



**HAL**  
open science

## Whole-genome single molecule real-time sequencing of SARS-CoV-2 Omicron

Florence Nicot, Pauline Trémeaux, Justine Latour, Romain Carcenac, Sofia Demmou, Nicolas Jeanne, Noémie Ranger, Clémentine de Smet, Stéphanie Raymond, Chloé Dimeglio, et al.

### ► To cite this version:

Florence Nicot, Pauline Trémeaux, Justine Latour, Romain Carcenac, Sofia Demmou, et al.. Whole-genome single molecule real-time sequencing of SARS-CoV-2 Omicron. *Journal of Medical Virology*, 2023, 95 (2), pp.e28564. <10.1002/jmv.28564>. <hal-04976056>

**HAL Id: hal-04976056**

**<https://ut3-toulouseinp.hal.science/hal-04976056v1>**

Submitted on 4 Mar 2025



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 CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License

# Whole-genome single molecule real-time sequencing of SARS-CoV-2 Omicron

Florence Nicot<sup>1</sup>  | Pauline Trémeaux<sup>1</sup> | Justine Latour<sup>1</sup> | Romain Carcenac<sup>1</sup> | Sofia Demmou<sup>1</sup> | Nicolas Jeanne<sup>1</sup> | Noémie Ranger<sup>1</sup> | Clémentine De Smet<sup>1</sup> | Stéphanie Raymond<sup>1,2</sup> | Chloé Dimeglio<sup>1,2</sup>  | Jacques Izopet<sup>1,2</sup>

<sup>1</sup>Virology Laboratory, Toulouse University Hospital, Toulouse, France

<sup>2</sup>INSERM UMR 1291–CNRS UMR 5051, Toulouse Institute for Infectious and Inflammatory Diseases (INFINITY), Toulouse, France

## Correspondence

Florence Nicot, Virology Laboratory, Toulouse University Hospital, Hôpital Purpan 330, Ave de Grande-Bretagne, 31059 Toulouse, France. Email: [nicot.f@chu-toulouse.fr](mailto:nicot.f@chu-toulouse.fr)

## Abstract

New variants and genetic mutations of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) genome can only be identified using accurate sequencing methods. Single molecule real-time (SMRT) sequencing has been used to characterize Alpha and Delta variants, but not Omicron variants harboring numerous mutations in the SARS-CoV-2 genome. This study assesses the performance of a target capture SMRT sequencing protocol for whole genome sequencing (WGS) of SARS-CoV-2 Omicron variants and compared it to that of an amplicon SMRT sequencing protocol optimized for Omicron variants. The failure rate of the target capture protocol (6%) was lower than that of the amplicon protocol (34%,  $p < 0.001$ ) on our data set, and the median genome coverage with the target capture protocol (98.6% [interquartile range (IQR): 86–99.4]) was greater than that with the amplicon protocol (76.6% [IQR: 66–89.6], [ $p < 0.001$ ]). The percentages of samples with >95% whole genome coverage were 64% with the target capture protocol and 19% with the amplicon protocol ( $p < 0.05$ ). The clades of 96 samples determined with both protocols were 93% concordant and the lineages of 59 samples were 100% concordant. Thus, target capture SMRT sequencing appears to be an efficient method for WGS, genotyping and detecting mutations of SARS-CoV-2 Omicron variants.

## KEYWORDS

long read sequencing, Omicron, SARS-CoV-2 genotyping, SMRT sequencing, whole genome sequencing

## 1 | INTRODUCTION

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a member of the *Coronaviridae* family that caused the COVID-19 pandemic,<sup>1,2</sup> has a single-strand positive-sense RNA genome approximately 29.9 kb long that encodes several proteins, including the spike

(S) structural protein.<sup>3</sup> The 1273 amino-acid (AA) long S protein attaches the virus to the host cell receptor via its receptor binding domain (RBD, residues 319–529).<sup>4</sup> The virus genome accumulates mutations that are associated with transmissibility, escape from neutralizing antibodies and virulence.<sup>5–7</sup> The SARS-CoV-2 genome has diverged during the pandemic to produce several variants that

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Journal of Medical Virology* published by Wiley Periodicals LLC.

differ in their biology and/or geographical distribution. Five of these variants have been declared variants of concern (VOCs)<sup>8</sup>: Alpha (B.1.1.7),<sup>9</sup> Beta (B.1.351),<sup>10</sup> Gamma (P.1),<sup>11</sup> Delta (B.1.617.2), and Omicron (B.1.1.529).<sup>12,13</sup> To date, only Omicron variants are circulating worldwide; they are divided in several clades: 21K/BA.1, 21L/BA.2, 22A/BA.4 and sublineages, 22B/BA.5 and sublineages, 22C/BA.2.12.1, 22D/BA.2.75 and sublineages, 22E/BQ.1 and sublineages (<https://Nextstrain.org/nextclade/SARS-CoV-2>).

Clade 21K/BA.1 caused most of infections that occurred in France between January and March 2022, followed by 21L/BA.2 between March 2022 and mid-June 2022, only to be replaced by 22B/BA.5 from mid-June 2022. The Omicron variants are more highly mutated than other VOCs, with 50 initial mutations throughout the genome (including 32 in the spike protein).<sup>14,15</sup> These mutations can affect the biological characteristics of the virus, increasing its transmissibility and enabling it to escape the immune system.<sup>16</sup>

Genomic surveillance and clinical management require the rapid identification of variants that are efficiently transmitted or that escape the host immune response. Next generation sequencing protocols, mostly based on multiplex amplicon or hybrid capture methods running on the Illumina and Oxford Nanopore Technologies (ONT) platforms,<sup>17-23</sup> have been developed to study the genomic diversity of SARS-CoV-2 worldwide.<sup>24-26</sup> We have recently shown that the Pacific Biosciences (PacBio) single molecule real-time (SMRT) sequencing method, which provides accurate whole genome sequences (WGS), can be used for genotyping and detecting mutations in Alpha and Delta variants.<sup>27</sup> However, no Omicron variants were included in this study. As reported for ARTIC v4 primers set and short-read sequencing or MIDNIGHT v1 primers set and long-read sequencing of 1.2 kb tiled amplicons,<sup>14</sup> Omicron mutations can decrease the enrichment efficiency of polymerase chain reaction (PCR) amplification. Primer sets have been adapted.<sup>28,29</sup>

This study assesses the performance of the target capture SMRT sequencing protocol for genotyping and detecting mutations of SARS-CoV-2 BA.2 and BA.5 Omicron variants and compared the results with those obtained with a 1.2 kb amplicon SMRT sequencing protocol using suitably adapted primers.

## 2 | MATERIALS AND METHODS

### 2.1 | Samples

We sequenced 172 nasopharyngeal samples from SARS-CoV-2 RNA positive patients (cycle thresholds, Ct 11.8–34.9). These samples had been stored at –80°C before sequencing in the Toulouse University Hospital Virology Laboratory. The Ct values were obtained with ID™ SARS-CoV-2/VOC Revolution Pentaplex (ID solutions) using one target sequence in the N gene and two in the RdRp gene. The consecutive samples were taken between May 30 and June 19, 2022 during an Omicron wave in France, at the beginning of the switch from BA.2 to BA.5 variants. Two strains of SARS-CoV-2 B.1.1.254/20B (EPI\_ISL\_804374) and B.1.617.2/21I (EPI\_ISL\_10316331) from a

culture supernatant and one SARS-CoV-2 negative sample were included as controls in each 96-well plate.

### 2.2 | SARS-CoV-2 RNA extraction

Virus RNA was extracted from 180 µL transport medium with the MGIEasy Nucleic Acid extraction kit on the MGI SP 960 system (Beijing Genome Institute) according to the manufacturer's instructions.

### 2.3 | Target capture PacBio SMRT sequencing

We used the SARS-CoV-2 Enrichment (PacBio) kit following the PacBio HiFiViral for SARS-CoV-2 workflow, High-Throughput Multiplexing for Full-Viral Genome Sequencing of SARS-CoV-2 version 3 (hereafter referred to as target capture). This protocol uses molecular inversion probes with a 675 bp target insert and was conducted as previously described.<sup>27</sup> Briefly, the first step was simultaneous complementary DNA (cDNA) synthesis and probe hybridization, followed by circularization and barcoded amplification. Samples were then pooled for library construction and the pool purified with AMPure PacBio beads (Pacific Bioscience). The DNA was quantified with the Quantifluor DSDNA system running on a Roche LC480 instrument. Last, a SMRT bell library was prepared and sequenced with the SMRTbell Express Template Prep 2.0 kit according to the manufacturer's instructions running on a Sequel IIe platform.

The resulting data were analyzed with SMRT Link software using the HiFiViral SARS-CoV-2 Analysis Application with the default parameters ([https://www.pacb.com/wp-content/uploads/SMRT\\_Link\\_User\\_Guide\\_v11.0.pdf](https://www.pacb.com/wp-content/uploads/SMRT_Link_User_Guide_v11.0.pdf)), except for the minimum base coverage, which was changed to 10 reads (ECDC recommendations for SARS-CoV-2 sequencing).<sup>30</sup> Consensus sequences were then analyzed with Pangolin lineages (v4.1.1, <https://cov-lineages.org/resources/pangolin.html>) and Nextstrain clades (v2.3.1, <https://nextstrain.org/sars-cov-2>). Finally, customized python script (v3.8.8) analysis was used to generate a user-friendly report, including the number of mapped reads, median coverage, list of S gene mutations (substitutions, insertions and deletions), S gene missing positions, multiple strains probability and previously found clades and lineages.

### 2.4 | 1.2 kb amplicons PacBio SMRT sequencing

We used the PacBio HiFiViral for SARS-CoV-2 workflow, High-Throughput Multiplexing 1.2 kb Amplicons for Full-Viral Genome Sequencing of SARS-CoV-2, based on 29 1.2 kb amplicons across the SARS-CoV-2 genome (hereafter referred to as amplicon) with improved primers for amplicons A1, A5, A10, A15, A21, A24, and A28, which have mutation-induced reduced coverage in Omicron variants (Table 1). The protocol was done as previously described.<sup>27</sup> Briefly, we first generated cDNAs and then obtained the 29 amplicons by two multiplex PCRs (with two primers pools, Pool 1 and Pool 2). These underwent another

PCR using barcoding primers tailed with the universal M13 sequence. Samples were pooled (1 µL of each sample from Pool 1 and Pool 2), quantified with the Quantifluor DSDNA system running on Roche LC480 instrument and normalized. SMRT bell libraries were constructed by pooling the normalized barcoded samples. SMRT bell libraries were sequenced with SMRTbell Express Template Prep 2.0 kits according to the manufacturer's instructions on a Sequel IIe platform.

The data were analyzed with the Snakemake pipeline for complete genome construction. This starts by demultiplexing HiFi reads with lima (v.2.6.0, <https://github.com/PacificBiosciences/barcoding>) and then creates the VCF files from pbAA clusters (PacBio tool, v.0.1.3 <https://github.com/PacificBiosciences/pbAA>). These files, along with the samtools (v1.11) depth file are used to build consensus sequences (CoSA, coronavirus Sequence Analysis, v9.0, <https://github.com/Magdoll/CoSA>). The minimum base coverage required to define a position was 10 reads. The consensus sequences were then analyzed with Pangolin (v4.1.1, <https://cov-lineages.org/resources/pangolin.html>) for lineages and Nextstrain (v2.3.1, <https://nextstrain.org/sars-cov-2>) for clades. The presence of multiple strains or of sample cross-over was checked by examining the number of mutations representing 10% to 90% of the quasispecies over the complete genome. The cut-off was more than 10 mutations between 10% and 90% of the quasispecies. We used our in-house python script (v3.8.8) analysis to generate a user-friendly report that included the number of mapped reads, the median coverage, the number of mutations between 10% and 90% of the quasispecies, a list of S gene mutations (substitutions, insertions and deletions), S gene missing positions, and the previously found clades and lineages.

See Supporting Information: Table S1 for GISAID accession numbers.

## 2.5 | Statistical analysis

A McNemar test was used for categorical variables and a Mann–Witney or a Wilcoxon sign-rank sum test for continuous variables. *p* values of <0.05 were considered to be significant.

**TABLE 1** Primers changed.

Primer name	Left	Right	Mutated positions (nt) (based on Wuhan-Hu-1 sequence NC_045512.2)
SARSCoV_1200_1	<b>TAAACGAACTTTAAATCTGTGTGGC</b>	GGTTGCATTCATTTGGTGACGC	C44T
SARSCoV_1200_5	<b>ACCTACTAAAAGGCTAGTGGC</b>	AGCATCTTTAGAGCAGGTGGA	G4184A
SARSCoV_1200_10	TTTACCAGGAGTTTCTGTGGTGT	<b>AATTGTGTCTCATAGCACATTGGTA</b>	G10447A, C10449A
SARSCoV_1200_15	TTTTAAGGAATTAAGTTGTATGCTGCT	<b>AACACACAACAGCATCGTCAGA</b>	C15714T
SARSCoV_1200_21	TCTGTAGTTTCTAAGTTGTCAAAGTGA	<b>AACACCACGTGTGAAAGAATTAGTGTGA</b>	TTACCCCTG(del)21632T <sup>a</sup>
SARSCoV_1200_24	<b>GTTAATAGGGGCTGAATATGTCAACA</b>	ATGAGGTGCTGACTGAGGGAAG	C23525T <sup>b</sup>
SARSCoV_1200_28	<b>TGCTTTTAGCCTTTCTGTTATTC</b>	GTTTGGCCTTGTGTTGTTGGC	C27807T

Note: Primers optimized for Omicron variants are in bold.

<sup>a</sup>AA69 in Spike gene.

<sup>b</sup>AA654 in Spike gene.

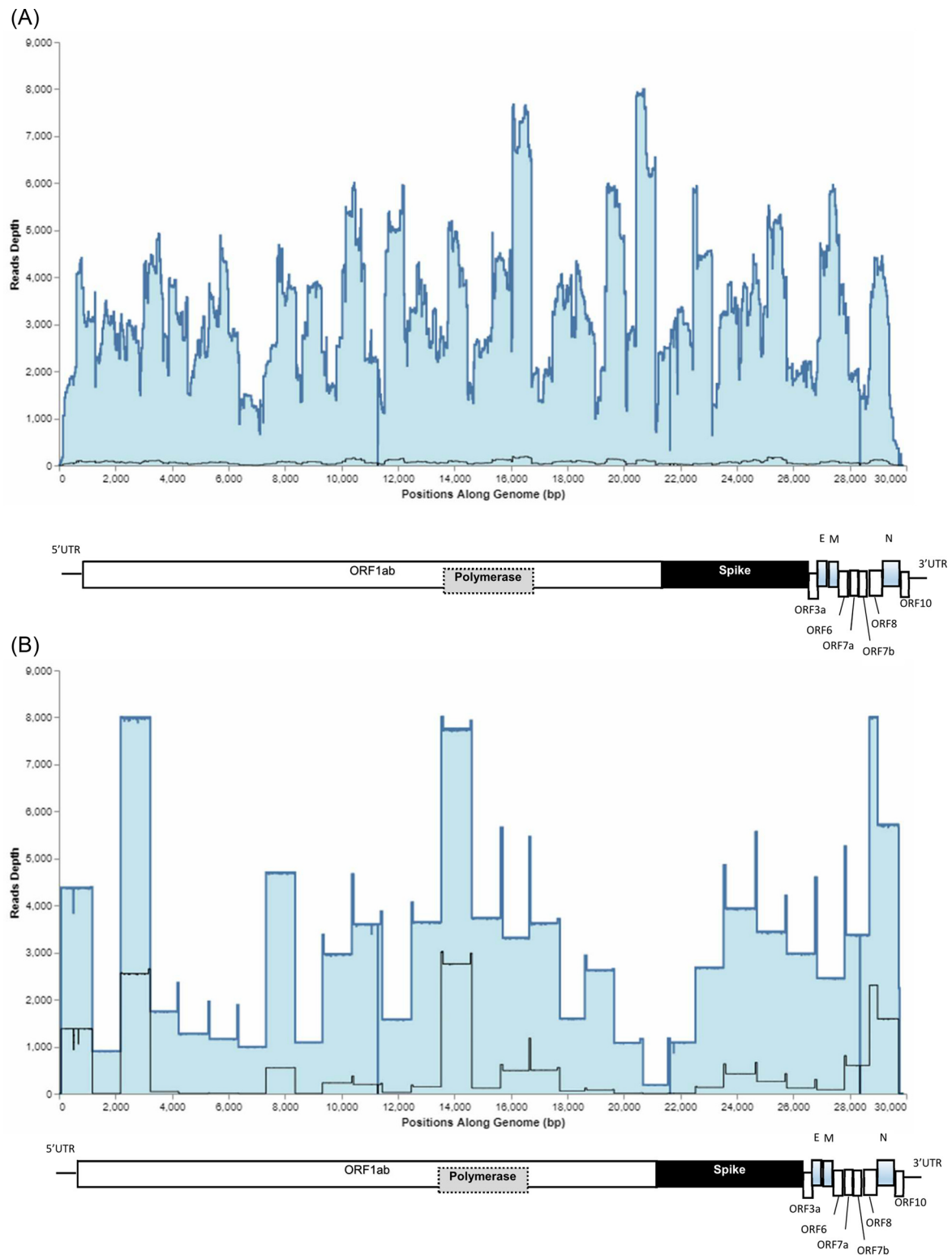
## 3 | RESULTS

### 3.1 | Clades and lineages from full-length genomes determined by PacBio SMRT target capture

We sequenced 172 samples (median Ct: 23.5 [interquartile range (IQR): 20.7–25.6]), 11/172 (6.2%) of which failed, so giving no genome coverage. SMRT link analysis flagged 15 samples (median Ct: 26.4 [IQR: 24.5–27.1]) as having multiple strains (probability >0.95). Sequence data were obtained for 146 (82%) samples with a median of 3828 reads per sample (IQR: 974–23 880), and a median read depth of 74 [IQR: 13–443] (Figure 1A and Supporting Information: Table S1). The median genome coverage was 97.1% [IQR: 60.9–99.2], and >95% for 80 (55%) samples. Most of the strains were clade 21L (Omicron) (52; 36%) and clade 22B (Omicron) (59; 40%) (Supporting Information: Table S1). Only the clade of 34 (27%) samples could be determined; their complete genome coverage was too low to determine the lineage (Supporting Information: Table S1) (11 [33%] classified recombinant by Nextstrain). Low genome coverage (<1%) prevented the determination of the clade of 4 (2.4%) samples (#34, #37, #88, and #154).

### 3.2 | Clades and lineages from full-length genomes determined by PacBio SMRT amplicon

The same 172 samples were sequenced with the amplicon protocol using suitably adapted primers to Omicron variants (Table 1). Sequencing failed for 59 samples (34%). Eleven samples (median Ct: 24.3 [IQR: 23.5–26.4]) had more than 10 positions along the genome with nucleotide variants representing 10%–90% of the quasispecies, suggesting the presence of multiple strains. Sequence data were obtained for 102 (59%) samples, with a median of 24 487 reads per sample (IQR: 9146–51 187) and a median read depth of 93 (IQR: 30–498) (Figure 1B and Supporting Information: Table S1). The



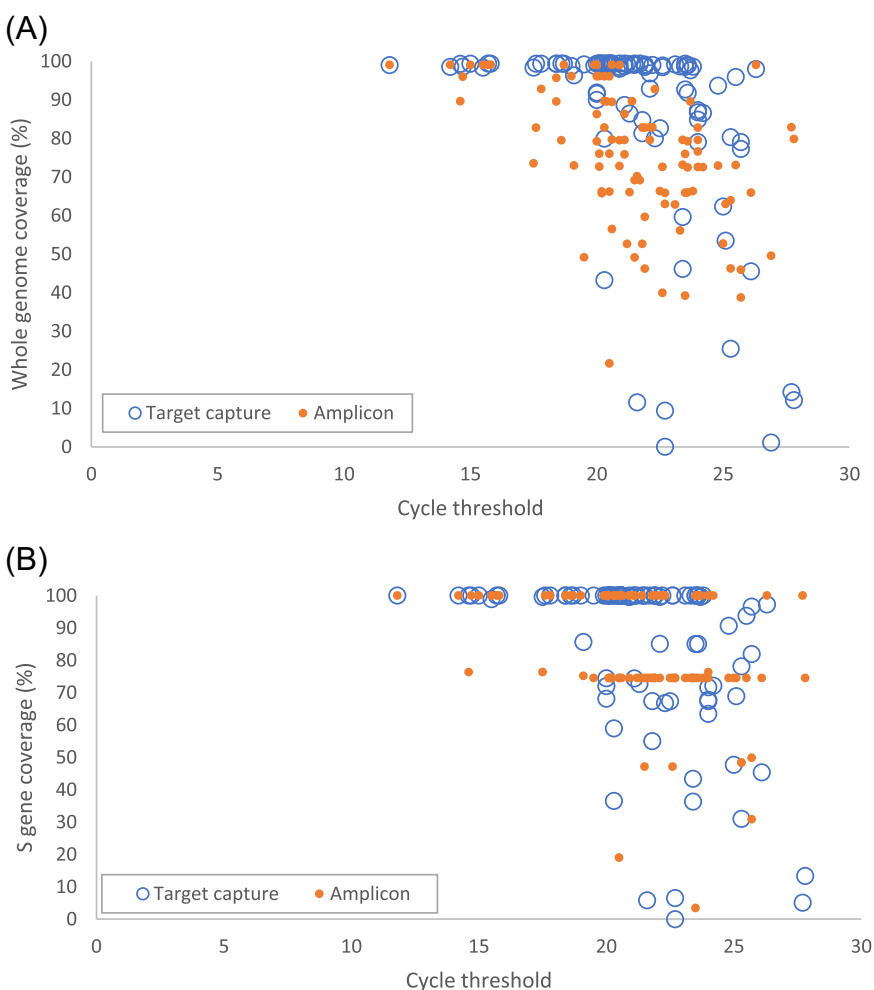
**FIGURE 1** Median read depth (blue line) and maximum coverage (blue area) along the entire severe acute respiratory syndrome coronavirus 2 genome (reference NC-045512.2) (A) target capture, (B) amplicon.

median genome coverage was 76% (IQR: 66–89.6), and >95% for 19 (19%) samples. The strains were clade 22B (Omicron) (52; 51%) and clade 21L (Omicron) (37; 37%). Only the clade of 46 samples (45%) was determined because their complete genome coverage was too low to determine the lineage.

### 3.3 | Comparison of the complete genome sequences obtained with the two methods

Multiple strains in at least one protocol was observed in 24 samples. Multiple strains were detected by both sequencing protocols in 2/24 samples, but the mutation patterns were unique for each protocol. The multiple strains detected with the target capture protocol in 6/24 samples were not detected with the amplicon protocol and those detected in 6/24 samples with the amplicon protocol were not detected with the target capture protocol. Lastly, 10/24 samples with multiple strains (7 obtained with the target capture protocol and three obtained with the amplicon protocol) failed to be sequenced with the other protocol. We therefore compared the two protocols by examining 148 samples. The target capture protocol produced fewer (nine, 6%) sequencing failures

than the amplicon protocol (51, 34%;  $p < 0.001$ ) but both failed to sequence nine samples. The amplicon protocol did not sequence 42 samples, including 21 (50%) that had a genome coverage of >80% with the target capture protocol, allowing clade and lineage assignment. Sequence data for 97 samples were obtained with both protocols. Whole genome coverage for samples with a high Ct value decreased for both protocols (Figure 2A). The Ct for samples that were sequenced by both protocols were significantly lower than those for samples that were not sequenced by either protocol (Figure 3). The Ct of the samples sequenced with target capture protocol but not with amplicon protocol were significantly lower than those of samples not sequenced by either protocol (Figure 3). The target capture protocol provided significantly greater median genome coverage (98.6% [IQR: 86.6–99.4]) than the amplicon protocol (76.6% [IQR: 66–89.6];  $p < 0.001$ ). Genome coverage was >95% for 62 (64%) samples analyzed by the target capture protocol and for 19 (19%) samples analyzed by the amplicon protocol ( $p < 0.05$ ). The clades of the 96 samples determined with both protocols were 93% concordant. The seven samples with low genome coverage (<60%) with one of the two protocols gave discordant results. The lineages of 59 samples determined with both protocols were 100% concordant. The sublineages of only six



**FIGURE 2** Whole genome (A) and S gene (B) coverage depending on the cycle threshold with both protocols.

samples (#14, #30, #70, #84, #118, and #131) differed, probably because the 1.2 kb amplicon protocol provided lower genome coverage than did the target capture protocol.

### 3.4 | Comparison of the spike sequences obtained with the two methods

The target capture protocol failed to sequence the S gene in 2/97 samples (#8 and #154), while the 1.2 kb amplicon protocol gave a spike genome coverage of 48% for #8 and 85% for #154. The median S gene coverage obtained with the target capture (100% [IQR: 72–100]) and the amplicon protocol (88% [IQR: 75–100]) for 95 samples were not significantly different. The coverage for the S gene of 61 samples (64%) analyzed with the target capture protocol and for 48 samples (50%) with the amplicon protocol was >95% ( $p=0.15$ ). The S gene coverage for samples with a high Ct value decreased for both protocols (Figure 2B). The mutations detected in 37/95 (39%) samples with both protocols were 100% concordant. The mutations detected in 15 (83%) of the 19 samples with  $Ct < 20$  were concordant. The target capture protocol missed positions in the S gene of 42 samples while the amplicon protocol missed positions in 48 samples. The target capture protocol missed positions impacting Omicron mutations in 33 samples and the amplicon protocol missed such positions in 47 samples. The target capture protocol detected mutations in 32 (33%) samples that were missed by the amplicon protocol and the amplicon protocol detected mutations missed by the target capture in 21 (20%) samples (Supporting Information: Table S2). Some mutations in 5 samples were missed by the target capture and/or amplicon protocols (Supporting Information: Table S2). The mutations missed by the target capture protocol were due to failure to sequence almost 600 initial amino acids in 18 (42%) samples (influencing key mutations in the RBD), or almost the whole S gene in eight (19%) samples. Mutations missed by the amplicon protocol were due to failure to amplify 1

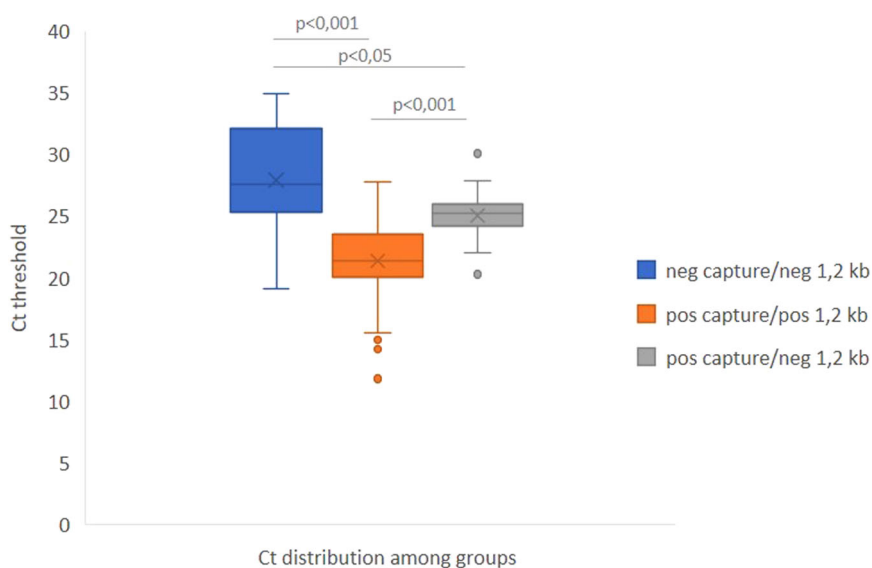
amplicon (covering the first 325 AA encoded by the S gene, which do not affect the key mutations in the RBD) in 39 (81%) samples or 2–4 amplicons in eight (17%) samples.

## 4 | DISCUSSION

Large scale, high-throughput WGS of SARS-CoV-2 is essential for rapid surveillance of the spread of new variants and their efficient follow-up, particularly immune-escaping variants that might interfere with vaccination or treatment. As we have shown that Pacbio SMRT sequencing is suitable for WGS of the early variants of SARS-CoV-2 produced by the pandemic, this report examines the suitability of the method for sequencing Omicron variants. We found that the target capture protocol was less affected by mutations that occur in BA.2 or BA.5 and that it appeared to perform better than the amplicon protocol.

The target capture protocol resulted in fewer sequence failures than did the amplicon protocol and gave a genome coverage >95% (the minimum recommended by ECDC)<sup>30</sup> of more sequences. The target capture protocol amplifies 675 bp short fragments of the SARS-CoV-2 and uses 1000 probes to tile the genome for 22-fold coverage, while the amplicon protocol amplifies 1.2 kb fragments. This is probably why the target capture protocol performed better, particularly in samples with a low viral load. A previous study showed that a short amplicon method is better at generating complete genomes from low viral load samples than a pooled long-amplicon method.<sup>31</sup>

The proportion of sample sequence failures obtained with the target capture method in this study of Omicron variants was similar to that obtained for the Alpha/Delta variants<sup>27</sup> and the complete genome coverages of the Omicron (median >98%) and Alpha/Delta variants (median >99%) were also similar.<sup>27</sup> However, the proportion of samples with a complete genome coverage of >95% was slightly lower (55%) than for the Alpha/Delta variants (63%), perhaps because



**FIGURE 3** Ct distributions for samples not sequenced by either method, sequenced by both methods, and sequenced by target capture but not by amplicon sequencing.

the Omicron samples of this study had lower viral loads. The median genome coverage of those Alpha/Delta and Omicron variant samples with a Ct < 20 using the target capture protocol were identical (99.4%) (data not shown), and the genome coverage was >98% for 94% of the Omicron variant samples. The target capture protocol provided data for the Alpha and Delta variants<sup>22</sup> that are similar to those obtained with Illumina or ONT sequencing.<sup>21,32-36</sup> The slight differences of genome coverage between these studies could be due to the protocol used (short/long fragment sequencing, primer sets), the variants sequenced, or the sample viral load. Feng et al.<sup>29</sup> showed that whole genome coverage of Delta and Omicron variants was > 99.7% with ARTICV4.1 or VarSkip2 using Illumina sequencing. We obtained lower genome coverages with the target capture SMRT sequencing protocol, probably because we used slightly longer amplicons. A limitation of this study is the absence of comparison with data obtained on Illumina sequencing platform on the same samples.

The co-circulation of various clade/lineage have been observed at different times and/or different geographic areas raising the questions of possible co-infections and so recombination events. Co-infections with a prevalence of 0.2%–0.6% have been described in several areas since the beginning of the pandemic.<sup>28,37,38</sup> The prevalence of Delta/Omicron or BA.1/BA.2 co-infections in France was estimated to be 0.18%–0.26%.<sup>28,39</sup> Recombinant SARS-CoV-2 variants have also been identified in both immunocompetent and immunocompromised individuals.<sup>39,40</sup> The target capture protocol and/or the amplicon protocol found several samples harboring multiple strains but this was for only one sequencing result (target capture or amplicon protocol). Any detected by both techniques had only a single specific mutation pattern, suggesting sample cross-over, as previously reported,<sup>41</sup> rather than multiple strains.

The mutated primer-binding sites of newly-emerging strains with more mutations may cause amplicon dropout or uneven coverage, making sequencing difficult. Primers must be updated periodically and re-validated to keep pace with virus evolution. The mutations that appeared in the Omicron variant at the end of 2021 decrease the enrichment efficiency of PCR amplification for both short-read and long-read sequencing.<sup>42</sup> ARTIC Network V4.1 primers or VarSkip2 primers were designed to optimize the short amplicon sequencing of Omicron variants<sup>28,29</sup> and MIDNIGHT V2 primers were produced for long amplicon sequencing based on nanopore technology.<sup>28</sup> It is much easier to adjust the primers for long-amplicon protocols because the primer sets are much smaller than those for short-amplicon protocols. However, in our study the genome coverage was low even for samples with a Ct < 20 (median genome coverage of 90%), despite optimization of the 1.2 kb amplicon protocol primers. Moreover, nine samples with low Ct values (Ct: 19–21) were insufficiently covered with the amplicon protocol and did not allow lineage assignment. This was probably due to technical issue during the process leading to misamplification. Higher genome coverage allowing lineage assignment was obtained in an independent reaction except for one sample (data not shown). Drop-outs are frequent in amplicon A21, covering

NC\_0455512.2: 20 582–21 642 nucleotide region with the new set of primers; this needs to be resolved. In contrast, the capture protocol primers have not been changed to accommodate Omicron variants.

New mutations in the Spike region accompany the emergence of new sub-lineages and this influences the activity of most anti-Spike monoclonal antibodies.<sup>43,44</sup> Accurate S gene sequencing is also needed to monitor SARS-CoV-2 in immunocompromised patients treated with monoclonal antibodies.<sup>45-47</sup> An *in silico* study compared the SARS-CoV-2 whole genome and spike gene detection rates of three primer sets yielding amplicon sizes of 400, 1200, and 2500 bp with the complete SARS-CoV-2 genomes sequences available in the NCBI in February 2022.<sup>48</sup> The primer set that yielded 400 bp amplicons covered the S gene in 59% of the genomes, the primer set that yielded 1200 bp amplicons covered the S gene in 76% of the genomes, and the primer set that yielded 2500 bp amplicons covered the S gene in 92% of the genomes, suggesting that long amplicon analysis gives better S gene coverage. The S gene can be sequenced with the 1.2 kb amplicon protocol using only four amplicons, one covering the RBD, which allows haplotyping of the main mutations in the S region. This is important for investigating the impact of certain mutations on variant evolution by modulating the effects of mutations at other sites, which could lead to antibody-escape substitutions in the RBD-domain.<sup>49</sup> PacBio SMRT sequencing of the S gene with amplicon lengths of 2.5–6.1 kb has been used to study spike gene quasispecies.<sup>45,50-52</sup> Sun et al.<sup>52</sup> showed that the virus population may consist of one dominant haplotype plus numerous minor haplotypes and that the complexity of quasispecies varies from one tissue to another, suggesting independent replication. Long amplicon sequencing can be used to accurately monitor SARS-CoV-2 quasispecies for compartmentalization studies, surveillance of escape-mutations in variants from patients treated with mAbs, and the emergence of new variants. It is clearly important to optimize amplicon primers, in particular those affecting regions of interest, such as the S gene.

To conclude, we find the target capture PacBio SMRT technology suitable for WGS of SARS-CoV-2, identifying circulating SARS-CoV-2 variants and detecting mutations in Omicron variants. The target capture protocol does not seem to be influenced by new mutations. Conversely, the accurate long-reads produced by the amplicon SMRT protocol are essential for studying SARS-CoV-2 quasispecies.

## AUTHOR CONTRIBUTIONS

Florence Nicot, Jacques Izopet, and Pauline Trémeaux designed the project. Romain Carcenac, Clémentine De Smet and Noémie Ranger carried out the experiments. Justine Latour, Sofia Demmou, and Nicolas Jeanne performed bioinformatics analyses. Florence Nicot, Pauline Trémeaux, and Jacques Izopet analyzed the study data. Chloé Dimeglio carried out the statistical analyses. Florence Nicot and Jacques Izopet wrote the manuscript. All the authors have approved the manuscript.

## ACKNOWLEDGMENTS

We thank Dr Owen Parkes for editing the English text. Target capture reagents were provided by Pacific Biosciences.

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

## DATA AVAILABILITY STATEMENT

The raw data are available on demand. Sequences are available on GISAID (accession numbers in Supporting Information: Table S1).

## ORCID

Florence Nicot  <http://orcid.org/0000-0003-2722-2159>

Chloé Dimeglio  <http://orcid.org/0000-0002-3713-4913>

## REFERENCES

- Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. *Nature*. 2020;579:265-269.
- Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;579:270-273.
- Khailany RA, Safdar M, Ozaslan M. Genomic characterization of a novel SARS-CoV-2. *Gene Rep*. 2020;19:100682.
- Wan Y, Shang J, Graham R, Baric RS, Li F. Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. *J Virol*. 2020;94:e00127-20.
- Rotondo JC, Martini F, Maritati M, et al. SARS-CoV-2 infection: new molecular, phylogenetic, and pathogenetic insights. Efficacy of current vaccines and the potential risk of variants. *Viruses*. 2021;13:1687.
- Khan MZI, Nazli A, Al-Furas H, et al. An overview of viral mutagenesis and the impact on pathogenesis of SARS-CoV-2 variants. *Front Immunol*. 2022;13:1034444.
- Starr TN, Greaney AJ, Hannon WW, et al. Shifting mutational constraints in the SARS-CoV-2 receptor-binding domain during viral evolution. *Science*. 2022;377:420-424.
- National Center for Immunization and Respiratory Diseases (U.S.): Division of Viral Diseases. SARS-CoV-2 variant classifications and definitions. 2021. Updated April 26, 2022. <https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html>
- Davies NG, Jarvis CI, Edmunds WJ, Jewell NP, Diaz-Ordaz K, Keogh RH. Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. *Nature*. 2021;593:270-274.
- Tegally H, Wilkinson E, Giovanetti M, et al. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature*. 2021;592:438-443.
- Voloch CM, da Silva Francisco R, de Almeida LGP, et al. Genomic characterization of a novel SARS-CoV-2 lineage from Rio de Janeiro, Brazil. *J Virol*. 2021;95(10):e00119-21. doi:10.1128/JVI.00119-21
- Karim SSA, Karim QA. Omicron SARS-CoV-2 variant: a new chapter in the COVID-19 pandemic. *Lancet*. 2021;398:2126-2128.
- Peng J, Liu J, Mann SA, et al. Estimation of secondary household attack rates for emergent spike L452R SARS-CoV-2 variants detected by genomic surveillance at a community-based testing site in San Francisco. *Clin Infect Dis*. 2022;74(1):32-39. doi:10.1093/cid/ciab283
- Ma W, Yang J, Fu H, et al. Genomic perspectives on the emerging SARS-CoV-2 Omicron variant. *Genom Insights*. 2022;20:60-69.
- Alkhatib M, Salpini R, Carioti L, et al. Update on SARS-CoV-2 Omicron variant of concern and its peculiar mutational profile. *Microbiol Spectr*. 2022;10:e0273221.
- Tian D, Sun Y, Xu H, Ye Q. The emergence and epidemic characteristics of the highly mutated SARS-CoV-2 Omicron variant. *J Med Virol*. 2022;94:2376-2383.
- Bal A, Destras G, Gaymard A, et al. Molecular characterization of SARS-CoV-2 in the first COVID-19 cluster in France reveals an amino acid deletion in nsp2 (Asp268del). *Clin Microbiol Infect*. 2020;26:960-962.
- Bhoyar RC, Jain A, Sehgal P, et al. High throughput detection and genetic epidemiology of SARS-CoV-2 using COVIDSeq next-generation sequencing. *PLoS One*. 2021;16:e0247115.
- Harilal D, Ramaswamy S, Loney T, et al. SARS-CoV-2 whole genome amplification and sequencing for effective population-based surveillance and control of viral transmission. *Clin Chem*. 2020;66:1450-1458.
- Bull RA, Adikari TN, Ferguson JM, et al. Analytical validity of nanopore sequencing for rapid SARS-CoV-2 genome analysis. *Nat Commun*. 2020;11:6272.
- Charre C, Ginevra C, Sabatier M, et al. Evaluation of NGS-based approaches for SARS-CoV-2 whole genome characterisation. *Virus Evol*. 2020;6:veaa075.
- Nasir JA, Kozak RA, Aftanas P, et al. A comparison of whole genome sequencing of SARS-CoV-2 using amplicon-based sequencing, random hexamers, and bait capture. *Viruses*. 2020;12:895.
- Xiao M, Liu X, Ji J, et al. Multiple approaches for massively parallel sequencing of SARS-CoV-2 genomes directly from clinical samples. *Genome Med*. 2020;12:57.
- Andrés C, Piñana M, Borràs-Bermejo B, et al. A year living with SARS-CoV-2: an epidemiological overview of viral lineage circulation by whole-genome sequencing in Barcelona city (Catalonia, Spain). *Emerg Microbes Infect*. 2022;11:172-181.
- Goswami C, Sheldon M, Bixby C, et al. Identification of SARS-CoV-2 variants using viral sequencing for the Centers for Disease Control and Prevention genomic surveillance program. *BMC Infect Dis*. 2022;22:404.
- Morang'a CM, Ngoi JM, Gyamfi J, et al. Genetic diversity of SARS-CoV-2 infections in Ghana from 2020-2021. *Nat Commun*. 2022;13:2494.
- Nicot F, Tremeaux P, Latour J, et al. Whole-genome sequencing of SARS-CoV-2: comparison of target capture and amplicon single molecule real-time sequencing protocols. *J Med Virol*. 2023;95(1):e28123. doi:10.1002/jmv.28123
- Bal A, Simon B, Destras G, et al. Detection and prevalence of SARS-CoV-2 co-infections during the Omicron variant circulation in France. *Nat Commun*. 2022;13:6316.
- Feng S, Ali MS, Evdokimova M, et al. Sequencing during times of change: evaluating SARS-CoV-2 clinical samples during the transition from the delta to Omicron wave. *Viruses*. 2022;14:1408.
- ECDC. Sequencing of SARS-CoV-2: first update. January 18, 2021. European Centre for Disease Prevention and Control. Stockholm; 2021.
- Lam C, Gray K, Gall M, et al. SARS-CoV-2 genome sequencing methods differ in their abilities to detect variants from low-viral-load samples. *J Clin Microbiol*. 2021;59:e0104621.
- Arana C, Liang C, Brock M, et al. A short plus long-amplicon based sequencing approach improves genomic coverage and variant detection in the SARS-CoV-2 genome. *PLoS One*. 2022;17:e0261014.
- Brinkmann A, Ulm SL, Uddin S, et al. AmpliCoV: rapid whole-genome sequencing using multiplex PCR amplification and real-time Oxford Nanopore MinION sequencing enables rapid variant identification of SARS-CoV-2. *Front Microbiol*. 2021;12:651151.
- Li J, Wang H, Mao L, et al. Rapid genomic characterization of SARS-CoV-2 viruses from clinical specimens using nanopore sequencing. *Sci Rep*. 2020;10:17492.
- Lambisia AW, Mohammed KS, Makori TO, et al. Optimization of the SARS-CoV-2 ARTIC network V4 primers and whole genome sequencing protocol. *Front Med*. 2022;9:836728.
- Kairov U, Amanzhanova A, Karabayev D, et al. A high scale SARS-CoV-2 profiling by its whole-genome sequencing using Oxford Nanopore Technology in Kazakhstan. *Front Genet*. 2022;13:906318.

37. Zhou HY, Cheng YX, Xu L, et al. Genomic evidence for divergent co-infections of co-circulating SARS-CoV-2 lineages. *Comput Struct Biotechnol J*. 2022;20:4015-4024.
38. Dezordi FZ, Resende PC, Naveca FG, et al. Unusual SARS-CoV-2 intrahost diversity reveals lineage superinfection. *Microb Genom*. 2022;8(9):000751.
39. Combes P, Bisseux M, Bal A, et al. Evidence of co-infections during Delta and Omicron SARS-CoV-2 variants co-circulation through prospective screening and sequencing. *Clin Microbiol Infect*. 2022;28:1503.e5-1503.e8.
40. Burel E, Colson P, Lagier JC, et al. Sequential appearance and isolation of a SARS-CoV-2 recombinant between two major SARS-CoV-2 variants in a chronically infected immunocompromised patient. *Viruses*. 2022;14:1266.
41. Rosenthal SH, Gerasimova A, Ruiz-Vega R, et al. Development and validation of a high throughput SARS-CoV-2 whole genome sequencing workflow in a clinical laboratory. *Sci Rep*. 2022;12:2054.
42. Ma W, Yang J, Fu H, et al. Genomic perspectives on the emerging SARS-CoV-2 Omicron variant. *Genomics Proteomics Bioinformatics*. 2022;20(1):60-69. doi:10.1016/j.gpb.2022.01.001
43. Focosi D, McConnell S, Casadevall A. The Omicron variant of concern: diversification and convergent evolution in spike protein, and escape from anti-Spike monoclonal antibodies. *Drug Resist Updates*. 2022;65:100882.
44. Sheward DJ, Kim C, Ehling RA, et al. Neutralisation sensitivity of the SARS-CoV-2 Omicron (B.1.1.529) variant: a cross-sectional study. *Lancet Infect Dis*. 2022;22:813-820.
45. Vellas C, Del Bello A, Debard A, et al. Influence of treatment with neutralizing monoclonal antibodies on the SARS-CoV-2 nasopharyngeal load and quasispecies. *Clin Microbiol Infect*. 2022;28:139.e5-139.e8.
46. Vellas C, Kamar N, Izopet J. Resistance mutations in SARS-CoV-2 Omicron variant after tixagevimab-cilgavimab treatment. *J Infect*. 2022;85:e162-e163.
47. Ko KKK, Yingtaweessittikul H, Tan TT, et al. Emergence of SARS-CoV-2 Spike mutations during prolonged infection in immunocompromised hosts. *Microbiol Spectr*. 2022;10:e0079122.
48. Liao YC, Chen FJ, Chuang MC, et al. High-integrity sequencing of spike gene for SARS-CoV-2 variant determination. *Int J Mol Sci*. 2022;23:3257.
49. Starr TN, Greaney AJ, Stewart CM, et al. Deep mutational scans for ACE2 binding, RBD expression, and antibody escape in the SARS-CoV-2 Omicron BA.1 and BA.2 receptor-binding domains. *PLoS Pathog*. 2022;18:e1010951.
50. Ko SH, Bayat Mokhtari E, Mudvari P, et al. High-throughput, single-copy sequencing reveals SARS-CoV-2 spike variants coincident with mounting humoral immunity during acute COVID-19. *PLoS Pathog*. 2021;17:e1009431.
51. Lhomme S, Latour J, Jeanne N, et al. Prediction of SARS-CoV-2 variant lineages using the S1-Encoding region sequence obtained by PacBio single-molecule real-time sequencing. *Viruses*. 2021;13:2544.
52. Sun F, Wang X, Tan S, et al. SARS-CoV-2 quasispecies provides an advantage mutation pool for the epidemic variants. *Microbiol Spectr*. 2021;9:e0026121.

#### SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Nicot F, Trémeaux P, Latour J, et al. Whole-genome single molecule real-time sequencing of SARS-CoV-2 Omicron. *J Med Virol*. 2023;95:e28564. doi:10.1002/jmv.28564