



Rapid Genomic Characterization of SARS-CoV-2 by Direct Amplicon-Based Sequencing Through Comparison of MinION and Illumina iSeq100TM System

Véronique Hourdel, Aurelia Kwasiborski, Charlotte Balière, Séverine Matheus, Christophe Frédéric Batéjat, Jean-Claude Manuguerra, Jessica Vanhomwegen, Valérie Caro

► To cite this version:

Véronique Hourdel, Aurelia Kwasiborski, Charlotte Balière, Séverine Matheus, Christophe Frédéric Batéjat, et al.. Rapid Genomic Characterization of SARS-CoV-2 by Direct Amplicon-Based Sequencing Through Comparison of MinION and Illumina iSeq100TM System. *Frontiers in Microbiology*, 2020, 11, pp.571328. 10.3389/fmicb.2020.571328 . pasteur-02976015

HAL Id: pasteur-02976015

<https://pasteur.hal.science/pasteur-02976015>

Submitted on 23 Oct 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Rapid Genomic Characterization of SARS-CoV-2 by Direct Amplicon-Based Sequencing Through Comparison of MinION and Illumina iSeq100™ System

OPEN ACCESS

Edited by:

Jason Kindrachuk,
University of Manitoba, Canada

Reviewed by:

Rachele Cagliani,
Eugenio Medea (IRCCS), Italy
Tung Phan,
University of Pittsburgh Medical
Center, United States
Rachel Graham,
University of North Carolina at Chapel
Hill, United States
Wenjie Tan,
National Institute for Viral Disease
Control and Prevention (China CDC),
China

*Correspondence:

Valérie Caro
valerie.caro@pasteur.fr

†These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Virology,
a section of the journal
Frontiers in Microbiology

Received: 18 June 2020

Accepted: 02 September 2020

Published: 25 September 2020

Citation:

Hourdel V, Kwasiborski A,
Balière C, Matheus S, Batéjat CF,
Manuguerra J-C, Vanhomwegen J
and Caro V (2020) Rapid Genomic
Characterization of SARS-CoV-2 by
Direct Amplicon-Based Sequencing
Through Comparison of MinION
and Illumina iSeq100™ System.
Front. Microbiol. 11:571328.
doi: 10.3389/fmicb.2020.571328

Véronique Hourdel[†], Aurelia Kwasiborski[†], Charlotte Balière[†], Séverine Matheus,
Christophe Frédéric Batéjat, Jean-Claude Manuguerra, Jessica Vanhomwegen and
Valérie Caro*

Environment and Infectious Risks Unit, Institut Pasteur, Paris, France

Global human health is increasingly challenged by emerging viral threats, especially those observed over the last 20 years with coronavirus-related human diseases, such as the Severe Acute Respiratory Syndrome (SARS) and the Middle East Respiratory Syndrome (MERS). Recently, in late December 2019, a novel *Betacoronavirus*, SARS-CoV-2, originating from the Chinese city of Wuhan, emerged and was then identified as the causative agent of a new severe form of pneumonia, COVID-19. Real-time genome sequencing in such viral outbreaks is a key issue to confirm identification and characterization of the involved pathogen and to help establish public health measures. Here, we implemented an amplicon-based sequencing approach combined with easily deployable next-generation sequencers, the small and hand-held MinION sequencer and the latest most compact Illumina sequencer, the iSeq100™ system. Our results highlighted the great potential of the amplicon-based approach to obtain consensus genomes of SARS-CoV-2 from clinical samples in just a few hours. Both these mobile next-generation sequencers are proven to be efficient to obtain viral sequences and easy to implement, with a minimal laboratory environment requirement, providing useful opportunities in the field and in remote areas.

Keywords: COVID-19, real-time genome sequencing, amplicon-based sequencing, iSeq100™ system, MinION, SARS-CoV-2

INTRODUCTION

Global human health is increasingly challenged by emerging viral threats, especially those observed over the last 20 years with the Severe Acute Respiratory Syndrome (SARS) and the Middle East Respiratory Syndrome (MERS) caused by coronaviruses. Recently, in late December 2019, a novel *Betacoronavirus*, SARS-CoV-2, originating from Wuhan, Hubei province, China, emerged and was identified as the causative agent of a new form of severe pneumonia, named coronavirus disease 2019 (COVID-19) (Wu et al., 2020; Zhou et al., 2020). The outbreak spread further to Europe and America mid-March 2020 to become a global public health emergency, with the vast majority of countries and territories reporting confirmed cases. By June 2, 2020, the number of

confirmed cases in the world has increased to over 6,306,746 (Dong et al., 2020). SARS-CoV-2, an enveloped positive-sense (+) single-stranded (ss) RNA virus, is a new member of the *Coronaviridae* family, *Betacoronavirus* genus, and the seventh coronavirus known to infect humans. Genetically distinct from SARS-CoV and MERS-CoV, SARS-CoV-2 was included in the monophyletic clade 2, within subgenus *Sarbecovirus*, due to its closest sequence similarity with Chinese Bat SARS-like coronaviruses (Lu et al., 2020; Wu et al., 2020; Ye et al., 2020). Despite its RNA proofreading capacity during replication like all members of the *Coronaviridae* family, SARS-CoV-2 could nevertheless acquire some nucleotide mutations along the genome, allowing for tracking its spread and thus clade definition.

The implementation of short turnaround time and user-friendly sequencing tools in affected regions to track the spread, analyze variants, and identify clusters of transmission of emerging viruses, like SARS-CoV-2, is becoming essential. Although whole genome sequencing of viral pathogens has become a routine procedure in epidemiological monitoring and surveillance, these molecular investigations essential for public health support are still challenging in remote areas with poor technical resources. Thus, the development of easily deployable viral genome toolkits can be proposed to rapidly obtain SARS-CoV-2 genomic sequence data in the field without the requirement to send biological samples to suitably equipped and reference laboratories. These molecular investigations also depend on efficient and direct sequencing of viral material from clinical samples, i.e., often with low viral genetic loads and a high concentration of contaminant host nucleic acid background. To circumvent this difficulty, different techniques of enrichment can be achieved, either directly through laborious and time-consuming culture isolation or, even better, the use of specific primers targeting the whole genome of the virus (Houldcroft et al., 2017). Moreover, specific amplification-based approaches are proven successful since they can generate sufficient quantities of the genetic material needed for next-generation sequencing (Quick et al., 2017).

Recently, mobile next-generation sequencers have provided new opportunities in infectious diseases diagnostics and surveillance, such as the rapid sequencing of viral genomes during outbreaks. The pocket size nanopore technology sequencer MinION (Oxford Nanopore Technologies, ONT) is portable and field-deployable and enables real-time outbreak surveillance of threatening pathogens, such as Ebola, Lassa, and Zika viruses (Faria et al., 2016; Hoenen et al., 2016; Kafetzopoulou et al., 2019). Similarly, the recent Illumina iSeq100™ system is the most compact, accessible, and affordable next-generation Illumina sequencer. This easy-to-use system is ideal for small whole-genome sequencing, e.g., viruses, and specific genomic targeted approaches (Colman et al., 2019).

In this study, we aimed at implementing an amplicon-based sequencing approach to obtain SARS-CoV-2 consensus genomes directly from clinical specimens, adaptable into the field conditions, with the two easily manageable next-generation sequencers, the nanopore MinION and the Illumina iSeq100™ system. We also evaluated the performance of these two sequencers, in terms of ease of use, hands-on time,

simplification of the sequencing process, and capacity for real-time genome sequencing in settings lacking laboratory resources for outbreak response.

MATERIALS AND METHODS

Clinical Samples

The two samples used in this study were sputum (CIBU-200107 and CIBU-200132) collected from French suspected COVID-19 patients and addressed to the Laboratory for Urgent Response to Biological Threats for diagnostic testing. Sputum samples were handled in an enhanced biosafety level 2 laboratory, with full personal protective equipment (including FFP3 respirators) and procedures adapted to airborne pathogens, by trained staff, as recommended by WHO [Laboratory biosafety guidance related to coronavirus disease (COVID-19), interim guidance, 19 March 2020].

Virus Culture

Vero E6 cells (mycoplasma-free) seeded in 24-well plates (4×10^5 cells/well) were cultured in DMEM (Dulbecco's modified Eagle medium; Thermo Fisher Scientific, United States) containing 1% PS (penicillin 10,000 U/ml; streptomycin 10,000 µg/ml) and supplemented with 5% FBS (fetal bovine serum). Samples were diluted 1:10 in DMEM 1% PS without FBS and supplemented with 1 µg/ml TPCK-trypsin (Sigma-Aldrich, United States), added to the cell monolayers and incubated for 1 h at 37°C in the presence of 5% CO₂. The inoculum was then removed and replaced with fresh DMEM 1% PS containing 1 µg/ml TPCK-trypsin. After incubation for 3 days at 37°C in the presence of 5% CO₂, cells were observed for the presence of a cytopathic effect (CPE) under the microscope, and culture supernatants were harvested, aliquoted, and stored at -80°C.

Extraction of Genomic RNA From Specimens

Total RNA from clinical samples was extracted using the NucleoMag kit on the KingFisher automate (Macherey Nagel, Germany) and from isolates using the NucleoSpin DX Virus (Macherey Nagel, Germany), following the manufacturer's instructions.

Genome Detection by Real-Time RT-PCR

Sputum specimens and respective isolates RNA extracts were tested with the SARS-CoV-2 real-time *RdRP* gene duplex reverse transcription (RT)-PCR developed by the French National Reference Center for Respiratory Viruses and the real-time *E* gene RT-PCR from the Charité protocol (see WHO Coronavirus disease COVID-19 technical guidance: Laboratory testing for 2019-nCoV in humans, available from <https://www.who.int/docs/default-source/coronaviruse/whoinhouseassays.pdf>). Samples were considered SARS-CoV-2 positive if at least two out of three SARS-CoV-2 gene targets were detected by the RT-PCR assays.

Amplicon-Based Protocol

For each sample, the ProtoScript II First Strand cDNA synthesis kit (New England Biolabs, NEB, United States) was used to obtain single-strand cDNA from RNA extracts using random hexamers. RT was performed at 48°C for 15 min and 80°C for 5 min. Pools of specific primer sets were used to generate 36 amplicons from the cDNA using the Q5 Hot Start High-Fidelity DNA Polymerase (NEB, United States) (**Supplementary Table**). These multiplex primer sets (divided into two separate pools) were designed using the Primal scheme software¹ on the SARS-CoV-2 reference sequence (Genbank accession number NC_045512) to sequentially amplify 1,000 bp using the following PCR conditions: 98°C for 30 s, 40 cycles of 98°C for 15 s, 65°C for 5 min, and ended by 65°C for 5 min. Clean-up of PCR products was performed with AMPure XP magnetic beads (Beckman Coulter, United States). The quantity of amplicons was measured with the Qubit 2.0 fluorometer using the dsDNA HS Assay Kit (Thermo Fisher Scientific, United States), and the quality was assessed by electrophoresis in E-Gel® EX 1% agarose (Invitrogen®) and visualized with E-Gel PowerSnap Electrophoresis System (Invitrogen®). The amplicons obtained from the two clinical samples and their corresponding isolates were normalized to equal concentrations, multiplexed (X2), and sequenced using the MinION and Illumina iSeq100™ System (**Figure 1**).

¹<http://primal.zibraproject.org/>

MinION Sequencing

Library preparation for the MinION sequencing was performed using the Ligation Sequencing kit SQK-LSK109 and Natives Barcoding kits EXP-NBD104/EXP-NBD114 (ONT) according to the manufacturer's instructions and modifications as per Quick et al. (2017). Briefly, PCR amplicons pools were end-repaired and dA-tailed using an UltraII End Prep Reaction Module (NEB, United States) followed by ligation of native barcodes using the NEBNext UltraII Ligation module (NEB, United States). All the pooled barcoded libraries were purified using AMPure XP beads (Beckman Coulter) followed by adapter ligation with the NEBNext UltraII Ligation module. Library clean-up was performed using AMPure XP beads and short fragment buffer (SFB) and then eluted in 15 µl of ONT's elution buffer. The library was loaded onto an R9.4 flow cell (FLO-MIN106) and sequenced on a MinION Mk1B device within 4 h (**Figure 1**). ONT MinKNOW software (version 19.12.2) was used to collect raw data and perform basecalling (Guppy v3.4). The RAMPART tool (Read Assignment, Mapping, and Phylogenetic Analysis in Real Time) developed by the ARTIC network has been used to visualize genome coverage in real time and reference matching for each barcode.

MinION Bioinformatics Workflow

From the fastq files, only reads with a minimum Q score of 7 were selected for the analysis. Data were demultiplexed using the guppy_barcode with the option require_barcode_both_ends to

SARS-CoV-2 sequencing workflow

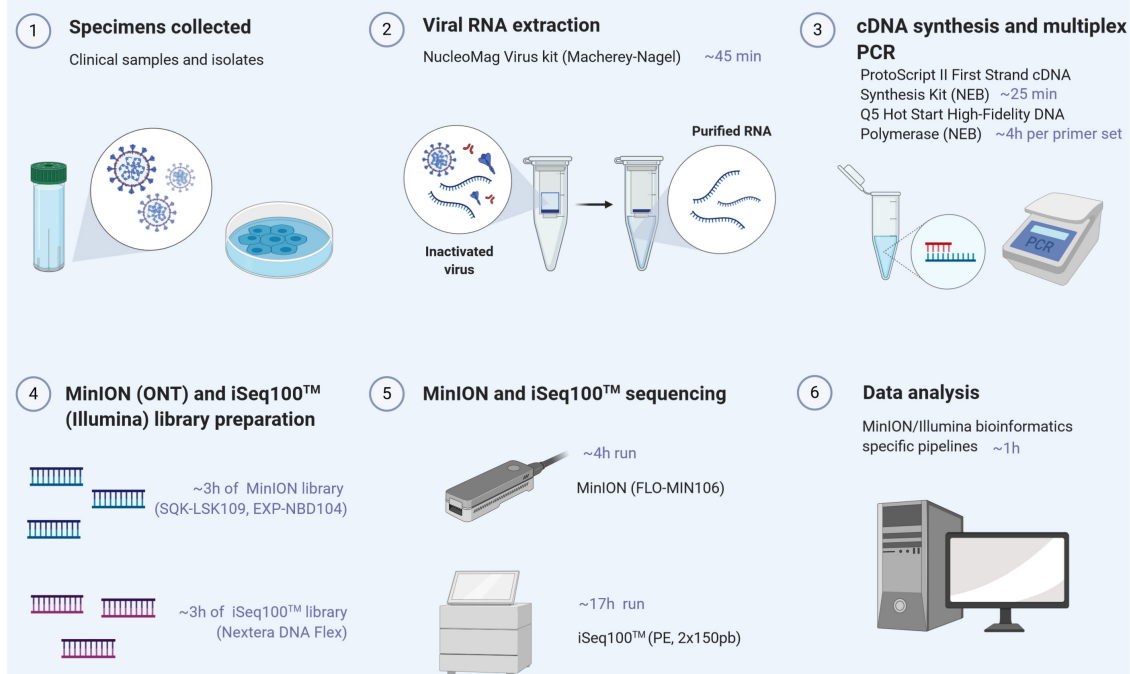


FIGURE 1 | SARS-CoV-2 next-generation sequencing workflow, from the sample to sequence analysis. Created with biorender.com.

ensure barcodes are present at each end of the fragment. The reads were filtered on the expected length of amplicons. Reads between 900 and 1,200 base pairs were kept; thus, potential chimeric reads were removed. Selected reads were mapped against SARS-CoV-2 reference (NC_045512) using Minimap2 (v2.9) (Li, 2018). SAMtools (v1.9) were used to sort the aligned BAM files, to obtain coverage data and a consensus sequence. Alignment statistics were also calculated with SAMtools. Data were manually inspected using Tablet (v1.19) (Milne et al., 2013).

Illumina iSeq100™ Sequencing

The appropriate volume of each pool of amplicons was adjusted to have a total quantity of 100 ng of amplicon input (50 ng of each) per sample. The sequencing-ready libraries were prepared using the Nextera DNA Flex Prep kit (Illumina, United States). The libraries were qualified on an Agilent Technologies 2100 Bioanalyzer using a high-sensitivity DNA chip following the manufacturer's instructions and quantified with the Qubit 2.0 fluorometer using the dsDNA HS Assay Kit (Thermo Fisher Scientific, United States). The resulting libraries were sequenced using the iSeq100™ system (Illumina, United States) and in two multiplexed runs were performed generating 2×151 bp read length data during a 17 h run time (Figure 1).

Illumina Bioinformatics Workflow

To remove low quality reads, trim off low-quality and contaminant residues, and filter out duplicated reads, fqCleaner v.0.5.0 was used, with Phred quality score of 28. Filtered reads were mapped against SARS-CoV-2 reference (NC_045512) using Burrows-Wheeler Aligner MEM algorithm (BWA-MEM) (v0.7.7). SAMtools (v1.9) was used to sort BAM files and to generate alignment statistics, coverage data, and a consensus sequence. Data were manually inspected using Tablet (v1.19) (Milne et al., 2013).

RESULTS

Specimens' Diagnostics

The two sputum samples CIBU-200107 and CIBU-200132, were tested with the specific SARS-CoV-2 real-time RT-PCR protocol. They were both found positive with the three SARS-CoV-2 gene targets. The threshold cycle (Ct) values for CIBU-200107 and CIBU-200132 were respectively 24-24 and 12-12 for the *RdRP* gene duplex and 25 and 13 for the *E* gene. The two clinical specimens were also inoculated on Vero E6 cells, and after incubation for 3 days, a clear cytopathic effect was observed. Culture supernatants of the first passage were harvested and tested for the presence of the virus with the specific SARS-CoV-2 real-time RT-PCR protocol. The Ct values of the two corresponding isolates CIBU-200107C1 and CIBU-200132C1 were respectively 21-21 and 10-10 for the *RdRP* gene duplex and 21 and 10 for the *E* gene, confirming the isolation of SARS-CoV-2.

Amplicon-Based Protocol

As soon as the first SARS-CoV-2 reference genome was released in Genbank (NC_045512), we designed two sets of primers to

generate a tiling path along the genome using the Primal Scheme tool. The full-length genome was amplified directly from the RNA extracts from the clinical samples and the corresponding isolates, to obtain PCR products overlapping by 120 bp and covering the entire 29,903 bp viral genomic sequence. Thirty-six amplicons, split into two pools, were successfully generated for each specimen with a size expected range between 900 and 1,098 bp (with an average length of 972 bp) and controlled by gel electrophoresis.

MinION Sequencing

The amplicons obtained from the two clinical samples CIBU-200107 and CIBU-200132 were multiplexed and sequenced using the MinION platform with R9.4.1 flow cell during 4 h. The MinION run produced 45,625 and 122,113 raw reads with a good quality score, respectively, for the two clinical samples, with an average of 1,074 bp read length. Thus, the mapping against Wuhan SARS-CoV-2 reference (NC_045512), by using Minimap2, retrieved 43,952 (96.3%) and 122,108 (99.9%) reads, respectively, for the two clinical samples (Table 1). The run yielded a median read depth of 436X and 3,418X for the two clinical samples, respectively (Table 1 and Figure 2), with the average base error rate ranging between 7.9 and 8.3%. A near full-length consensus sequence of 26,852 and 28,391 bp was generated for CIBU-200107 and CIBU-200132, respectively. In the same way, a 4 h MinION run, applied to the two corresponding isolates CIBU-200107C1 and CIBU-200132C1, produced 107,532 and 120,958 high-quality raw reads, respectively, with an average of 1,022 bp read length (Table 1). The run yielded a median read depth of 2,979X and 3,607X for the two isolates, respectively (Table 1 and Figure 2). The mapping against Wuhan SARS-CoV-2 reference (NC_045512) using Minimap2 used 99.9% of raw reads in both cases and allowed to generate the genome consensus sequence of 28,445 and 28,958 bp, for the two isolates, respectively. Regardless of using a read depth cutoff of 10X or 30X, we obtained around 89 and 94% of SARS-CoV-2 genome coverage, for the clinical samples, respectively. The genome coverage variation between both clinical samples reflected the initial viral load difference, which was even more marked at 100X (Table 1), as is the difference between the median read depths. However, due to the high sequencing error rate around 8%, we observed that a reliable consensus sequence should be analyzed at a minimum depth coverage at 30X, or even 100X for some nucleotides. Conversely, the genome coverage for the isolates was quite similar, with at 30X 94.9% and 95.3%, respectively (Table 1). For the clinical sample with the lower viral load CIBU-200107, the genome coverage was significantly improved by the isolation step, less so for the second clinical sample with a very low Ct value (Figure 2). Concerning the bioinformatics analysis, the RAMPART tool (Read Assignment, Mapping and Phylogenetic Analysis in Real Time) developed by the ARTIC network was helpful and time-saving to visualize genome coverage in real time and allowed to stop the run when sufficient data were generated. Globally, using our amplicon-based approach, combined with the MinION platform, we were able to obtain the near full-length genome of the studied viral specimens in around 8 h, from samples to sequences data.

TABLE 1 | Overview of MinION and iSeq100TM system sequences data obtained for both clinical specimens and corresponding isolates analyzed in this study.

	Nanopore MinION sequences		iSeq100 TM system sequences	
	Clinical sample	Isolate	Clinical sample	Isolate
CIBU-200107				
Total raw reads	45,625	107,532	3,990,761	3,770,873
Total mapped reads (% raw reads)	43,952 (96.3%)	107,529 (99.9%)	2,752,222 (69.0%)	3,764,987 (99.8%)
Average reads length	1,074	1,021	140	144
Base error rate	8.3%	7.9%	0.2%	0.1%
Median read depth	436X	2,979X	4,833X	20,054X
Genome coverage at reads depth 10X	89.8%	95.1%	94.7%	97.9%
Genome coverage at reads depth 30X	89.7%	94.9%	92.8%	97.8%
Genome coverage at reads depth 100X	76.0%	94.9%	89.5%	97.7%
CIBU-200132				
Total raw reads	122,113	120,958	4,046,340	3,790,945
Total mapped reads (% raw reads)	122,108 (99.9%)	120,956 (99.9%)	4,038,699 (99.8%)	3,786,372 (99.9%)
Average reads length	1,074	1,024	141	144
Base error rate	8.0%	7.9%	0.2%	0.1%
Median read depth	3,418X	3,607X	20,229X	19,163X
Genome coverage at reads depth 10X	94.9%	96.8%	97.8%	99.5%
Genome coverage at reads depth 30X	94.9%	95.3%	97.3%	98.7%
Genome coverage at reads depth 100X	94.9%	94.9%	96.7%	98.0%

Illumina iSeq100TM System Sequencing

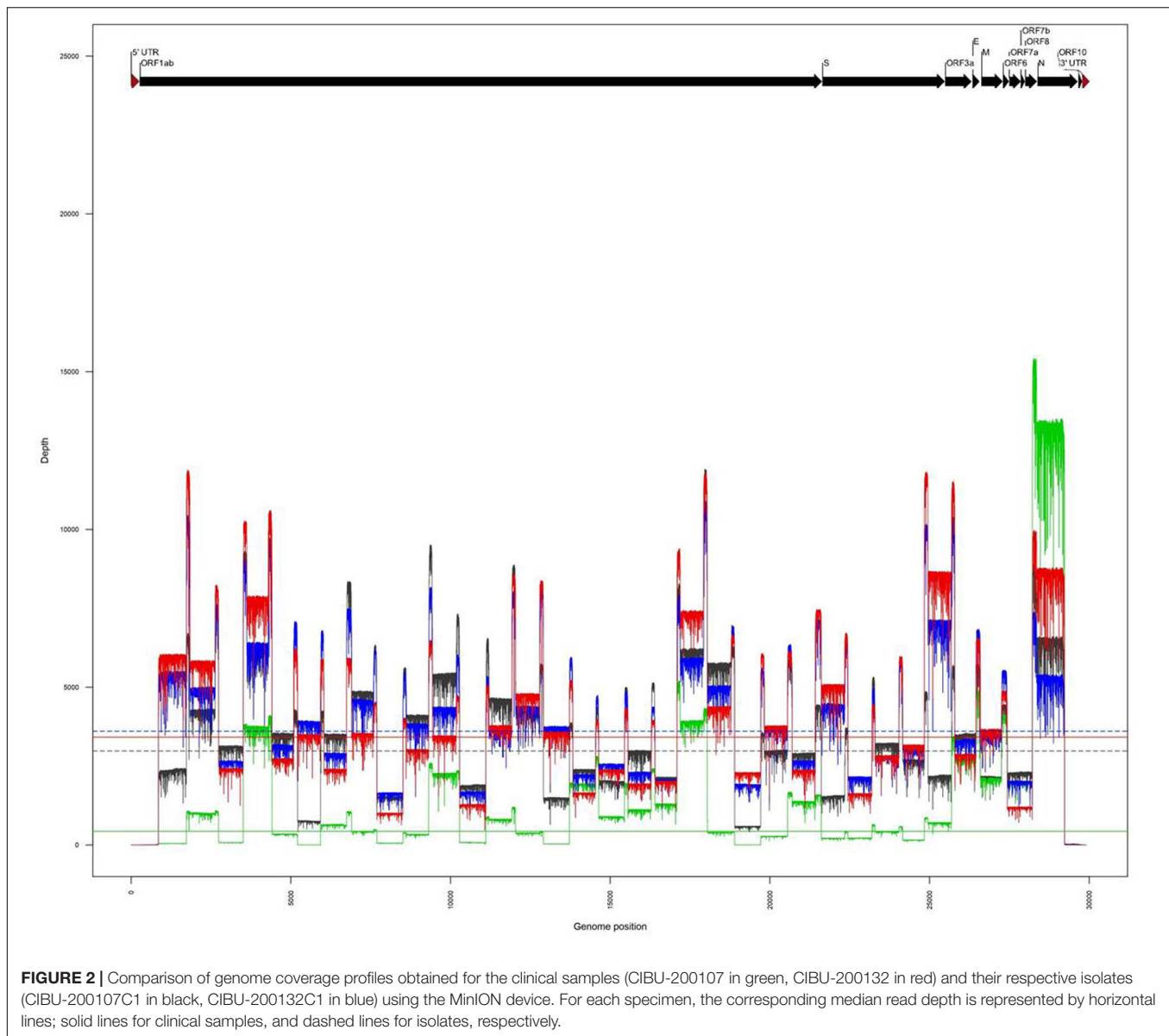
The same amplicons, obtained from the two clinical samples CIBU-200107 and CIBU-200132, were multiplexed and sequenced using the Illumina iSeq100TM system during a run of 17 h. Illumina run generated 3,990,761 and 4,046,340 raw reads with a Phred quality score of 28, for the two clinical samples, respectively, with an average of 140 bp read length. The average base error rate was around 0.2%. The mapping against Wuhan SARS-CoV-2 reference (NC_045512), by using BWA-MEM, retrieved 2,752,222 (69%) and 4,038,699 reads (99.8%) for the two clinical specimens, respectively (Table 1). A near full-length consensus sequence of 28,243 and 29,253 bp was generated for the two clinical samples, respectively. Concerning the amplicons obtained from both isolates CIBU-200107C1 and CIBU-200132C1, they were also multiplexed and run independently in the same conditions. The run produced 3,770,873 and 3,790,945 of high-quality raw reads, with a very low base error rate of 0.1%, with an average of 144 bp read length (Table 1). By using BWA-MEM, the mapping against Wuhan SARS-CoV-2 reference (NC_045512) used almost all the produced raw reads (99.9%) in both cases and allowed to generate the genome consensus sequence of 29,283 and 29,727 bp, for the two isolates, respectively. With a read depth of 10X, 94.7 and 97.8% of genome coverage was obtained with a reliable consensus sequence, for the two clinical samples, respectively. Despite the viral load difference between the two clinical specimens, the quite high genome coverage is probably due to the low base error rate and the huge number of obtained reads. Finally, the median read depth (ranging from 19,163X to 20,229X) was quite similar for the high viral load clinical sample and both isolates, except CIBU-200107 with 4,833X median value (Figure 3). These observed variations in read

depth are the direct consequence of the variable amplification efficiency of the primer pairs used in the multiplex PCR pool, impacted by the viral load of the samples as well as the secondary structures of the viral genome which affect primer binding. These results underlined that high-quality sequences could be obtained with greater samples multiplexing. Using an amplicon-based approach, combined with the iSeq100TM system platform, we were able to obtain the near full genome of the studied SARS-CoV-2 in around 24 h.

Mobile Next-Generation Sequencing Platforms Comparison for SARS-CoV-2 Genome Procurement

Viral consensus genomic sequences were rapidly and easily obtained for the two SARS-CoV-2 clinical specimens and their respective isolates, by using the two different sequencing platforms, MinION and iSeq100TM system. We noted that the sequences obtained for the same specimen are identical, regardless of the sequencer used. Sequence analysis of the SARS-CoV-2 consensus genomes showed that they belong to the proposed phylogenetic clade A2², which is consistent with the phylogenetic analysis of the other human French SARS-CoV-2 sequences (Gámbaro et al., 2020). Compared to the Wuhan reference sequence (NC_045512), the viral genomic sequences CIBU-200107 and CIBU-200132 (GISAID Accession numbers EPI_ISL_437689 and EPI_ISL_437690) display seven and six nucleotide mutations, respectively. Both share five nucleotide mutations (C241T, C3037T, C14408T, A23403G, and G25563T), among which three have led to amino acid mutations P4715L in ORF1ab, D614G in S, and Q57H in ORF3a. It is noteworthy that

²<https://nextstrain.org/>



the mutation D614G in the Spike glycoprotein is specific of clade A2, encompassing most of the French SARS-CoV-2 sequences. The CIBU-200107 additionally contains two more mutations in ORF1ab, C1059T (T265I) and C10582T, whereas the CIBU-200132 includes an additional one, C2416T. Comparison of the genomes obtained for the clinical specimens and their respective isolates show that the sequences are strictly identical, suggesting that culture of the SARS-CoV-2 on Vero E6 cells did not induce any molecular change, at least during the first passage.

With many challenges still to be overcome in real-time genomics in rapid-response diagnostics, we evaluated the performances of these both benchtop instruments. While the library preparation time is similar (approximately 3 h) between the two approaches and require the same benchtop laboratory equipment (**Figure 1**), the MinION is more efficient and versatile in terms of run time, with the procurement of robust

results after 2–4 h, compared to an incompressible run time of 17 h for the iSeq100™ system. It is important to take into account the thawing time of the Illumina cartridge reagents which takes at least 6 h in comparison to the immediately available MinION loading reagents which are stored at room temperature. MinION needs a powerful laptop computer and a stable broadband internet connection, whereas Illumina system is stand-alone at least in the first bioinformatics steps, as basecalling and demultiplexing. Concerning the sequencers themselves, the two systems are equivalent in terms of ease of use and maintenance and are therefore both suitable for experiments in resource-limited settings. The read length was fixed to 150 nt by the Illumina chemistry, and Nanopore reads were longer and entirely covered the length of the generated amplicons with an average of 1,074 nt, after trimming with Porechop (**Table 1**). While ONT improved sequencing accuracy, a still relatively high error rate of

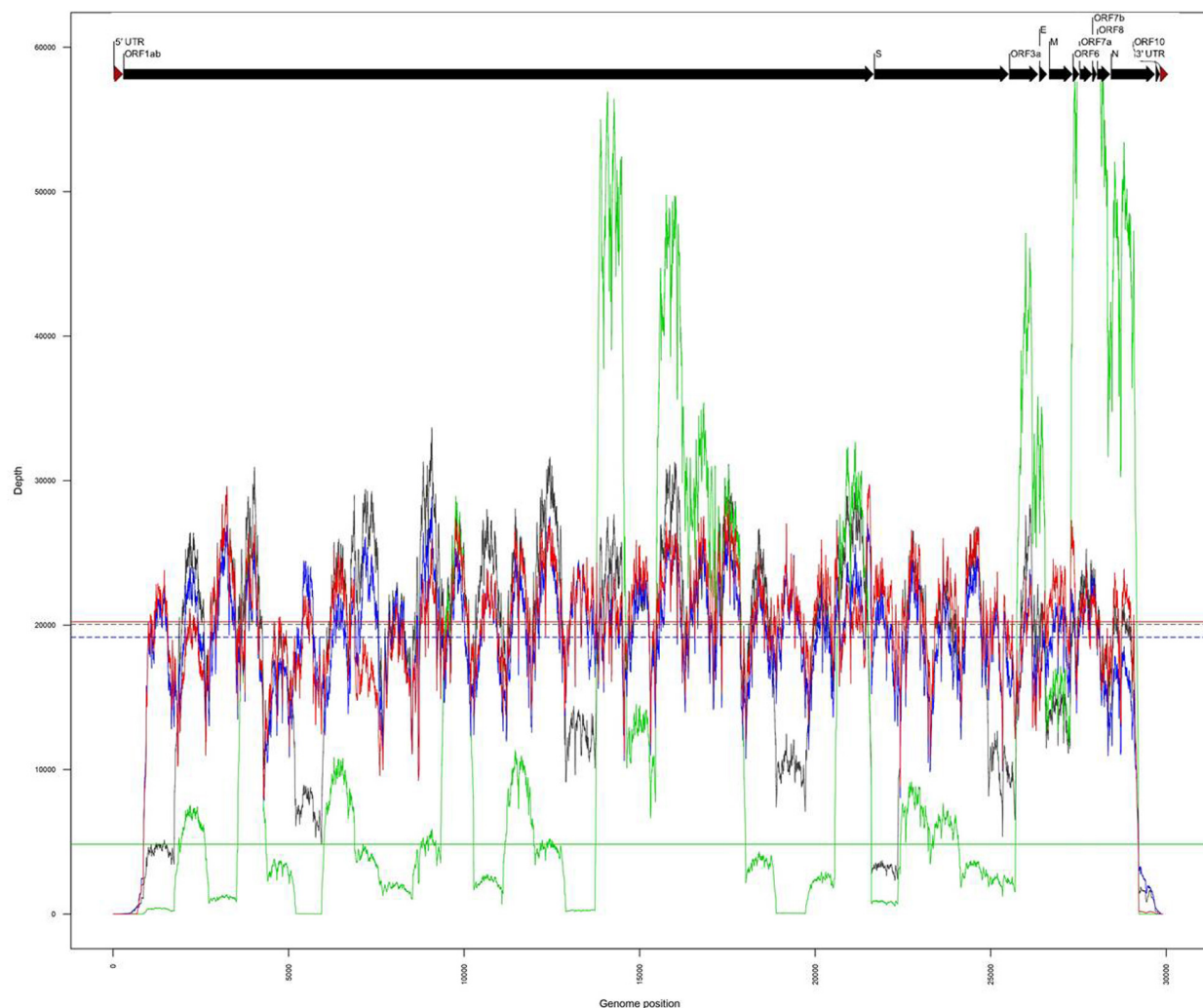


FIGURE 3 | Comparison of genome coverage profiles obtained for the clinical samples (CIBU-200107 in green, CIBU-200132 in red) and their respective isolates (CIBU-200107C1 in black, CIBU-200132C1 in blue) using Illumina iSeq100™ system. For each specimen, the corresponding median read depth is represented by horizontal lines; solid lines for clinical samples and dashed lines for isolates, respectively.

the MinION raw reads (around 8%) was noticed in comparison with Illumina ones (around 0.2%), requiring a higher read depth to obtain a confident consensus sequence. Overall, for each sample, we obtained a better coverage of the genome with the Illumina approach, and this as early as 10X of read depth. By comparing MinION and iSeq100™ system produced consensus genomes, we observed that the necessary coverage threshold is higher for MinION than for iSeq100™ system.

DISCUSSION

Novel mammalian coronaviruses are regularly identified and involved in epidemics and severe diseases, such as SARS-CoV, which emerged in southern China in 2002 (Drosten et al., 2003), and MERS-CoV, first described in Saudi Arabia in 2012 and still circulating (Song et al., 2019). At the source of viral outbreaks,

readily adaptable methods with rapid turn-around times are necessary for pathogen identification, surveillance programs, and public health strategies. Whole genome sequencing with next-generation sequencers is now widely used in specialized laboratories, which have required equipment and long experience for this implementation. However, it is also essential that this approach is accessible at the pen-side of infected humans to investigate outbreaks in remote areas. In this regard, we therefore evaluated two space-saving and easily portable next-generation sequencers, MinION and iSeq100™ system, through the current pandemic of COVID-19.

Although the initial investment gives an advantage to the MinION, the cost of reagents on the other hand is similar, and both approaches require the identical minimum of laboratory equipment to generate the libraries. By using the Primal scheme software on the first released SARS-CoV-2 reference sequence, we immediately developed an amplicon-based approach to sequence

around 1,000 bp amplicons on our small portable sequencers. It has been reported that the ampliseq panel proposed by ARTIC network, based on 109 amplicons, may lead to coverage bias due to dimer formation between primers (Itokawa et al., 2020). Our choice to produce 36 longer amplicons, with lower number of primers, could be more efficient. Nevertheless, for all the specimens, the extremities of the genome could not be obtained. As often in the genome sequencing of single-strand RNA viruses, the 3' and 5' non-coding regions are quite difficult to directly acquire, which would rather require an approach such as rapid amplification of cDNA ends PCR (RACE-PCR).

We successfully proved that such targeted approach performs well directly with clinical samples. The counterpart of this workflow requires prior knowledge of the pathogen but, on the other hand, could be promptly adapted to all viruses of interest, including those with very large genomes. We found that viral genomic sequencing, even with low-throughput sequencers, could be performed with high confidence and be achieved directly from clinical specimens with significantly shortened run time. Although higher samples multiplexing is suitable, these relatively low throughput sequencers can also run economically with a small number of samples, which is a great advantage when the information is needed as quickly as possible. Both sequencers generated the correct consensus genomes, with a slightly lower coverage for MinION and a higher error rate than the iSeq100TM system. However, we were able to obtain the genome of the studied specimens in around 8 h with the MinION, illustrating the quasi real-time sequencing capacity to effectively support health authorities. The main defect of the iSeq100TM system finally lies in its run time. Nanopore is regularly considered to have the advantage over other technologies of high portability and fast turnaround time (Gardy and Loman, 2018). It should be noted that today Illumina offers a sequencer, which tends toward the same interesting features, while maintaining the high quality of the reads, which is at the origin of Illumina's success. The MinION retains the advantage of greater versatility, with the possibility of analyzing data in true real time, particularly with tools such as RAMPART, and thus being able to stop the run after a few hours as soon as the amount of data is sufficient. Concerning the further generation of the alignment of amplicon reads against the reference and the variant calling, both devices usually offer tools in the cloud, such as Local Run Manager (DNA amplicon analysis module) for iSeq100TM system, and EPI2ME for MinION. In outbreak situations, these two sequencers therefore make rapid sequencing accessible to a very large number of laboratories, requiring a minimum of molecular biology skills. Nevertheless, the improvement of the dedicated bioinformatics user-friendly resources should still be developed to take into account the low network and informatics resources in the field.

In conclusion, our investigation led to the development of an amplicon-based sequencing approach and the characterization of the viral genome as SARS-CoV-2 within a few hours, adaptable in field conditions, with two easily manageable next-generation sequencers, the nanopore MinION and the Illumina iSeq100TM system. They provide faster and more

cost-effective small-scale runs, which avoid outsourcing for laboratories in remote areas, and allow control of the whole sequencing process. The MinION and iSeq100TM system bring the power of next-generation sequencing to virtually any laboratory, with amplicon-based whole-genome sequencing protocols rapidly transposable to all molecular investigations of epidemic prone pathogens.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

VH performed all the data analysis. AK and CB conducted the experiments. VH, AK, CB, JV, and VC conceived the study. VH, AK, CB, and VC analyzed the results and prepared the manuscript. SM, CB, and J-CM provided scientific guidance. All authors reviewed the manuscript.

FUNDING

This work was made possible thanks to financial support obtained through the «URGENCE Nouveau Coronavirus» fundraising campaign of Institut Pasteur. This project also received funding from the French Institut Carnot Pasteur “Microbes et Santé” (ANR 19 CARN 0023-01) and Institut Carnot “France Futur Élevage” (ANR 19 CARN 0012-01), in the framework of the FIELD project.

ACKNOWLEDGMENTS

We acknowledged all the members of the Laboratory for Urgent Response to Biological Threats (Institut Pasteur) who are involved in the COVID-19 diagnosis and India Leclercq for her support in viral isolation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2020.571328/full#supplementary-material>

REFERENCES

- Colman, R. E., Mace, A., Seifert, M., Hetzel, J., Mshael, H., Suresh, A., et al. (2019). Whole-genome and targeted sequencing of drug-resistant *Mycobacterium tuberculosis* on the iSeq100 and MiSeq: a performance, ease-of-use, and cost evaluation. *PLoS Med.* 16:e1002794. doi: 10.1371/journal.pmed.1002794
- Dong, E., Du, H., and Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* 20, 533–534. doi: 10.1016/S1473-3099(20)30120-1
- Drosten, C., Günther, S., Preiser, W., van der Werf, S., Brodt, H.-R., Becker, S., et al. (2003). Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.* 348, 1967–1976. doi: 10.1056/NEJMoa030747
- Faria, N. R., Sabino, E. C., Nunes, M. R. T., Alcantara, L. C. J., Loman, N. J., and Pybus, O. G. (2016). Mobile real-time surveillance of Zika virus in Brazil. *Genome Med.* 8:97. doi: 10.1186/s13073-016-0356-2
- Gámbaro, F., Behillil, S., Baidaliuk, A., Donati, F., Albert, M., Alexandru, A., et al. (2020). Introductions and early spread of SARS-CoV-2 in France. *BioRxiv* [Preprint] doi: 10.1101/2020.04.24.059576
- Gard, J. L., and Loman, N. J. (2018). Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat. Rev. Genet.* 19, 9–20. doi: 10.1038/nrg.2017.88
- Hoenen, T., Groseth, A., Rosenke, K., Fischer, R. J., Hoenen, A., Judson, S. D., et al. (2016). Nanopore sequencing as a rapidly deployable ebola outbreak tool. *Emerg. Infect. Dis.* 22:331. doi: 10.3201/eid2202.151796
- Houldcroft, C. J., Beale, M. A., and Breuer, J. (2017). Clinical and biological insights from viral genome sequencing. *Nat. Rev. Microbiol.* 15, 183–192. doi: 10.1038/nrmicro.2016.182
- Itokawa, K., Sekizuka, T., Hashino, M., Tanaka, R., and Kuroda, M. (2020). A proposal of alternative primers for the ARTIC Network's multiplex PCR to improve coverage of SARS-CoV-2 genome sequencing. *bioRxiv* [Preprint] doi: 10.1101/2020.03.10.985150
- Kafetzopoulou, L. E., Pullan, S. T., Lemey, P., Suchard, M. A., Ehichioya, D. U., Pahlmann, M., et al. (2019). Metagenomic sequencing at the epicenter of the Nigeria 2018 Lassa fever outbreak. *Science* 363, 74–77. doi: 10.1126/science.aau9343
- Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi: 10.1093/bioinformatics/bty191
- Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., et al. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 395, 565–574. doi: 10.1016/S0140-6736(20)30251-8
- Milne, I., Stephen, G., Bayer, M., Cock, P. J. A., Pritchard, L., Cardle, L., et al. (2013). Using Tablet for visual exploration of second-generation sequencing data. *Brief Bioinform.* 14, 193–202. doi: 10.1093/bib/bbs012
- Quick, J., Grubaugh, N. D., Pullan, S. T., Claro, I. M., Smith, A. D., Gangavarapu, K., et al. (2017). Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat. Protoc.* 12, 1261–1276. doi: 10.1038/nprot.2017.066
- Song, Z., Xu, Y., Bao, L., Zhang, L., Yu, P., Qu, Y., et al. (2019). From SARS to MERS, thrusting coronaviruses into the spotlight. *Viruses* 11:59. doi: 10.3390/v11010059
- Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., et al. (2020). A new coronavirus associated with human respiratory disease in China. *Nature* 579, 265–269. doi: 10.1038/s41586-020-2008-3
- Ye, Z.-W., Yuan, S., Yuen, K.-S., Fung, S.-Y., Chan, C.-P., and Jin, D.-Y. (2020). Zoonotic origins of human coronaviruses. *Int. J. Biol. Sci.* 16, 1686–1697. doi: 10.7150/ijbs.45472
- Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 270–273. doi: 10.1038/s41586-020-2012-7

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Hourdel, Kwasiborski, Balière, Matheus, Batéjat, Manuguerra, Vanhomwegen and Caro. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.