



RESEARCH ARTICLE

Multiplex sequencing of SARS-Cov-2 genome directly from clinical samples using the Ion Personal Genome Machine (PGM)

Tan, K.K.¹, Tiong, V.¹, Tan, J.Y.¹, Wong, J.E.¹, Teoh, B.T.¹, Abd-Jamil, J.¹, Johari, J.¹, Nor'e, S.S.¹, Khor, C.S.¹, Yaacob, C.N.¹, Zulkifli, M.M.S.¹, CheMatSeri, A.¹, Mahfodz, N.H.¹, Azizan, N.S.¹, AbuBakar, S.^{1,2*}

¹Tropical Infectious Diseases Research and Education Centre (TIDREC), Higher Institution Centre of Excellence (HiCoE), University of Malaya, 50603 Kuala Lumpur, Malaysia

²Department of Medical Microbiology, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia

*Corresponding author: sazaly@um.edu.my

ARTICLE HISTORY

Received: 6 May 2021
Revised: 17 June 2021
Accepted: 17 June 2021
Published: 31 July 2021

ABSTRACT

Various methods have been developed for rapid and high throughput full genome sequencing of SARS-CoV-2. Here, we described a protocol for targeted multiplex full genome sequencing of SARS-CoV-2 genomic RNA directly extracted from human nasopharyngeal swabs using the Ion Personal Genome Machine (PGM). This protocol involves concomitant amplification of 237 gene fragments encompassing the SARS-CoV-2 genome to increase the abundance and yield of viral specific sequencing reads. Five complete and one near-complete genome sequences of SARS-CoV-2 were generated with a single Ion PGM sequencing run. The sequence coverage analysis revealed two amplicons (positions 13 751-13 965 and 23 941-24 106), which consistently gave low sequencing read coverage in all isolates except 4Apr20-64-Hu. We analyzed the potential primer binding sites within these low covered regions and noted that the 4Apr20-64-Hu possess C at positions 13 730 and 23 929, whereas the other isolates possess T at these positions. The genome nucleotide variations observed suggest that the naturally occurring variations present in the actively circulating SARS-CoV-2 strains affected the performance of the target enrichment panel of the Ion AmpliSeq™ SARS CoV 2 Research Panel. The possible impact of other genome nucleotide variations warrants further investigation, and an improved version of the Ion AmpliSeq™ SARS CoV 2 Research Panel, hence, should be considered.

Keywords: SARS-CoV-2; COVID-19; infectious diseases; next generation sequencing; Malaysia.

INTRODUCTION

The Coronavirus Disease 2019 (COVID-19) pandemic is a serious public health crisis (World Health Organization, 2020b). The disease was first reported as a flu-like illness, causing pneumonia in Wuhan, Hubei Province, China, in late December 2019 (Li *et al.*, 2020; Wang *et al.*, 2020). A novel betacoronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) which was initially named as 2019 novel coronavirus (2019-nCoV) is the etiologic agent of this ongoing COVID-19 pandemic (Wu *et al.*, 2020). Similar to other coronaviruses, SARS-CoV-2 is primarily transmitted between people via close contact with the infected person, exposure to infectious respiratory droplets or contaminated surface (World Health Organization, 2020a). Genetically, the SARS-CoV-2 clustered closely with viruses recovered in bats and pangolins, suggesting its potential zoonotic origin (Zhang *et al.*, 2020; Zhou *et al.*, 2020). This SARS-CoV-2 related lineage together with severe acute respiratory syndrome coronavirus

(SARS-CoV) and the SARS-CoV-related bats viruses forms the sarbecovirus subgenus (Lu *et al.*, 2020).

The first SARS-CoV-2 complete genome sequence, Wuhan-Hu-1 (MN908947), was derived from RNA extracted directly from the patient's bronchoalveolar lavage fluid using the whole genome meta-transcriptomics approach (Wu *et al.*, 2020). This unbiased sequencing approach allows culture-free detection of novel (Wu *et al.*, 2020) or known pathogens (Marcelino *et al.*, 2019). The meta-transcriptomics approach, however, is not always able to produce the entire genome sequence due in part to the presence of host RNAs that contribute to the majority of the generated sequencing reads (Wu *et al.*, 2020). The performance of the meta-transcriptomic approach also relies heavily on the virus titer and types of sample (Fischer *et al.*, 2015). The use of this meta-transcriptomics approach for the sequencing of a large number of SARS-CoV-2 samples is therefore limited. After releasing the first SARS-CoV-2 complete genome sequence, extensive efforts have been made to develop various

sequence-based next-generation sequencing (NGS) library preparation methods to increase the yield of viral specific sequencing reads (Campos *et al.*, 2020; Pillay *et al.*, 2020; Quick, 2020). PCR amplification and hybridization are two commonly used sequence-based approaches that enable SARS-CoV-2 target enrichment (Xiao *et al.*, 2020). This is an effective step to ensure successful complete genome sequencing, especially in samples with low virus titer. The Ion AmpliSeq™ SARS CoV 2 Research Panel (Ion Torrent, Thermo Scientific) is an amplicon-based SARS-CoV-2 target enrichment approach for full genome sequencing of SARS-CoV-2. The published workflow however, was designed to be used with an Ion GeneStudio™ S5 Series System (MAN0019277 Rev.A.O).

Here, we described i) a modified Ion Torrent library preparation of Ion AmpliSeq™ SARS CoV 2 Research Panel (Ion Torrent, Thermo Scientific), ii) the sequencing workflow using the Ion PGM, and iii) data analysis using the Torrent Suite™ Software (Ion Torrent, Thermo Scientific). From the exercise, we generated six complete and near-complete SARS-CoV-2 genome sequences from RNA samples extracted from patient's nasopharyngeal swab in a single Ion PGM (Ion Torrent, Thermo Scientific) run. Our findings revealed the possible impact of genome sequence variations affecting the performance of the Ion AmpliSeq™ SARS CoV 2 Research Panel. An improved version of the Ion AmpliSeq™ SARS CoV 2 Research Panel should be considered to allow better and faster sequencing of SARS-CoV-2 complete genome in a single run.

MATERIALS AND METHODS

Ethics statement

The study was approved by the Medical Ethics Committee of the University Malaya Medical Centre (MREC ID no.: 20201228-9626) and informed consent was waived based on the basis that this is a retrospective study using anonymous samples and data. All protocols were performed in accordance with relevant guidelines and regulations.

SARS-CoV-2 RNA samples

The SARS-CoV-2 RNA samples used were obtained from the COVID-19 Testing Laboratory at the Tropical Infectious Diseases Research and Education Centre (TIDREC), Universiti Malaya (UM) as part of the Ministry of Science, Technology, and Innovation (MOSTI)-Ministry of Higher Education (MOHE) COVID-19 testing initiative. Six RNA samples tested positive using the real-time reverse-transcription polymerase chain reaction (rRT-PCR) were randomly chosen for the sequencing study. No identifiers linking to any specific individual were provided along with the RNA samples.

Amplicon-based enrichment of SARS-CoV-2 template preparation and sequencing

For the sequencing, the cDNA was synthesized using the SuperScript® IV Reverse Transcriptase (Invitrogen, Thermo Scientific). The SARS-CoV-2 full genome amplicon libraries were generated with Ion AmpliSeq™ SARS CoV 2 Research Panel (Ion Torrent, Thermo Scientific) using the Ion AmpliSeq™ Library Kit Plus (Ion Torrent, Thermo Scientific): Ion AmpliSeq™ SARS CoV 2 Research Panel (MAN0019277). Briefly, the cDNA was used to generate two amplification pools with amplicon length ranging from 125-275 bp, covering 99% of the SARS-CoV-2 genome. After amplification, the amplified products of Pool 1 and Pool 2 were combined. The partial digestion of amplicons was performed to remove primers and phosphorylate the amplicons. The selected barcode and adaptor were ligated to amplicons for each sample to allow

sample multiplexing for sequencing. The barcode-ligated libraries were purified using Agencourt™ AMPure™ XP Reagent (Beckman Coulter). The purified amplified libraries were quantitated using the Qubit™ dsDNA HS Assay Kit (Invitrogen, Thermo Scientific). Quantified libraries underwent template amplification and enrichment using Ion OneTouch™ 2 System (Ion Torrent, Thermo Scientific) according to the manufacturer's protocol. The sequencing reaction was established using the Ion PGM Hi-Q view sequencing kit. The enriched template was loaded onto the Ion 318™ Chip Kit v2 BC as implemented using the Ion PGM (Ion Torrent, Thermo Scientific). A flow of 550 was set in order to gain complete sequencing of longer fragments.

Genome assembly and analysis using Torrent Suite Software

The generated sequencing reads were analyzed using the in-built workflow and reference file provided by the manufacturer in Torrent Suite™ Software (Ion Torrent, Thermo Scientific). The generated reads were mapped to SARS-CoV-2 reference strains, Wuhan-Hu-1 (GenBank accession number: MN908947), and the five human expression control's gene sequences provided by the manufacturer. The extend of sequence coverage on the target genome regions was performed using coverageAnalysis v5.12.0.0. Genome assembly was conducted using the IRMAreport v1.2.1.0.

RESULTS

Overall, 5.1 million usable reads with the mean read length of 209 bp for six SARS-CoV-2 isolates were generated in a single run on the Ion 318™ Chip Kit v2 BC. The Ion AmpliSeq™ SARS CoV 2 Research Panel comprised 237 amplicons specific to the SARS CoV-2 genome and ten amplicons specific to five human expression controls. The on target sequence coverage (SARS-CoV-2 and human expression control) was 99.99% for all six SARS-CoV-2 isolates used in this study with an average mean depth of more than 1000X and coverage uniformity of approximately 96% and above. The amplicon targets drop out occurred on the human expression control (Supplementary Table 1 available upon request). In order to reflect the actual read distribution on the SARS-CoV-2 genome, we reanalyzed the data after removing the human expression control amplicons from the reference genome file. As expected, the mean base depth and base coverage uniformity for all six isolates improved to 100% on target mapping and more than 97% base coverage (Table 1).

In order to assess the amplification uniformity, we looked into the data on the total number of amplicons mapped onto each target region. Among the 237 SARS-CoV-2 specific amplicon targets, 13 had less than 20% of average reads per amplicon in at least one of the SARS-CoV-2 (Table 2 and Table 3). Among the 13 amplicon targets, two of the targets, r1_1.14.786182 and r1_1.25.388943, corresponded to amplicons at positions 13 751-13 965 and 23 941-24 106, respectively, showed extremely low coverage (<5% of the average reads per amplicon) in most of the isolates except 4Apr20-64-Hu (Table 3). To investigate the potential contribution of genome variations, we analyzed the two low covered regions and the primer binding sites for all six isolates. Since the primer sequences were not made available, we assumed the 30 bp upstream and downstream of the amplicons as the potential primer binding sites. Based on these assumptions, it was revealed that 4Apr20-64-Hu possessed C at positions 13 730 (21bp upstream of amplicon r1_1.14.786182) and 23 929 (12bp upstream of amplicon r1_1.25.388943), while other isolates possessed T at these positions (Figure 1), suggesting the presence of

Table 1. Genome coverage assembly analysis

Sample Name	Generated Reads	Mapping on SARS-CoV-2 and human expression control				Mapping on SARS-CoV-2			
		Mapped Reads	On Target	Mean Base Depth (x)	Base Coverage Uniformity	Mapped Reads	On Target	Mean Base Depth (x)	Base Coverage Uniformity
5Apr20-64-Hu	216151	216056	99.99%	1436	97.46%	216118	100%	1458	98.85%
21Apr20-130-Hu	1513292	1512025	99.99%	10002	96.03%	1512749	100%	10154	97.40%
21Apr20-211-Hu	771351	770852	99.99%	5046	97.20%	771200	100%	5123	98.64%
21Apr20-101-Hu	1202333	1201474	99.99%	7997	97.58%	1202067	100%	8119	99.03%
21Apr20-209-Hu	179227	179148	99.99%	1180	95.94%	179214	100%	1198	97.31%
2May20-132-Hu	1168738	1167742	99.99%	7728	97.14%	1168100	100%	7845	98.62%

Table 2. Amplicon targets with low read coverage

Isolate	Uniformity of amplicon coverage (%)	Average reads per amplicon	20% of the average reads per amplicon		5% of the average reads per amplicon	
			Number of reads per amplicon	Number of targets with less than 20% of the average reads per amplicon	Number of reads per amplicon	Number of targets with less than 5% of the average reads per amplicon
5Apr20-64-Hu	99.16	911.9	182.32	2	45.60	0
21Apr20-130-Hu	97.05	6383	1276.60	7	319.15	2
21Apr20-211-Hu	98.73	3254	650.8	3	162.70	2
21Apr20-101-Hu	99.16	5072	1014.4	2	253.6	1
21Apr20-209-Hu	97.47	756.2	151.24	6	37.81	2
2May20-132-Hu	97.89	4929	985.8	5	246.45	3

Table 3. Amplicon targets with low read coverage

Target	Target location (referring to Wuhan-Hu-1, MN908947)	Number of reads mapped to amplicon target					
		5Apr20-64-Hu	21Apr20-130-Hu	21Apr20-211-Hu	21Apr20-101-Hu	21Apr20-209-Hu	2May20-132-Hu
r1_1.2.1238102	839-1060	607	7990	2714	5139	682	537
r1_1.5.1289446	4019-4239	678	1249	1322	3025	263	2023
r1_1.8.440758	6883-7099	855	1043	1285	3631	348	1971
r1_1.11.528369	10453-10679	163	2161	747	1151	117	1301
r1_1.12.1126179	10993-11218	959	679	1421	3445	359	1827
r1_1.14.786182	13751-13965	587	40	104	626	24	101
r1_1.15.464755	14252-14472	435	871	1054	2876	183	1622
r1_1.15.1421280	14410-14550	114	10586	6223	8658	1531	8949
r1_1.21.1100292	20649-20875	555	3646	1832	3159	308	73
r1_1.23.179592	22623-22834	324	405	497	1790	124	852
r1_1.25.388943	23941-24106	599	27	30	20	26	28
r1_1.27.793950	26389-26605	240	1882	1127	1489	146	1508
r1_1.30.329795	28995-29206	274	2028	837	2627	148	1541

Red – number of reads with less 20% of the average reads per amplicon.

Blue – number of reads with less 5% of the average reads per amplicon.

these nucleotide variations within the potential primer binding sites.

Genome assemblies were completed using the IRMAreport plugin, resulting in five complete and one near-complete genome consensus sequences covering at least from positions 26 to 29 860. A 16-nucleotide gap at position 23 993 to 24 008 (comparing against Wuhan-Hu-1, MN908947)

was detected in 21Apr20-130-Hu. This gap fell within the amplicon target region of r1_1.25.388943 (position 23 941-24 106, Table 3). The 16-bp sequencing gap in 21Apr20-130-Hu was filled by Sanger sequencing with primers; (Primer F: 5'-CAACTTACTCCTACTTGGCGTGT-3' and Primer R: 5'-TGTGTACAAAACCTGCCATATTGCA-3') as previously published (Quick, 2020).

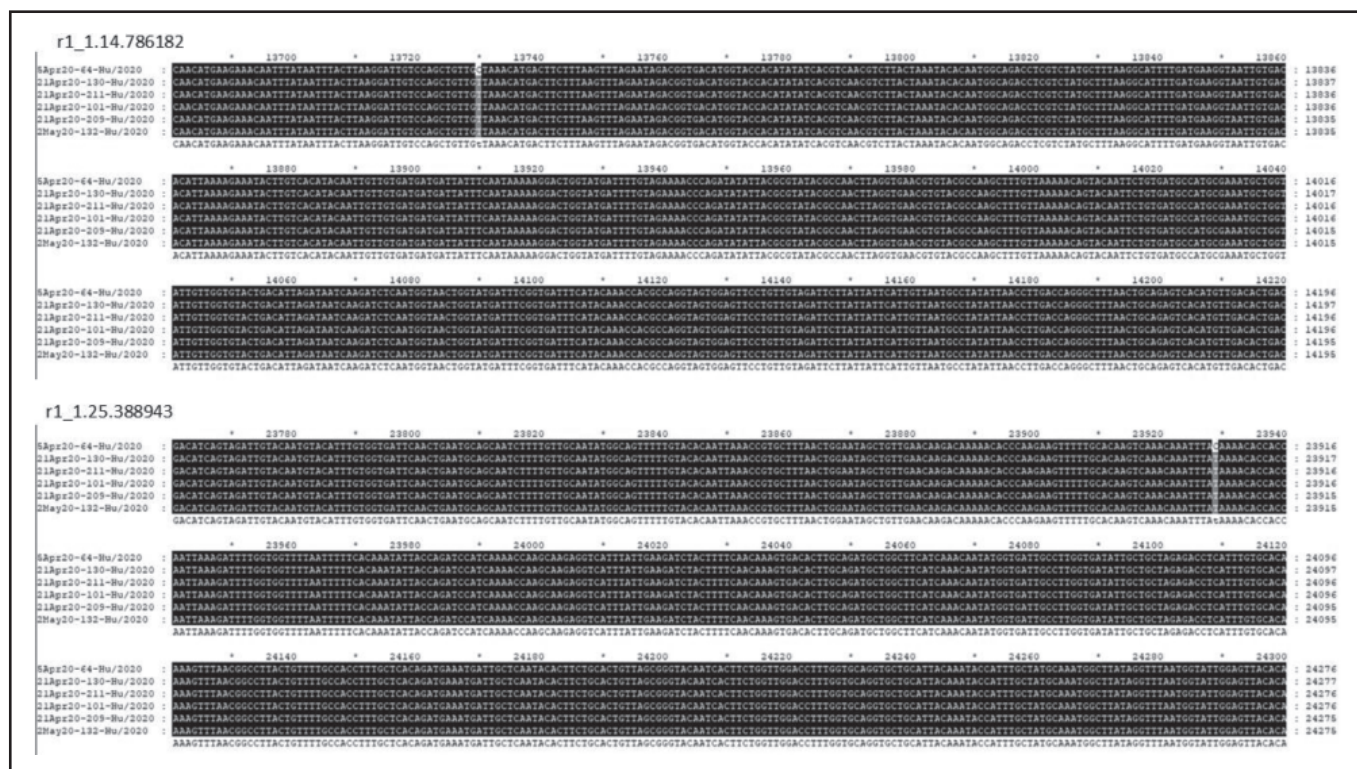


Figure 1. Genomic sequence variations among the SARS-CoV-2 isolates for amplicon r1_1.14.786182 and r1_1.25.388943.

DISCUSSION

The number of SARS-CoV-2 genome sequences in the public database is growing rapidly (Shu & McCauley, 2017). The exponential growth of the publicly available SARS-CoV-2 genome sequences is attributable to the rapid genome sequencing, development of data analysis workflow, and data sharing by researchers worldwide (Campos *et al.*, 2020; Pillay *et al.*, 2020; Quick, 2020). Currently, most of the sequencing workflows are created for the use of the Oxford Nanopore and Illumina’s sequencing platforms (Colson *et al.*, 2020; Nasir *et al.*, 2020; Resende *et al.*, 2020). The ARTIC protocol implemented for the Oxford Nanopore platform, one of the most widely used sequencing protocols (Quick, 2020). In some cases, researchers used both the Oxford Nanopore and Illumina sequencing platforms to generate consensus genome sequences (Colson *et al.*, 2020). The Ion Torrent sequencing platform, one of the previously popular platforms used for viral genome sequencing (Tan *et al.*, 2015, 2018; Marine *et al.*, 2020), however, was not widely used in SARS-CoV-2 sequencing. Several Ion-Torrent based SARS-CoV-2 sequencing workflows were reported (Alessandrini *et al.*, 2020; Campos *et al.*, 2020; Capobianchi *et al.*, 2020) but not very popular among the researchers. Unlike the ARTIC protocol, the published protocol for some of these in-house Ion-Torrent based assays were not detailed enough for it to be replicated in other laboratory settings (Campos *et al.*, 2020; Capobianchi *et al.*, 2020). The recently launched Ion AmpliSeq™ SARS CoV 2 Research Panel user guide contained the detailed and optimized sequencing protocol for the Ion S5 sequencing platform (MAN0019277 Rev.A.O). In the current study, we adopted and modified the Ion AmpliSeq™ SARS CoV 2 Research Panel and applied it to the Ion PGM sequencing platform. When adopting or establishing a new protocol, it is critical to harmonize the steps written in the user manual with the existing Standard Operating Procedures

(SOP) in a laboratory. This is to ensure that we obtain high-quality data and results. If a different platform or different reagents from the user manual are used, the specific SOP should be carefully optimized and accessed to maximize the output and data reproducibility of this newly established protocol. In a single sequencing run with the Ion 318™ Chip Kit v2 BC, five complete and one near-complete genome sequences of SARS-CoV-2 derived from RNA samples directly extracted from human nasopharyngeal swabs were generated. According to the manufacturer’s protocol, 1 million sequencing reads were recommended for every sample (MAN0019277 Rev.A.O). Our sequencing run, however, generated approximately 180 000 to 1 500 000 reads per sample, suggesting a lesser number of total reads were sufficient to generate a complete genome sequence. Hence, the number of multiplexed samples can be increased to reduce the sequencing cost using the Ion PGM.

For NGS, an ideal protocol should generate results with high on-target specificity and read coverage uniformity. In this study, 99.9% of the generated reads of all six isolates were mapped to SARS-CoV-2 genomes with more than 96% coverage uniformity. Uneven read distribution is a common issue, and intrinsic factors affecting data quality of NGS (Maurier *et al.*, 2019). In fact, uneven reads distribution was also reported in the previous version of SARS-CoV-2 tiling PCR amplification method (Quick, 2020). Two target regions (r1_1.14.786182 and r1_1.25.388943) of the Ion AmpliSeq™ SARS CoV 2 Research Panel consistently resulted in low sequencing read coverage in most samples isolated from human and infected cell culture supernatant (unpublished data). Clearly, this problem cannot be random or sample type dependent. Generally, increased amount of sequencing output is the easiest way to improve the coverage in the low read depth region. The 4Apr20-64-Hu with sequencing reads of 216 151 had good sequence coverage (>500X) for both regions. Neither the sample with lower sequencing reads

(21Apr20-209-Hu) nor samples with a higher number of sequencing reads demonstrated good coverage at these two regions. The high amount of sequencing reads will only lead to the over-sequencing of the adequately covered regions, causing higher sequencing costs. Therefore, increasing the overall number of sequencing reads will not be suitable for solving the read depth problem for r1_1.14.786182 and r1_1.25.388943. Other factors, such as genome sequence variation and variability of GC content, are common factors that could affect the efficiency of the target enrichment process. We observed that the sequence nucleotide variations between 4Apr20-64-Hu (>500X read coverage at both regions) and other isolates at positions 13 730 and 23 929 at the potential primer binding sites for target regions, respectively. Hence, low coverage regions reported herein could be genome sequence dependent, and the specific genome sequence variations could lead to inefficient primer annealing during the multiplex amplification process. An improved version of the Ion AmpliSeq™ SARS CoV 2 Research Panel tackling these two low coverage regions or other genome sequence variations present in the currently circulating SARS-CoV-2 strains should be considered.

Taken together, we report a rapid complete genome sequencing protocol for the SARS-CoV-2 to be used with the Ion PGM. The much lesser amount of sequencing reads than the recommended 1 million reads was sufficient to produce a complete SARS-CoV-2 genome sequence. Six samples or more can be included in a single sequencing run using the Ion 318™ Chip Kit v2 BC. Our findings nonetheless revealed that using the Ion AmpliSeq™ SARS CoV 2 Research Panel, two potential dropout regions would occur, and increasing of the sequencing reads would not be useful. An improved version of the Ion AmpliSeq™ SARS CoV 2 Research Panel that addresses the potential genome nucleotide variations at the primer binding sites could improve the read coverage uniformity and usability of the sequencing data.

Data availability

The generated genome sequences of SARS-CoV-2 were deposited into the European Nucleotide Archive (ENA) database with study number PRJEB40971.

Abbreviations

COVID-19, coronavirus disease 2019; ISPs, Ion SphereParticles; MOHE, Ministry of Higher Education; MOSTI, Ministry of Science, Technology, and Innovation; NGS, Next-generation sequencing; PGM, Personal Genome Machine; rRT-PCR, real-time reverse-transcription polymerase chain reaction; SARS-CoV-2, Severe acute respiratory syndrome coronavirus 2; TIDREC, Tropical Infectious Diseases Research and Education Centre; UM, Universiti Malaya.

ACKNOWLEDGMENTS

We are part of the COVID-19 research and laboratory testing team at the Tropical Infectious Diseases Research and Education Centre (TIDREC), Universiti Malaya (UM), and WHO Collaborating Centre for Arbovirus Reference and Research (MAA-12 [WPRO]). We are grateful to all the staffs from TIDREC who were involved in the COVID-19 testing, members, and staffs from the Malaysia Ministry of Science, Technology, and Innovation (MOSTI)-Ministry of Higher Education (MOHE) COVID-19 testing initiative and the Malaysia Ministry of Health who were involved in the sample collection and logistics. The manuscript had been deposited as preprint at Research Square (doi: 10.21203/rs.3.rs-141611/v1).

This study was supported in parts by the Ministry of Science, Technology and Innovation, Malaysia (www.mosti.gov.my), the special funding for COVID-19 testing (UM.0000345/KWJ.AK), Ministry of Higher Education, Malaysia (www.mohe.gov.my), the funding for niche area research under the Higher Institution Centre of Excellence (HICoE) program (MO002-2019) and the funding under Fundamental Research Grant Scheme: FRGS-MRSA/1/2018/SKK08/UM/01/1 (MO012-2017), and University of Malaya (www.um.edu.my; University Malaya RU grant: RU002-2019). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of interest

The author declares that they have no conflict of interests.

REFERENCES

- Alessandrini, F., Caucci, S., Onofri, V., Melchionda, F., Tagliabracci, A., Bagnarelli, P., Di Sante, L., Turchi, C. & Menzo, S. (2020). Evaluation of the Ion AmpliSeq SARS-CoV-2 research panel by massive parallel sequencing. *Genes* **11**: 929. <https://doi.org/10.3390/genes11080929>
- Campos, G.S., Sardi, S.I., Falcao, M.B., Belitardo, E., Rocha, D., Rolo, C.A., Menezes, A.D., Pinheiro, C.S., Carvalho, R.H. & Almeida, J.P.P. (2020). Ion torrent-based nasopharyngeal swab metatranscriptomics in COVID-19. *Journal of Virological Methods* **282**: 113888. <https://doi.org/10.1016/j.jviromet.2020.113888>
- Capobianchi, M.R., Rueca, M., Messina, F., Giombini, E., Carletti, F., Colavita, F., Castilletti, C., Lalle, E., Bordi, L. & Vairo, F. (2020). Molecular characterization of SARS-CoV-2 from the first case of COVID-19 in Italy. *Clinical Microbiology and Infection* **26**: 954-956. <https://doi.org/10.1016/j.cmi.2020.03.025>
- Colson, P., Lagier, J.C., Baudoin, J.P., Bou Khalil, J., La Scola, B. & Raoult, D. (2020). Ultrarapid diagnosis, microscope imaging, genome sequencing, and culture isolation of SARS-CoV-2. *European Journal of Clinical Microbiology & Infectious Disease* **39**: 1601-1603. <https://doi.org/10.1007/s10096-020-03869-w>
- Fischer, N., Indenbirken, D., Meyer, T., Lutgehetmann, M., Lellek, H., Spohn, M., Aepfelbacher, M., Alawi, M. & Grundhoff, A. (2015). Evaluation of unbiased next-generation sequencing of RNA (RNA-seq) as a diagnostic method in influenza virus-positive respiratory samples. *Journal of Clinical Microbiology* **53**: 2238-2250. <https://doi.org/10.1128/JCM.02495-14>
- Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K.S.M., Lau, E.H.Y. & Wong, J.Y. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *New England Journal of Medicine* **382**: 1199-1207. <https://doi.org/10.1056/NEJMoa2001316>
- Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B. & Zhu, N. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet* **395**: 565-574. [https://doi.org/10.1016/S0140-6736\(20\)30251-8](https://doi.org/10.1016/S0140-6736(20)30251-8)
- Marcelino, V.R., Irinyi, L., Eden, J.-S., Meyer, W., Holmes, E.C. & Sorrell, T.C. (2019). Metatranscriptomics as a tool to identify fungal species and subspecies in mixed communities – a proof of concept under laboratory conditions. *IMA Fungus* **10**. <https://doi.org/10.1186/s43008-019-0012-8>

- Marine, R.L., Magana, L.C., Castro, C.J., Zhao, K., Montmayeur, A.M., Schmidt, A., Diez-Valcarce, M., Ng, T.F.F., Vinje, J. & Burns, C.C. (2020). Comparison of Illumina MiSeq and the Ion Torrent PGM and S5 platforms for whole-genome sequencing of picornaviruses and caliciviruses. *Journal of Virological Methods* **280**: 113865. <https://doi.org/10.1016/j.jviromet.2020.113865>
- Maurier, F., Beury, D., Flechon, L., Varre, J.S., Touzet, H., Goffard, A., Hot, D. & Caboche, S. (2019). A complete protocol for whole-genome sequencing of virus from clinical samples: Application to coronavirus OC43. *Virology* **531**: 141-148. <https://doi.org/10.1016/j.virol.2019.03.006>
- Nasir, J.A., Kozak, R.A., Aftanas, P., Raphenya, A.R., Smith, K.M., Maguire, F., Maan, H., Alruwaili, M., Banerjee, A. & Mbareche, H. (2020). A comparison of whole genome sequencing of SARS-CoV-2 using amplicon-based sequencing, random hexamers, and bait capture. *Viruses* **12**: 895. <https://doi.org/10.3390/v12080895>
- Pillay, S., Giandhari, J., Tegally, H., Wilkinson, E., Chimukangara, B., Lessells, R., Mattison, S., Moosa, Y., Gazy, I. & Fish, M. (2020). Whole genome sequencing of SARS-CoV-2: Adapting Illumina protocols for quick and accurate outbreak investigation during a pandemic. *Genes* **11**: 949. <https://doi.org/10.3390/genes11080949>
- Quick, J. (2020). nCoV-2019 sequencing protocol. Protocols. <https://www.protocols.io/view/ncov-2019-sequencing-protocol-v3-locost-bh42j8ye>. Accessed 20 June 2021.
- Resende, P.C., Motta, F.C., Roy, S., Appolinario, L., Fabri, A., Xavier, J., Harris, K., Matos, A.R., Caetano, B. & Orgeswalska, M. (2020). SARS-CoV-2 genomes recovered by long amplicon tiling multiplex approach using nanopore sequencing and applicable to other sequencing platforms. *BioRxiv* 2020.04.30.069039 [Preprint]. <http://doi.org/10.1101/2020.04.30.069039>
- Shu, Y. & McCauley, J. (2017). GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance* **22**: 30494. <http://dx.doi.org/10.2807/1560-7917.ES.2017.22.13.30494>
- Tan, K.K., Sy, A.K., Tandoc, A.O., Khoo, J.J., Sulaiman, S., Chang, L.Y. & AbuBakar, S. (2015). Independent emergence of the Cosmopolitan Asian Chikungunya Virus, Philippines 2012. *Scientific Reports* **5**: 12279. <https://doi.org/10.1038/srep12279>
- Tan, K.K., Zulkifle, N.I., Sulaiman, S., Pang, S.P., NorAmdan, N., MatRahim, N., Abd-Jamil, J., Shu, M.H., Mahadi, N.M. & AbuBakar, S. (2018). Emergence of the Asian lineage dengue virus type 3 genotype III in Malaysia. *BMC Evolutionary Biology* **18**: 58. <https://doi.org/10.1186/s12862-018-1175-4>
- Wang, C., Horby, P.W., Hayden, F.G. & Gao, G.F. (2020). A novel coronavirus outbreak of global health concern. *The Lancet* **395**: 470-473. [https://doi.org/10.1016/S0140-6736\(20\)30185-9](https://doi.org/10.1016/S0140-6736(20)30185-9)
- World Health Organization. (2020a). Mission summary: WHO Field Visit to Wuhan, China 20-21 January 2020. World Health Organization. <https://www.who.int/china/news/detail/22-01-2020-field-visit-wuhan-china-jan-2020>. Accessed 20 June 2021
- World Health Organization. (2020b). WHO Coronavirus Disease (COVID-19) Dashboard. World Health Organization. <https://covid19.who.int/table>. Accessed 20 June 2021.
- Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z.W., Tian, J.H. & Pei, Y.Y. (2020). A new coronavirus associated with human respiratory disease in China. *Nature* **579**: 265-269. <https://doi.org/10.1038/s41586-020-2008-3>
- Xiao, M., Liu, X., Ji, J., Li, M., Li, J., Yang, L., Sun, W., Ren, P., Yang, G. & Zhao, J. (2020). Multiple approaches for massively parallel sequencing of SARS-CoV-2 genomes directly from clinical samples. *Genome Medicine* **12**: 57. <https://doi.org/10.1186/s13073-020-00751-4>
- Zhang, T., Wu, Q. & Zhang, Z. (2020). Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak. *Current Biology* **30**: 1578. <https://doi.org/10.1016/j.cub.2020.03.063>
- Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B. & Huang, C.L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**: 270-273. <https://doi.org/10.1038/s41586-020-2012-7>