Type of the Paper: Virology

Rapid High Throughput Whole Genome Sequencing of SARS-CoV-2 by using One-step RT-PCR Amplification with Integrated Microfluidic System and Next-Gen Sequencing

Tao Li[1], Hye Kyung Chung[1], Papa K. Pireku[1], Brett F. Beitzel[2], Mark A. Sanborn[1], Cynthia Y. Tang[3,4,5,6], Richard Hammer[7], Detlef Ritter[7], XiuFeng Wan[3,4,5,6,8], Irina Maljkovic Berry[1], Jun Hang[1]

[1] Viral Diseases Branch, Walter Reed Army Institute for Research, Silver Spring, Maryland, USA;

[2] US Army Medical Research Institute of Infectious Disease Center for Genome Sciences, Ft. Detrick, Maryland, USA

[3] MU Center for Influenza and Emerging Infectious Diseases (CIEID), University of Missouri, Columbia, Missouri, USA;

[4] Department of Molecular Microbiology and Immunology, School of Medicine, University of Missouri, Columbia, Missouri, USA;

[5] Bond Life Sciences Center, University of Missouri, Columbia, Missouri, USA;

[6] MU Institute for Data Science and Informatics, University of Missouri, Columbia, Missouri, USA

[7] Department of Pathology, School of Medicine, University of Missouri, Columbia, Missouri, USA;

[8] Department of Electrical Engineering & Computer Science, College of Engineering, University of Missouri, Columbia, Missouri, USA

Correspondence: Jun.hang.civ@mail.mil; Tel. +1-301-319-9519

ABSTRACT

The long-lasting global COVID-19 pandemic demands timely genomic investigation of SARS-CoV-2 viruses. Here we report a simple and efficient workflow for whole genome sequencing utilizing one-step RT-PCR amplification on a microfluidic platform, followed by MiSeq amplicon sequencing. The method uses Fluidigm IFC and instruments to amplify 48 samples with 39 pairs of primers in a single step. Application of this method on RNA samples from both viral isolate and clinical specimens demonstrate robustness and efficiency of this method in obtaining the full genome sequence of SARS-CoV-2.

INTRODUCTION

Severe acute respiratory syndrome-coronavirus 2 (SARS-CoV-2) (family *Coronaviridae*, genus *Betacoronavirus*) is responsible for the global pandemic of coronavirus disease 2019 (COVID-19) (1-4). Since its emergence in Wuhan, China, in November/December 2019, the disease has rapidly spread worldwide. As of 31 October 2020, there have been over 45 million confirmed cases, and over one million deaths in 188 countries or regions (https://coronavirus.jhu.edu/) (5-8). In addition to the continuously growing number of SARS-CoV-2 infections, increased complexity and diversity of disease symptoms are also observed. Despite the large number of whole genome sequences of SARS-CoV-2 available in GenBank and other public sources, the unprecedented scale of the viral transmission and the complexity of its pathogenic mechanism demand more genomic data to be produced using cost-effective and quality-consistent methodology (9-12).

Targeted whole genome amplification and next-generation sequencing (NGS) techniques have been used in sequencing SARS-CoV-2 genomes (8, 10, 13). In contrast to the high throughput capacity of NGS, using a conventional multiplexing reverse transcription and polymerase chain reaction (RT-PCR) procedures to amplify the large 30 Kb viral genomic RNA of SARS-CoV-2 is tedious, technically challenging, and has variable contamination risks. A rapid and streamlined approach with fewer manual steps to obtain whole genome amplicons suitable for NGS is desired.

In this study, we utilize an integrated microfluidic nucleic acid amplification system (14), custom primer design, and a one-step RT-PCR program to amplify whole SARS-CoV-2 genomes. After NGS of the amplicons, an in-house developed bioinformatics pipeline is used to rapidly obtain genome sequences with graphic summaries for data and results visualization.

MATERIALS AND METHODS

SARS-CoV-2 RNA samples and quantitative RT-PCR

SARS-CoV-2 RNA samples used in this study include RNA extracted from viral isolate R4717 in the US Army Medical Research Institute of Infectious Disease (USAMRIID) and RNA

extracted from de-identified clinical respiratory specimens. A one-step RT-qPCR method targeting RNA dependent RNA polymerase (RdRp) was used to quantify SARS-CoV-2 RNA. Serially diluted *in vitro* transcripts (IVT), corresponding to nucleotide region 15431 to 15530 of NC_045512.2 for strain Wuhan-Hu-1, were prepared and used as real-time RT-PCR standards for quantification of genome equivalent copy number (GE) of SARS-CoV-2 RNA. Real-time RT-PCR was performed using the protocol from Carman et al. (15) with SuperScript III one-step RT-PCR System and Platinum Taq Polymerase. The QuantStudio 7 Flex Real-Time PCR System and software (ThermoFisher Scientific, Inc.) was used for data acquisition.

Whole genome RT-PCR amplification and Illumina sequencing

For whole genome RT-PCR amplification, 35 primer pairs covering the 29903 bp SARS-CoV-2 reference genome (NC_045512.2) were custom designed by Fluidigm Corporation (South San Francisco, CA) (Table 1). The RT-PCR products are approximately 1 kb for all amplicons. One-step RT-PCR amplification was performed using Fluidigm Access Array (AA) nucleic acids amplification system (Fluidigm Corporation, CA) and SuperScript III one-step RT-qPCR System with Platinum Taq High Fidelity (ThermoFisher Scientific, Inc.). Four additional pairs of primers selected from the ARTIC network protocol v3 (https://www.protocols.io/view/ncov-2019-sequencing-protocol-v3-locost-bh42j8ye).

SARS-CoV-2 R4717 RNA was serially diluted to the concentrations of 10^6, 10^5, 10^4, 10^3, 10^2, 10, and 1 GE/µl (genomic equivalence per microliter). For Fluidigm RT-PCR amplification, each sample well in the Integrated Fluidic Circuit (IFC) chip contained 1.45 µl of RNA sample and 4 µl of sample mixture solution, consisting of 3 µl of 2× Reaction Mix, 0.2 µl of DMSO, 0.05 µl of RNaseOUT, 0.05 µl of SuperScript III RT/Platinum Taq High Fidelity Enzyme Mix, 0.25 µl of 20 × AA Loading Reagent, and 0.7 µl of H2O. Each well of primer mix contained 6 µl of mixture containing one pair of primers at the final concentration of 4 µM for each primer and 2.5 mM MgSO4 in AA Loading Reagent. The primed IFC chip was loaded into Fluidigm FC1 cycler and amplified using the cycling conditions: 50°C for 30 mins (reverse transcription), 94°C for 2 mins, 35 cycles of 94°C for 30 secs, 53°C for 30 secs and 68°C for 90 secs, final extension at 68°C for 7 mins, and kept at 4°C. The RT-PCR products were purified using Agencourt AMPure XP beads (Beckman Coulter, CA, USA) and then analyzed by using the Agilent

Tapestation 4200 System and High Sensitivity DNA D5000 kit (Agilent Technologies, CA, USA) to determine quality and quantity of the amplicons.

The NGS libraries were prepared using Illumina DNA Flex Library kit (Illumina, CA, USA). DNA was fragmentized by tagmentation for 15 mins followed by indexing and library amplification, with 6 cycles used for 25 ng of amplicon or 12 cycles used for 1-9 ng of amplicon. The libraries were quantified by using Agilent Tapestation and DNA D5000 kit and then pooled with equal molar ratio for each library. Pooled libraries were denatured and diluted to a final concentration of 13.5 pM then sequenced using Illumina MiSeq System and Reagent Kit v3 (600 cycles).

Reference genome mapping assembly using ngsmapper program

The full genome of SARS-CoV-2 reference NC_045512.2 was downloaded from GenBank (16) and utilized for reference-based genome assembly using the WRAIR viral disease branch in-house bioinformatics pipeline ngs_mapper version 1.5.0 (https://github.com/VDBWRAIR/ngs_mapper). This pipeline incorporates a series of quality control and assembly processes for the fastq sequence reads retrieved from the Illumina MiSeq and other platforms. These processes includes filtration to drop poorly indexed reads, read trimming based on quality thresholds using Trimommatic (17), read mapping to the reference genome using the Burrows-Wheeler Aligner with maximal exact matches (BWA-MEM) (18), read tagging, variant calling file (VCF) generation using an in-house base caller, read mapping visualizations, fastq statistic generation, and consensus sequence generation from the VCF using an in-house script basecaller.py. The pipeline streamlines bioinformatic analysis by combining multiple tools and consolidating output required for data validation and sequence curation. A minimum Phred base quality score of 35 and a minimum depth of coverage of 10 were utilized as the configuration parameters for this project. To identify variants present at a frequency of 20% or higher, an 80/20 ambiguous position threshold was used. The assembled genomes were further manually curated utilizing bam, consensus and variant calling files generated from the ngs_mapper pipeline. Geneious R10 software, integrated genome viewer (IGV) (19), and MEGA version 7 (20) were used for quality control of ambiguous calls, insertions, deletions and primer-induced mutations. The genomes were processed to only include coding sequence regions by clipping the 5' and 3' untranslated regions. MEGA7 was used to

118 align genomes using default parameters with Multiple Sequence Comparison by Log Expectation
119 (MUSCLE) (21).

120

121 **RESULTS**

122 **Correlation of Fluidigm RT-PCR yield with amount of input SARS-CoV-2 RNA**

123 Fluidigm RT-PCR whole-genome amplification was done in the Access Array microfluidic chip
124 using only 1.45 µl of RNA input for each sample and with a maximum capacity of 48 samples
125 and 48 pairs of primers. In this study, a set of 35 primer pairs, i.e. nCOVF/R1-35 in Table 1,
126 were designed and tested for genome RT-PCR amplification by using serial dilutions of purified
127 RNA from SARS-CoV-2 isolate R4717. The experiment was done with four replicates processed
128 in parallel. A single band of amplicons with expected sizes of approximately 1 kb was seen for
129 all concentrations, with the band intensities correlated with copy numbers of SARS-CoV-2 in the
130 serial dilutions. The correlation between the concentrations of yielded amplicons and the input
131 SARS-CoV-2 genome copy numbers (**Figure 1**) was significant with p-value of 3.38E-03.

132 **Whole genome coverage and alignment depth of SARS-CoV-2 sequence assembly**

133 MiSeq data for the quadruplicated R4717 RNA serial dilutions described above were assembled
134 using ngs_mapper pipeline with SARS-CoV-2 complete genome sequence NC_045512.2 as the
135 mapping reference. As expected, genome assembly results correlated well with SARS-CoV-2
136 copy numbers in each sample. For reactions with 10^4 or higher SARS-CoV-2 RNA copies,
137 complete genome sequences were readily obtained, and importantly, they had uniform coverage
138 depth across the genome with the peaks matching the regions of amplicon alignment (**Figure 2**).
139 For reactions with 10^3 copies of SARS-CoV-2 RNA, the assembled genome sequences were
140 nearly complete, except for two small dip/gap at positions 1870-2500 and 16800-17700. The
141 dip/gap regions were successfully filled by adding two extra primer pairs for each region to the
142 panel of 35 primer pairs. These additional primers were selected from the ARTIC v3 protocol,
143 paired to cover the two regions and added into four separate Fluidigm primer wells. In total 39
144 pairs of primers were applied for whole genome amplification in a single RT-PCR reaction using
145 Fluidigm nucleic acids amplification system.

SARS-CoV-2 genome sequencing of clinical respiratory specimens

This method was subsequently used in genome sequencing of SARS-CoV-2 in RNA extracts purified from COVID-19 positive nasopharyngeal swabs. The set of 29 samples contained a wide range of titers, with RT-qPCR Ct values from 13.5 to 33.4 (379 to 2.72×10^8 GE/µl) of SARS-CoV-2. The RT-PCR cDNA yield has significant correlation with viral titers in the Ct range of 20 to 35, with p-value of 1.26E-05. For the samples with Ct values below 20, or exceedingly high concentrations of 1.0×10^7 or greater, RT-PCR yields were substantially lower than projected (**Figure 3**). Importantly, this observation suggests that severe suppression of PCR could occur when samples of extremely high titer are used. Nevertheless, full or nearly complete genome coverage was obtained for all clinical specimens with titers above 1.0×10^4 GE/µl, or approximate Ct value of 29.

Comparison of assembled consensus sequences and nucleotide variations

Fluidigm one-step RT-PCR protocol was applied to SARS-CoV-2 RNA of highly varied titers. For SARS-CoV-2 isolate R4717 RNA 10× dilutions (**Figure 1**), a total of 14 full genomes were assembled and curated. All the full genome consensus sequences were identical except for two replicates of 1.0×10^4 GE/µl. One sample had the ambiguous call Y (T or C) in the alignment nucleotide position 2105, while the other three replicates had a C. Another sample had a T in position 23260 while all the other three had the ambiguous call Y. These were the only ambiguous positions found in the samples, showing low sample diversity at a variant frequency of 20% or higher. Together this sequencing approach produces high accuracy results.

DISCUSSION

Even before the disease was named as COVID-19, the sequence was swiftly determined using next-gen sequencing technologies and SARS-CoV-2 was identified as the causative pathogen for the emerging acute respiratory disease (5, 22). The first sequence was made publicly available immediately with a massive number of sequences subsequently generated and shared, which has greatly facilitated research and development (23-25). All the efforts in developing, improving, and sharing materials and/or methods has played an essential role in sequence-based investigations. All the known sequencing protocols use conventional PCR apparatuses with

differences in design and selection of primers and reaction parameters. In this study, we applied one-step RT-PCR protocol on a microfluidic platform (14) to establish a convenient workflow with throughput, speed, simplicity, consistency, and yield suitable for COVID-19 genome sequencing. Fluidigm Access Array IFC holds 48 RNA samples (inlets) and 48 primer pairs (inlets). Steps for mixing sample with primers are obviated, which not only substantially reduces pipetting manipulation but also effectively mitigates the chance of sample-to-sample cross-contamination. In this report, we selected 39 primer pairs to obtain even genome coverage. Nine more individual pairs of primers can be easily added to the panel of 48 primer inlets to quickly address emerging SARS-CoV-2 genetic divergence. Moreover, the total number of primer pairs can be further increased without difficulty by pooling together several compatible primer pairs and adding them into one primer inlet. Each sample is mixed with individual primer pairs in IFC microfluidic chambers for nano-liter (nl) RT-PCR amplifications in a simplex independent reaction manner. In contrast, conventional PCR methods often need optimization of primer pooling and reaction parameters to circumvent primer-to-primer interference and to avoid highly variable yields among amplicons.

Since input RNA samples are partitioned into individual nl reaction chambers to cross-mix with individual primer pairs, microfluidic applications including Fluidigm IFC require sufficient genomic copies in order to achieve whole genome amplification. In consequence, Fluidigm amplification based whole genome sequencing has limitations in sequencing low titer samples. For RNA samples with 1000 GE/μl or lower concentration, multiplex RT-PCR based methods or SARS-CoV-2 hybridization-based enrichment method might be a more suitable choice for whole genome sequencing (26-28). The addition of a first-strand cDNA synthesis step prior to Fluidigm amplification, with a change of Fluidigm thermocycling program from RT-PCR to PCR, may help increase genome coverage for low titer samples. Many, if not all, COVID-19 specimens are tested with quantitative molecular tests and the Ct values or equivalent titer scores are readily available for deciding whether one-step Fluidigm amplification-based genome sequencing protocol is appropriate.

Using this method, very few sequence assembly errors were observed throughout the tested SARS-CoV-2 sample genomes. These errors might be due to PCR, sequencing or basecalling algorithm errors, as well as due to normal fluctuations in the minor variant

frequencies between the sample aliquots. When assembled sequences show potentially significant nucleotide alterations or indels, thorough examination of the data quality processing, primer trimming, curation of sequence assembly, and detailed laboratory records are needed. Whenever possible, running replicate samples, repeating the experiment entirely, or using other methods are important to validate genome variations and minor variants.

In conclusion, our study demonstrates a convenient SARS-CoV-2 whole genome sequencing protocol by incorporating one-step RT-PCR amplification, microfluidic technology, and next-generation sequencing to achieve a simple and fast workflow with consistent and quality data. The performance of the protocol was verified using viral isolate RNA and tested by sequencing clinical respiratory samples of varying viral titers.

Funding: Global Emerging Infections Surveillance and Response System (GEIS), Division of the Armed Forces Health Surveillance Branch.

Acknowledgments: We thank Mr. James S. Hilaire, Ms. Nicole R. Nicholas, Mr. Tuan K. Nguyen and Ms. April N. Griggs for their assistance in project management, sample tracking, storage and retrieval.

Conflicts of Interest: The authors declare no conflict of interest.

Disclaimer: Material has been reviewed by the authors' respective institutions. There is no objection to its presentation and/or publication. The views expressed here are those of the authors and do not reflect the official policy of the Department of the Army, Department of the Navy, Department of Defense or U.S. Government. This is the work of U.S. government employees and may not be copyrighted (17 USC 105).

Figure legends

Figure 1 Agilent TapeStation analysis of SARS-CoV2 whole genome RT-PCR amplification. A. TapeStation profile of RT-PCR products. From *left* to *right*, 1, 10, 10^2, 10^3, 10^4, 10^5, and 10^6 copies of SARS-CoV2 RNA from isolate R4717 were used. **B.** SARS-CoV-2 amplicon concentrations for RT-PCR with serial dilutions of R4717 RNA as inputs. The concentration values were the average and standard deviation from four replicates.

Figure 2 Genome coverage and sequence mapping alignment depth for MiSeq reads data. SARS-CoV-2 isolate R4717 RNA serial dilutions were subjected to Fluidigm RT-PCR amplification and MiSeq sequencing. Data were mapped using NC_045512.2 as reference. **A.** Genome coverage for 10× serial dilutions of R4717 RNA from 1.0E+06 (*top*) to 1 (*bottom*) copies. Mapping depth was indicated in colors, *green* for normal (depth ≥ 10), *cyan* for dip (depth 1-10), *red* for gap. Results for four replicates were shown. **B.** Sequence mapping graphs for 10^6, 10^5, 10^4 and 10^3 copies of RNA from one set of serial dilutions.

Figure 3 SARS-CoV2 whole genome RT-PCR amplification. Twenty nine RNA extracts from COVID-19 positive nasopharyngeal swabs were amplified and sequenced. RT-PCR yield was quantified by using Agilent TapeStation. Reference mapping coverage was determined by mapping MiSeq read data to reference genome sequence NC_045512.2.

REFERENCES

1. Hu B, Guo H, Zhou P, Shi ZL. 2020. Characteristics of SARS-CoV-2 and COVID-19. Nat Rev Microbiol doi:10.1038/s41579-020-00459-7.
2. Hon KL, Leung KKY, Leung AKC, Sridhar S, Qian S, Lee SL, Colin AA. 2020. Overview: The history and pediatric perspectives of severe acute respiratory syndromes: Novel or just like SARS. Pediatr Pulmonol 55:1584-1591.
3. Habibzadeh P, Stoneman EK. 2020. The Novel Coronavirus: A Bird's Eye View. Int J Occup Environ Med 11:65-71.
4. Gralinski LE, Menachery VD. 2020. Return of the Coronavirus: 2019-nCoV. Viruses 12.
5. Wang C, Horby PW, Hayden FG, Gao GF. 2020. A novel coronavirus outbreak of global health concern. Lancet 395:470-473.
6. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, Bi Y, Ma X, Zhan F, Wang L, Hu T, Zhou H, Hu Z, Zhou W, Zhao L, Chen J, Meng Y, Wang J, Lin Y, Yuan J, Xie Z, Ma J, Liu WJ, Wang D, Xu W, Holmes EC, Gao GF, Wu G, Chen W, Shi W, Tan W. 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet 395:565-574.
7. Fauver JR, Petrone ME, Hodcroft EB, Shioda K, Ehrlich HY, Watts AG, Vogels CBF, Brito AF, Alpert T, Muyombwe A, Razeq J, Downing R, Cheemarla NR, Wyllie AL, Kalinich CC, Ott IM, Quick J, Loman NJ, Neugebauer KM, Greninger AL, Jerome KR, Roychoudhury P, Xie H, Shrestha L, Huang ML, Pitzer VE, Iwasaki A, Omer SB, Khan K, Bogoch, II, Martinello RA, Foxman EF, Landry ML, Neher RA, Ko AI, Grubaugh ND. 2020. Coast-to-Coast Spread of SARS-CoV-2 during the Early Epidemic in the United States. Cell 181:990-996.e5.
8. Worobey M, Pekar J, Larsen BB, Nelson MI, Hill V, Joy JB, Rambaut A, Suchard MA, Wertheim JO, Lemey P. 2020. The emergence of SARS-CoV-2 in Europe and North America. Science doi:10.1126/science.abc8169.
9. Rockett RJ, Arnott A, Lam C, Sadsad R, Timms V, Gray KA, Eden JS, Chang S, Gall M, Draper J, Sim EM, Bachmann NL, Carter I, Basile K, Byun R, O'Sullivan MV, Chen SC, Maddocks S, Sorrell TC, Dwyer DE, Holmes EC, Kok J, Prokopenko M, Sintchenko V. 2020. Revealing COVID-19 transmission in Australia by SARS-CoV-2 genome sequencing and agent-based modeling. Nat Med 26:1398-1404.
10. Oude Munnink BB, Nieuwenhuijse DF, Stein M, O'Toole Á, Haverkate M, Mollers M, Kamga SK, Schapendonk C, Pronk M, Lexmond P, van der Linden A, Bestebroer T, Chestakova I, Overmars RJ, van Nieuwkoop S, Molenkamp R, van der Eijk AA, GeurtsvanKessel C, Vennema H, Meijer A, Rambaut A, van Dissel J, Sikkema RS, Timen A, Koopmans M. 2020. Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. Nat Med 26:1405-1410.
11. Licastro D, Rajasekharan S, Dal Monego S, Segat L, D'Agaro P, Marcello A. 2020. Isolation and Full-Length Genome Characterization of SARS-CoV-2 from COVID-19 Cases in Northern Italy. J Virol 94.
12. Islam MR, Hoque MN, Rahman MS, Alam A, Akther M, Puspo JA, Akter S, Sultana M, Crandall KA, Hossain MA. 2020. Genome-wide analysis of SARS-CoV-2 virus strains circulating worldwide implicates heterogeneity. Sci Rep 10:14004.
13. Stefanelli P, Faggioni G, Lo Presti A, Fiore S, Marchi A, Benedetti E, Fabiani C, Anselmo A, Ciammaruconi A, Fortunato A, De Santis R, Fillo S, Capobianchi MR, Gismondo MR, Ciervo A, Rezza G, Castrucci MR, Lista F, On Behalf Of Iss Covid-Study G. 2020. Whole genome and phylogenetic analysis of two SARS-CoV-2 strains isolated in Italy in January and February 2020: additional clues on multiple introductions and further circulation in Europe. Euro Surveill 25.

14. Wang M, Escudero-Ibarz L, Moody S, Zeng N, Clipson A, Huang Y, Xue X, Grigoropoulos NF, Barrans S, Worrillow L, Forshew T, Su J, Firth A, Martin H, Jack A, Brugger K, Du MQ. 2015. Somatic Mutation Screening Using Archival Formalin-Fixed, Paraffin-Embedded Tissues by Fluidigm Multiplex PCR and Illumina Sequencing. J Mol Diagn 17:521-32.
15. Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DK, Bleicker T, Brunink S, Schneider J, Schmidt ML, Mulders DG, Haagmans BL, van der Veer B, van den Brink S, Wijsman L, Goderski G, Romette JL, Ellis J, Zambon M, Peiris M, Goossens H, Reusken C, Koopmans MP, Drosten C. 2020. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Euro Surveill 25.
16. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang YZ. 2020. A new coronavirus associated with human respiratory disease in China. Nature 579:265-269.
17. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-20.
18. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754-60.
19. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. Nat Biotechnol 29:24-6.
20. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 33:1870-4.
21. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792-7.
22. Eurosurveillance Editorial T. 2020. Note from the editors: novel coronavirus (2019-nCoV). Euro Surveill 25.
23. Corman VM, Muller MA, Costabel U, Timm J, Binger T, Meyer B, Kreher P, Lattwein E, Eschbach-Bludau M, Nitsche A, Bleicker T, Landt O, Schweiger B, Drexler JF, Osterhaus AD, Haagmans BL, Dittmer U, Bonin F, Wolff T, Drosten C. 2012. Assays for laboratory confirmation of novel human coronavirus (hCoV-EMC) infections. Euro Surveill 17.
24. Nalla AK, Casto AM, Huang MW, Perchetti GA, Sampoleo R, Shrestha L, Wei Y, Zhu H, Jerome KR, Greninger AL. 2020. Comparative Performance of SARS-CoV-2 Detection Assays Using Seven Different Primer-Probe Sets and One Assay Kit. J Clin Microbiol 58.
25. Yuan M, Wu NC, Zhu X, Lee CD, So RTY, Lv H, Mok CKP, Wilson IA. 2020. A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. Science 368:630-633.
26. Itokawa K, Sekizuka T, Hashino M, Tanaka R, Kuroda M. 2020. Disentangling primer interactions improves SARS-CoV-2 genome sequencing by multiplex tiling PCR. PLoS One 15:e0239403.
27. Paden CR, Tao Y, Queen K, Zhang J, Li Y, Uehara A, Tong S. 2020. Rapid, Sensitive, Full-Genome Sequencing of Severe Acute Respiratory Syndrome Coronavirus 2. Emerg Infect Dis 26:2401-2405.
28. Nasir JA, Kozak RA, Aftanas P, Raphenya AR, Smith KM, Maguire F, Maan H, Alruwaili M, Banerjee A, Mbareche H, Alcock BP, Knox NC, Mossman K, Wang B, Hiscox JA, McArthur AG, Mubareka S. 2020. A Comparison of Whole Genome Sequencing of SARS-CoV-2 Using Amplicon-Based Sequencing, Random Hexamers, and Bait Capture. Viruses 12.

340 **Table 1** Primers for RT-PCR amplification of SARS-CoV-2 genome using Fluidigm system

Forward Primer	Forward Primer Sequence	Reverse Primer	Reverse Primer Sequence	Amplicon Length	From	To
nCoVF1	TTCCCAGGTAACAAACCAACCAA	nCoVR1	AGGTGTCTGCAATTCATAGCTCTT	1011	17	1027
nCoVF2	CCGAACAACTGGACTTTATTGACAC	nCoVR2	GCCTTCTGTAAAACACGCACAGA	1048	912	1959
nCoVF3	ACTGAGTCCTCTTTATGCATTTG	nCoVR3	AGCATCTGCCACAACACA	1048	1852	2899
nCoVF32	AGAAGAAACTGGCCTACTCATGC	nCoVR32	ACATTGGCTGCATTAACAACCAC	1017	2937	3953
nCoVF33	TTTGGAATTTGGTGCCACTTCTG	nCoVR33	CCTCTTGAACAACATCACCCACT	1004	3642	4645
nCoVF4	AGAGTTTGTGTAGATACTGTTCG	nCoVR4	TATAGGAACCAGCAAGTGAGATG	1048	3752	4799
nCoVF5	GTTTCTGTTTCTTCACCTGATGC	nCoVR5	TGGTGCTGACATCATAACAAAAG	1008	4688	5695
nCoVF6	CCTTGTACGTGTGGTAAACAAGC	nCoVR6	GCTAAACCATGAGTAGCAAGGGT	1028	5621	6648
nCoVF7	TACAGAAGAGGTTGGCCACACAG	nCoVR7	TGTACATTCGACTCTTGTTGCTCT	1015	6523	7537
nCoVF8	ATGTGCATGTTGTAGACGGTTGT	nCoVR8	GCAGCACTACGTATTTGTTTTCGT	1009	7455	8463
nCoVF9	GCAGGTAGCAAAAAGTCACAACA	nCoVR9	AGATGCTGATATGTCCAAAGCAC	1036	8368	9403
nCoVF10	GGAGTTTTCTGTGGTGTAGATGC	nCoVR10	AGGTGTCTTAGGATTGGCTGTAT	1041	9311	10351
nCoVF11	TCTAAAGTTGCGTAGTGATGTGCT	nCoVR11	TCCAGTTTGAGCAGAAAGAGGTC	994	9835	10828
nCoVF12	GTTTGTTCGCATTCAACCAGGAC	nCoVR12	ACACTCTCCTAGCACCATCATCA	1031	10360	11390
nCoVF34	ATATGCCTGCTAGTTGGGTGATG	nCoVR34	CTGCATCACGGTCAAATTCAGAT	1026	11726	12751
nCoVF35	GCCTCAGAGTTTAGTTCCCTTCC	nCoVR35	ATTAGTGATTGGTTGTCCCCCAC	1047	12598	13644
nCoVF13	AAGCTGGTAATGCAACAGAAGTG	nCoVR13	TTTCGCATGGCATCACAGAATTG	1011	13023	14033
nCoVF14	TGTAGAAAACCCAGATATATTACGC	nCoVR14	ATTTGTCTAGGTTGTTGACGATG	1007	13935	14941
nCoVF15	TTGATTGTTACGATGGTGGCTGT	nCoVR15	AGGTACACATAATCATCACCCTG	1048	14879	15926
nCoVF16	TGTCTGAAGCAAAATGTTGGACT	nCoVR16	CAACAGCATCACCATAGTCACCT	1044	15821	16864
nCoV-2019_54_LEFT	TGAGTTAACAGGACACATGTTAGACA	nCoV-2019_56_RIGHT	ACACTATGCGAGCAGAAGGGTA	1033	16119	17152
nCoV-2019_55_LEFT	ACTCAACTTTACTTAGGAGGTATGAGCT	nCoV-2019_57_RIGHT	GTAATTGAGCAGGGTCGCCAAT	1035	16417	17452
nCoVF17	ACCTAGACCACCACTTAACCGAA	nCoVR17	CAGCTTTTCTCCAAGCAGGGTTA	1016	16749	17764
nCoV-2019_56_LEFT	ACCTAGACCACCACTTAACCGA	nCoV-2019_59_RIGHT	AAGAGTCCTGTTACATTTTCAGCTTG	1313	16749	18062
nCoV-2019_58_LEFT	TGATTTGAGTGTTGTCAATGCCAGA	nCoV-2019_60_RIGHT	GGTACCAACAGCTTCTCTAGTAGC	966	17382	18348
nCoVF18	GCTTAAAGCACATAAAGACAAATCA	nCoVR18	GTGCGCTCAGGTCCTATTTT	1041	17616	18656
nCoVF19	GTCTTATGGGCACATGGCTTTGA	nCoVR19	AGCCACATTTTCTAAACTCTGAAGTC	1050	18589	19638
nCoVF20	ACATGATGATCTCAGCTGGCTTT	nCoVR20	TCACTTTGACAACCTTAGAAACTACA	1047	19535	20581
nCoVF21	TTGGAGAAGCCGTAAAAACACAG	nCoVR21	TTTATAGCCACGGAACCTCCAAG	1045	20123	21167
nCoVF22	TTTAAGACAGTGGTTGCCTACGG	nCoVR22	GGACTGGGTCTTCGAATCTAAAGT	1001	20910	21910
nCoVF23	TAAGGGGTACTGCTGTTATGTCTT	nCoVR23	TCAAGTGCACAGTCTACAGCATC	1025	21419	22443
nCoVF24	GGGTTATCTTCAACCTAGGACTT	nCoVR24	ACATCCTGATAAAGAACAGCAAC	1044	22363	23406
nCoVF25	ACAATTTGGCAGAGACATTGCTG	nCoVR25	AAACCTATAAGCCATTTGCATAGC	1030	23251	24280
nCoVF26	CGGGTACAATCACTTCTGGTTGG	nCoVR26	ACTATGGCAATCAAGCCAGCTAT	1048	24198	25245
nCoVF27	TTCAAAAAGAAATTGACCGCCTC	nCoVR27	CCGTCGATTGTGTGAATTTGGAC	1047	25098	26144
nCoVF28	ACATGTTACCTTCTTCATCTACA	nCoVR28	GACTGTATGCAGCAAAACCTG	1045	26070	27114
nCoVF29	GTGACATCAAGGACCTGCCTAAA	nCoVR29	ATAGGACACGGGTCATCAACTACAT	1012	26998	28009
nCoVF30	CTGTAGCTGCATTTCACCAAGAA	nCoVR30	CAAGCTGGTTCAATCTGTCAAGC	1037	27928	28964
nCoVF31	GAACTTCTCCTGCTAGAATGGC	nCoVR31	TCACATGGGGATAGCACTAC	961	28884	29844

341

342

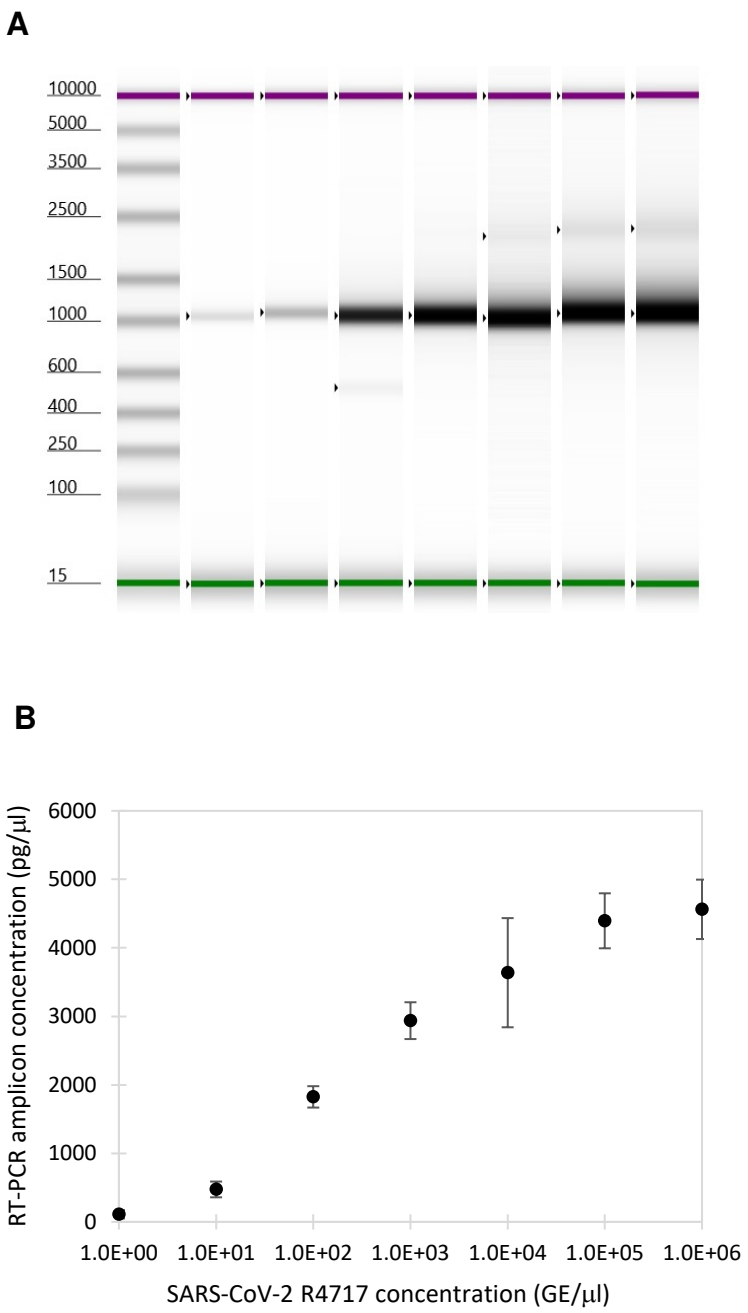

Figure 1 Agilent TapeStation analysis of SARS-CoV2 whole genome RT-PCR amplification. A. TapeStation profile of RT-PCR products. From *left* to *right*, 1, 10, 10^2, 10^3, 10^4, 10^5, and 10^6 copies of SARS-CoV2 RNA from isolate R4717 were used. **B.** SARS-CoV-2 amplicon concentrations for RT-PCR with serial dilutions of R4717 RNA as inputs. The concentration values were the average and standard deviation from four replicates.

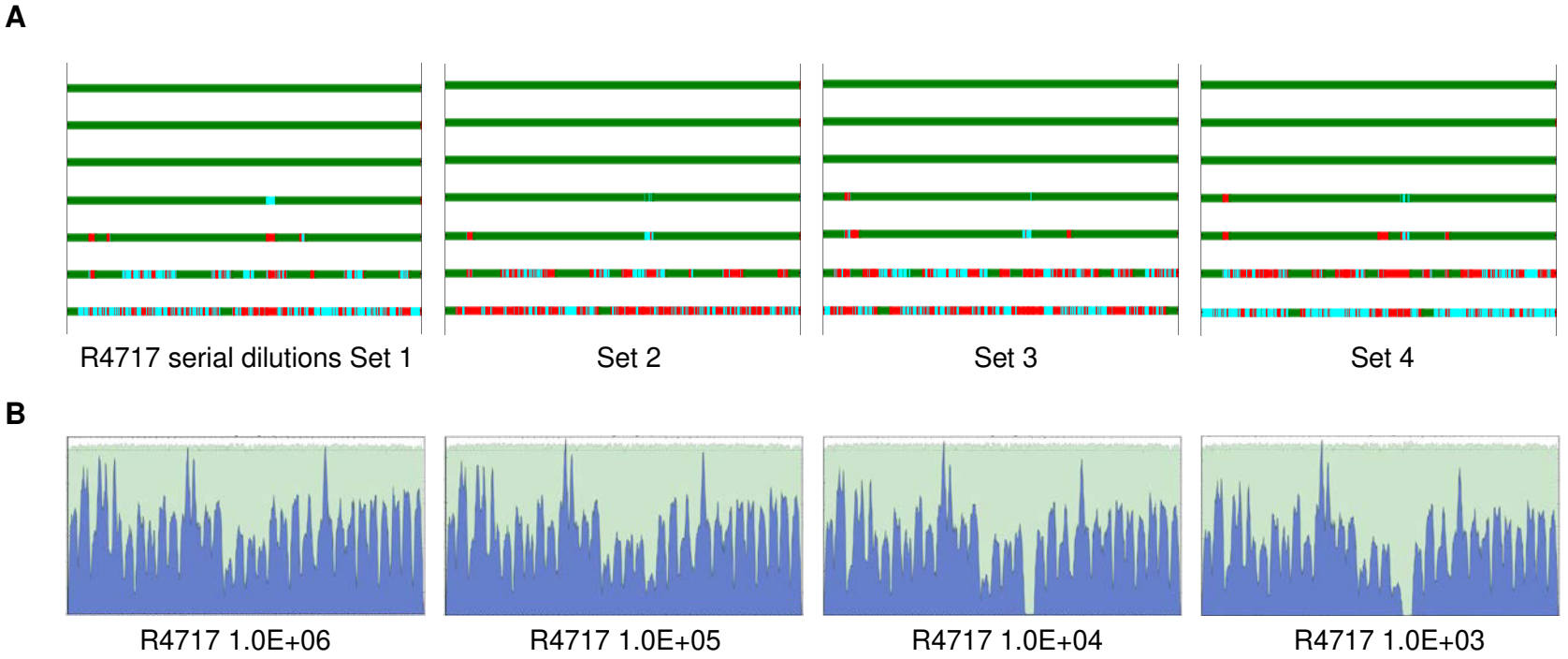

Figure 2 Genome coverage and sequence mapping alignment depth for MiSeq reads data. SARS-CoV-2 isolate R4717 RNA serial dilutions were subjected to Fluidigm RT-PCR amplification and MiSeq sequencing. Data were mapped using NC_045512.2 as reference. **A.** Genome coverage for 10× serial dilutions of R4717 RNA from 1.0E+06 (*top*) to 1 (*bottom*) copies. Mapping depth was indicated in colors, *green* for normal (depth ≥ 10), *cyan* for dip (depth 1-10), *red* for gap. Results for four replicates were shown. **B.** Sequence mapping graphs for 10^6, 10^5, 10^4 and 10^3 copies of RNA from one set of serial dilutions.

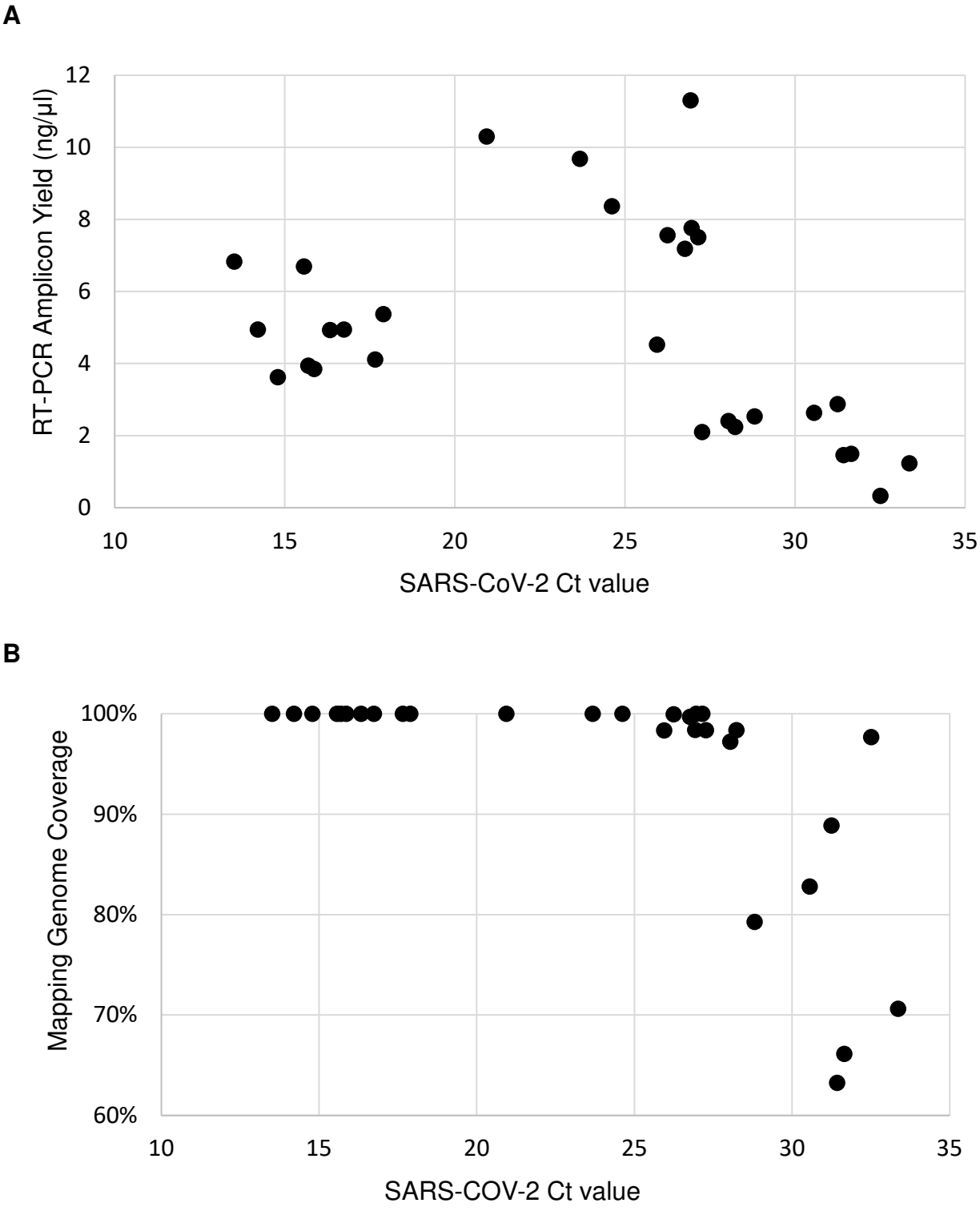

Figure 3 SARS-CoV2 whole genome RT-PCR amplification. Twenty nine RNA extracts from COVID-19 positive nasopharyngeal swabs were amplified and sequenced. RT-PCR yield was quantified by using Agilent TapeStation. Reference mapping coverage was determined by mapping MiSeq read data to reference genome sequence NC_045512.2.