostic kit manual.

### Workflows of Virus Genome Sequencing

Viral RNA extracted from clinical samples was used as a template to amplify and sequence the SARS-CoV-2 genome. Briefly, cDNA was synthesized from 11  $\mu$ L viral RNA using SuperScript™ IV First-Strand Synthesis System (Invitrogen) with random hexamers. PCR was performed using Q5 Hot Start High-Fidelity DNA Polymerase (NEB) and a set of primers targeting regions of the SARS-CoV-2 genome designed by ARTIC network (<https://artic.network/ncov-2019>). The PCR mixture was initially incubated 2 minutes at 98 °C

for the denaturation, followed by 35 cycles at 98 °C for 15 seconds, 65 °C for 1, 3 or 5 minutes (depending on the workflows). The amplified products were purified by equal volume of AMPure XP beads (Beckman Coulter) to exclude small nonspecific fragments.

According to eight-hour routine workflow (Figure 1), the purified DNA was repaired with NEBNext FFPE Repair Mix (NEB), followed by the DNA ends preparation using NEBNext End repair / dA-tailing Module (NEB) and the successively attachment of native barcodes and sequencing adapters supplied in the EXP-NBD104/114 kit (Nanopore) to the DNA ends. The DNA concentration was determined by a Qubit 3.0 using dsDNA HS Assay Kit (Thermo Fisher). After priming the flow cell, 60 ng DNA per sample of products were pooled in the DNA library with final volume of 65  $\mu$ L. Following the ligation sequencing kit (SQK-LSK109) protocol, MinION Mk1B was used to perform the genome sequencing on an R9.4.1 Flow Cell for 1 hour per sample. For rapid barcoding workflow, a fragmentation mix in the SQK-RBK004 kit (Nanopore) was used to attach the barcodes to the DNA ends, followed by the attachment of sequencing adapters.

### Reads Preprocessing and Consensus Building for Nanopore Sequencing

Base calling was performed by guppy (<https://community.nanoporetech.com>) with the parameter “-c dna\_r9.4.1\_450bps\_hac.cfg -x auto” and different samples were separated and adapters were trimmed with the additional parameter “-trim\_barcodes -barcodes EXP-NBD104/EXP-NBD114/SQK-RBK004”. FASTQ reads were filtered for quality control with the cutoff “length >= 200 and Phred value >= 7” using program “filtlong v0.2.0” (<https://github.com/rwick/Filtlong>).

After reads quality control, artic-ncov2019 pipeline(<https://artic.network/ncov-2019>) was applied to perform the sequences mapping, primers trimming, variations calling and consensus assembly building. Variations were called using Medaka 0.11.1 (<https://github.com/nanoporetech/medaka>). In the stage of consensus assembly building, the site with depth lower than 50  $\times$  will be masked by N, and reference will be substituted by the homozygous variations with phred quality >= 20. “Samtools depth” was used to calculate the depth of each site and “Samtools bedcov” was used to calculate the window depth scanning in the genome<sup>11</sup>.

# Reads Preprocessing and Variations Calling for Illumina Sequencing

Raw illumina PE reads were trimmed and quality controlled by the software fastp 0.20.0 with default parameters<sup>12</sup>. Bwa 0.7.17-r1188<sup>13</sup> was used to mapping the clean reads to SARS-CoV-2 reference genome and sam/bam files were manipulated by samtools 1.9<sup>11</sup>, variations were detected with the program “mpileup and calling” from bcftools 1.9<sup>14</sup>, variations will be considered as positive when the homozygous variations has a phred quality value  $\geq 20$  and a depth  $\geq 50$ .

## Phylogeny and Variations Analysis

In order to remove the bias from the gaps in uncompleted genome, sequence alignment from all SNPs site were chosen to build the phylogenetic tree. First, all SNPs were called aligned to SARS-CoV-2 reference sequences using “nucmer” and “dnadiff” programs from MUMmer 3.23<sup>15</sup>, the effect of SNPs were estimated using SnpEff 4.3t<sup>16</sup>. Second, all SNPs site were connected to a single sequence for every sample based on the variations calling results from the last step. Then, these sequences were combined to perform phylogeny analysis, maximum likelihood phylogenies were estimated by FastTree 2.1.10<sup>17</sup> with default parameters. The phylogenetic tree with variations heatmap matrix was drew by phyD3<sup>18</sup>. The group and clade numbers were assigned to achieve consistency with the earlier studies.

## Declarations

### Data Availability

All genome sequences in this study are available in GISAID (<http://gisaid.org>) (the full list of the accession numbers are available from table 1).

### Acknowledgments:

This work was supported by the Hangzhou Bureau of Science and Technology under Grant [number 20202013A02].

### Competing Interests:

The authors declare no competing interests.

### Authors' Contributions:

W.X.C., L.J., W.H.Q., M.L.F., P.J.C. and S.J.L. designed the study and wrote the manuscript. C.J.F., S.Z. involved in the sample collection. C.S.C., L.J., Y.X.F. and Q.X. did the viral detection. L.J., Y.H. and C.S. participated in the genomes sequencing on nanopore MinION sequencer and illumina Miseq platform. M.L.F., L.J. performed the data analysis. All authors reviewed the manuscript.

## References

1. Li, Q. *et al.* Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med* **382**, 1199-1207, doi:10.1056/NEJMoa2001316 (2020).
2. Zhu, N. *et al.* A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med*, doi:10.1056/NEJMoa2001017 (2020).
3. Su, S. *et al.* Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. *Trends Microbiol* **24**, 490-502, doi:10.1016/j.tim.2016.03.003 (2016).
4. Drosten, C. *et al.* Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N Engl J Med* **348**, 1967-1976, doi:10.1056/NEJMoa030747 (2003).
5. Zaki, A. M., van Boheemen, S., Bestebroer, M., Osterhaus, A. D. & Fouchier, R. A. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med* **367**, 1814-1820, doi:10.1056/NEJMoa1211721 (2012).
6. Zhang, , Wu, Q. & Zhang, Z. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. *Curr Biol*, doi:10.1016/j.cub.2020.03.022 (2020).
7. Wolfel, *et al.* Virological assessment of hospitalized patients with COVID-2019. *Nature*, doi:10.1038/s41586-020-2196-x (2020).
8. Rothe, *et al.* Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. *N Engl J Med*, doi:10.1056/NEJMc2001468 (2020).
9. Han, H., Luo, Q., Mo, , Long, L. & Zheng, W. SARS-CoV-2 RNA more readily detected in induced sputum than in throat swabs of convalescent COVID-19 patients. *Lancet Infect Dis*, doi:10.1016/S1473-3099(20)30174-2 (2020).
10. Riou, J. & Althaus, C. L. Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. *Euro Surveill* **25**, doi:10.2807/1560-7917.ES.2020.25.4.2000058 (2020).
11. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

12. Chen, S., Zhou, , Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884-i890, doi:10.1093/bioinformatics/bty560 (2018).
13. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
14. Narasimhan, *et al.* BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics* **32**, 1749-1751, doi:10.1093/bioinformatics/btw044 (2016).
15. Kurtz, S. *et al.* Versatile and open software for comparing large genomes. *Genome Biol* **5**, R12, doi:10.1186/gb-2004-5-2-r12 (2004).
16. Cingolani, *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80-92, doi:10.4161/fly.19695 (2012).
17. Price, M. N., Dehal, S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490, doi:10.1371/journal.pone.0009490 (2010).
18. Kreft, L., Botzki, A., Coppens, , Vandepoele, K. & Van Bel, M. PhyD3: a phylogenetic tree viewer with extended phyloXML support for functional genomics data visualization. *Bioinformatics* **33**, 2946-2947, doi:10.1093/bioinformatics/btx324 (2017).

## Tables

Table 1 The information of seventeen Hangzhou SARS-CoV-2 viruses from COVID-19 patients in this study.

Strain number	Source	Gender	Age	Date of onset	Collection date	History	qRT-PCR (Ct)		GISAID accession number	Average depth	Coverage(%)
							ORF	N			
HZCDC0001	Sputum	Male	31	2020-01-18	2020-01-19	Infected in Wuhan	26.51	27.03	EPI_ISL_407313	754.88	100.00
HZCDC0012	Nasal, oropharyngeal swab	Male	47	2020-01-18	2020-01-20	Spouse of HZCDC0013	22.2	21.67	EPI_ISL_421236	326.64	100.00
HZCDC0013	Nasal, oropharyngeal swab	Female	45	2020-01-15	2020-01-20	Infected in Wuhan	22.34	21.88	EPI_ISL_421235	418.66	100.00
HZCDC0025	Nasal, oropharyngeal swab	Male	51	2020-01-21	2020-01-21	Infected in Wuhan	35.42	34.83	EPI_ISL_421234	509.46	99.99
HZCDC0048	Nasal, oropharyngeal swab	Male	35	2020-01-16	2020-01-21	Contact to patient from Wuhan	27.95	27.72	EPI_ISL_421233	506.14	100.00
HZCDC0048L	Tracheal aspirate sample					Wuhan	19.7	19.58	EPI_ISL_421232	385.17	100.00
HZCDC0049L	Tracheal aspirate sample	Female	40	2020-01-17	2020-01-21	Contact to patient from Wuhan	25.11	24.64	EPI_ISL_421231	413.09	100.00
HZCDC0090	Nasal, oropharyngeal swab	Female	34	2020-01-17	2020-01-21	Contact to patient from Wuhan	Neg.	Neg.	EPI_ISL_421230	233.75	99.15
HZCDC0090L	Bronchoalveolar-lavage fluid					Wuhan	30.29	29.54	EPI_ISL_421229	555.60	100.00
HZCDC0091	Nasal, oropharyngeal swab	Male	31	2020-01-17	2020-01-21	Contact to patient from Wuhan	Neg.	37.14	EPI_ISL_421228	264.45	99.16
HZCDC0091L	Bronchoalveolar-lavage fluid					Wuhan	33.09	32.95	EPI_ISL_421227	535.46	99.97
HZCDC0119	Nasal, oropharyngeal swab	Female	41	2020-01-21	2020-01-22	Contact to patient from Wuhan	26.2	25.99	EPI_ISL_421226	464.11	100.00
HZCDC0135	Nasal, oropharyngeal swab	Male	62	2020-01-15	2020-01-22	Infected in Wuhan	35.5	34.79	EPI_ISL_421225	509.26	99.95
HZCDC0162	Nasal, oropharyngeal swab	Male	46	2020-01-21	2020-01-23	/	35.93	34.96	EPI_ISL_421224	488.18	99.61
HZCDC0167	Nasal, oropharyngeal swab	Female	53	2020-01-21	2020-01-23	/	33.33	32.85	EPI_ISL_421223	464.55	98.08
HZCDC6706	Nasal, oropharyngeal swab	Female	37	2020-03-12	2020-03-14	Infected in U.S.A.	26.71	28.10	EPI_ISL_421222	507.37	100.00
HZCDC6789	Nasal, oropharyngeal swab	Male	21	2020-03-13	2020-03-15	Infected in U.K.	22.79	22.41	EPI_ISL_421221	365.88	100.00

Table 2 The list of substitutions in all obtained 17 complete genomes of SARS-CoV-2

Site	Variations	Nanopore SNP quality value	Nanopore depth	Illumina SNP quality value	Illumina depth	Variation type	Gene	Amino acid change	Samples
241	CàT	30.9	447	225.0	7421	Intergenic region	5'UTR	None	HZCDC6706
6187	CàT	58.8	412	228.0	2645	Synonymous variant	orf1ab	His1974His	HZCDC0025
9190	GàT	56.0	299	225.0	7145	Synonymous variant	orf1ab	Val2975Val	HZCDC6789
11083	GàT	31.4/22.8/33.5/32.2	183/178/329/162	228.0/225.0/225.0/225.0	2909/4821/6846/6139	Missense variant	orf1ab	Leu3606Phe	HZCDC0012, HZCDC0013, HZCDC0167, HZCDC6789
14034	TàC	59.2	345	225.0	7981	Missense variant	orf1ab	Met4590Thr	HZCDC6789
14805	CàT	63.2	368	228.0	7689	Missense variant	orf1ab	Thr4847Ile	HZCDC6789
16468	CàT	54.2	647	228.0	1577	Synonymous variant	orf1ab	His5401His	HZCDC6789
17247	TàC	46.1	319	228.0	7958	Missense variant	orf1ab	Val5661Ala	HZCDC6789
19646	TàC	58.1	265	228.0	3637	Synonymous variant	orf1ab	Leu6461Leu	HZCDC6789
21282	AàG	56.2/44.0	221/274	225.0/225.0	1391/2248	Missense variant	orf1ab	Asp7006Gly	HZCDC0012, HZCDC0013
24133	CàT	61.5	691	228.0	6456	Synonymous variant	S	Gly857Gly	HZCDC6789
26144	GàT	35.4/24.9/58.8	412/407/324	228.0/228.0/228.0	7846/7748/7739	Missense variant	ORF3a	Gly251Val	HZCDC0025, HZCDC0167, HZCDC6789

## Figures

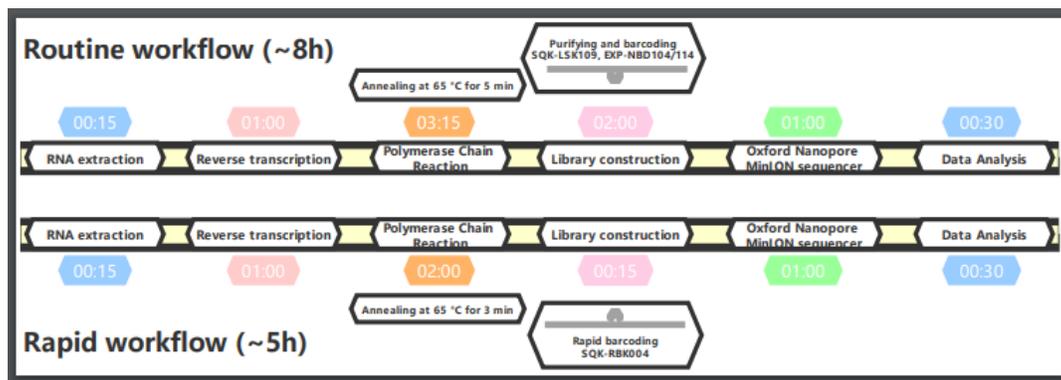


Figure 1

Overview of two nanopore sequencing workflows. The white boxes represent a series of tasks as the components of 8 hours routine workflow and 5 hours rapid workflow respectively. The numbers in the colorful boxes indicate the time required to complete the tasks.

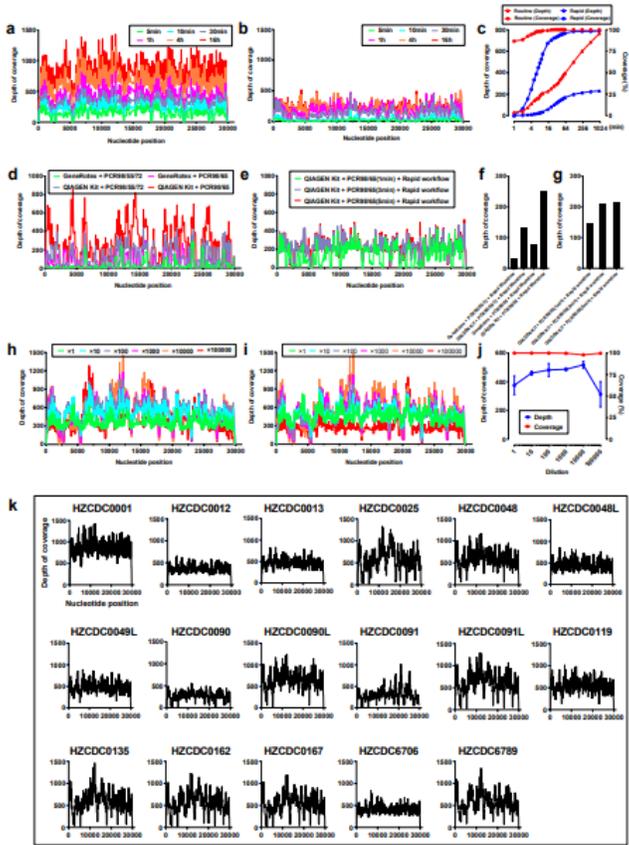


Figure 2

Analysis of genome sequencing data of SARS-CoV-2 viruses by Oxford Nanopore Technologies using MinION sequencer. (a) Trends of the data of depth appearing among 16 hours sequencing using Ligation Sequencing Kit 109. (b) The genome-wide depth of coverage while using Rapid Barcoding Kit 004. (c) Comparative analysis of the average depth and genome-wide coverage between routine workflow and rapid workflow. (d-g) Method optimization of nucleic acid extraction (magnetic beads or spin column method) and PCR amplification procedure. (h-j) Repeated assays of nanopore sequencing of viral RNA with 10 times dilution. (k) The depth of genome-wide coverage appears when applying the optimized methods in clinical specimens.

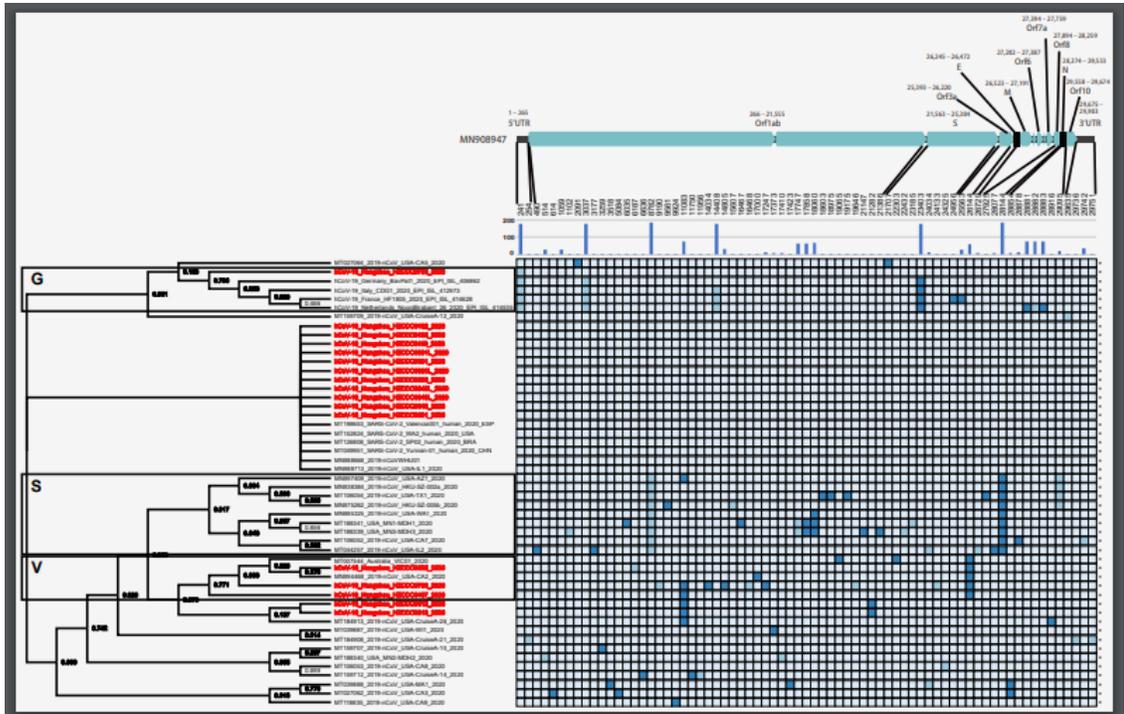


Figure 3

Phylogenetic analysis of seventeen Hangzhou SARS-CoV-2 strains with reference genome sequences. The histogram in the upper panel refer to the frequency of occurrence of these variations. In the lower panel, the missense and synonymous variants were marked as dark blue and light blue respectively. ORF: open reading frame; S: spike protein gene; E: envelope protein gene; M: membrane protein gene; N: nucleocapsid protein gene. Numbers above ORFs are genome positions according to the reference SARS-CoV-2 genome (Genbank accession number MN908947).